Molecular Genetic Analysis of the *Adenomatous Polyposis Coli* (*APC*) gene region

by

Garret Malcolm Hampton

A thesis
submitted in requirement
for the degree of
Doctor of Philosophy
at the
University of London

May, 1992

Cancer Genetics Laboratory
(formerly Director's Laboratory)
Imperial Cancer Research Fund
London

and

Department of Genetics and Biometry
University College London
London University

ProQuest Number: 10609182

All rights reserved

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

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

ProQuest 10609182

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

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

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

To my parents and my wife, Kate

Abstract

Familial Adenomatous Polyposis (FAP) is a rare, autosomal dominant predisposition to colorectal cancer, affecting about one in ten thousand individuals in all populations studied. The gene responsible for this syndrome, designated *APC* (for Adenomatous Polyposis Coli) was mapped to 5q21-q22 by linkage analysis following a cytogenetic report of a male patient with polyposis and an interstitial deletion on 5q. The high incidence of allele loss at 5q21-q22 in carcinomas of sporadic patients suggests that mutation of the *APC* gene is a very frequent step in the tumorigenic pathway to nonfamilial colorectal carcinomas and emphasises the importance of isolating the gene and identifying its function.

Attempts were made to identify this gene using a positional cloning strategy, on the basis of its genomic location rather than by a knowledge of its function. As a first step toward this goal, two approaches were taken to identify a large number of DNA probes mapping within the breakpoints of two, and later three independent deletions encompassing the *APC* gene. In the first approach, a novel method, termed '*alu*-PCR' was developed. By comparing the PCR patterns generated from normal and deleted chromosomes 5, a probe was identified mapping close to the *APC* gene. In the second approach, genomic libraries constructed from physically dissected DNA around chromosomal region 5q21-q22 were used to derive a large number of DNA probes. These probes facilitated the definition of a new minimally deleted region harbouring the gene and assisted with the construction of a physical map of this region.

For the second phase of the cloning project, these DNA probes, in addition to others sub-localised to this region, were used to isolate a collection of yeast artificial chromosomes (YACs) which in total cover some 4-megabase-pairs (Mb) of DNA.

Two of the YACs identified in this study, which cover a total distance of 1.1-Mb, were used indirectly to identify potential alterations, particularly small deletions, in homologs of chromosome 5 from FAP patients. One such deletion of 260-kb was identified and shown to be entirely overlapped by one of these YACs. This deletion, and the YAC used to identify it, contain the entire coding sequence of the *APC* gene.

Acknowledgements

I am indebted to a great number of people who have helped, supported, listened and provided a thoroughly enjoyable three and a bit years at ICRF.

Particularly, I am most grateful to Sir Walter Bodmer for taking me on board and for his support in writing fellowship applications, in addition to this thesis. To Dr. Sue Povey for her endless help and support and the sometimes unpleasant task of reading this thesis in its many forms. To Dr. David Markie, without whom, on many occasions I sincerely think I would still be in the dark. I thank him also for his incisive and honest comments on reading this thesis - may the Southern blot fairy soon visit. To Ms. Sally Cottrell for putting up with so much, and to all the many members of Walter's lab, both past and present, who have helped in so many different ways.

I would also like to thank my many colleagues in the Solomon, Trowsdale, Frischauf, Sheer, Lehrach and Medical Oncology labs. for providing an excellent environment in which to work, in addition to providing some interesting Friday nights.

Outside of ICRF, I owe a great dept to my parents, Marjorie and Archie Hampton, without whom, in all honesty, I know I would not be here now (post-conceptually, that is). I thank them for their continued and loving support. To my wife, Kate, who has suffered (on my behalf) a difficult and at times trying first year of marriage, but borne the scar with grace.

Contributions

The experiments described in this thesis are entirely my own work unless otherwise stated. I am grateful for the following contributions:

Dr. Wolfgang Balhaussen, Ms. Gabriele Leuteritz and Dr. Udo Trautmann for providing microclones and the photographs in figure 4.1.

Ms. Tania Jones for the yeast *in situ* hybridisation experiments and for providing the photographs in figure 5.3.

Ms. Sally Cottrell for many probe isolates, buffers and solutions that were always dependable.

Mr. Tristan Ward for the physical mapping information depicted in figure A7.1, and for providing YAC-containing yeast strains.

Ms. Kathy Howe for providing data and microclones.

Dr. Huw Thomas for providing somatic cell hybrids.

Drs. Mark Ross, Anthony Monaco and Hans Lehrach for providing the high density YAC filters.

Dr. Anna-Maria Frischauf and Ms. Sian Searle for providing the high density cosmid filters.

The ideas for the 'junction-trapping' experiments described in Chapter 7 of this thesis came equally from both myself and Dr. Ketan Patel. The experiments described in this chapter are, however, my own work.

Table of contents.

Title.		1
Dedication.		2
Abstract.		3
Acknowledgements.		4
Contributions.		5
Table of contents.		6
List of figures.		13
List of tables.		14
Abbreviations.		15
1.	**Introduction.**	18
	Part I. Cancer genetics.	18
1.1.	A genetic basis for cancer.	18
1.2.	Congenital predisposition to cancer.	19
1.3.	Germline and somatic mutations in cancer.	21
	Part II. Colorectal cancer.	26
1.4.	Common colorectal cancer.	26
1.5.	Familial Adenomatous Polyposis (FAP).	27
1.6.	Localisation of the *APC* gene.	31
1.7.	Loss of chromosome 5 alleles and the relationship between inherited and sporadic colorectal cancer.	32
1.8.	Other genetic alterations in the progression to colorectal neoplasia.	34
	Part III. Cloning human genes.	37
1.9.	Functional cloning.	38
1.10.	Positional cloning.	40
1.10.1.	Localisation of the gene: Linkage analysis.	40
1.10.2.	Localisation of the gene: Chromosomal aberration.	41
1.10.3.	Isolation of DNA probes.	43

1.10.4. Physical characterisation of defined chromosomal regions. 46
1.10.5. Yeast artificial chromosomes. 47
1.10.6. Identification of genes in cloned DNA. 50

1.11. Aims of this thesis. 52

2. Materials and Methods. 54

2.1. Materials. 54
2.1.1. Enzymes. 54
2.1.2. Chemicals. 54
2.1.3. Nucleic acids. 55
2.1.4. Solutions and buffers. 55
2.1.5. Media. 59

2.2. General DNA methods. 62
2.2.1. Preparation of plasmid DNA. 62
2.2.2. Extraction of genomic DNA. 63
2.2.3. Restriction endonuclease digestions. 64
2.2.4. Agarose gel electrophoresis. 65
2.2.5. Southern blotting. 65
2.2.6. Radiolabelling of DNA probes. 65
2.2.7. Hybridisation. 66
2.2.8. Competitive hybridisation. 66
2.2.9. Removal of radioactive probes. 67
2.2.10. Polymerase chain reaction (PCR). 67

2.3. *Alu*-PCR methods. 68
2.3.1. *Alu*-PCR. 68
2.3.2. Cloning of inter-*alu* PCR products. 68
2.3.3. Analysis of inter-*alu* product libraries. 69

2.4. Identification and isolation of cosmids. 70
2.4.1. Gridded cosmid library screening. 70
2.4.2. Isolation of unique cosmid sub-fragments. 70

2.5. Microclone methods. 70

2.6. Yeast Artificial Chromosome (YAC) methods. 71
2.6.1. YAC library screening. 71
2.6.2. Secondary screening. 72
2.6.3. Yeast DNA preparations. 72
2.6.4. Restriction digestions of yeast DNA. 73
2.6.6. Pulsed field gel electrophoresis. 74
2.6.7. Derivation of YAC physical maps. 74
2.6.8. Comparative yeast and human Southern blot hybridisations. 75

2.7. Isolation of YAC insert termini. 75
2.7.1. *Alu*-vector PCR. 75
2.7.2. Vectorette PCR. 76
2.7.3. 'Junction-trapping' of YAC insert termini. 77

Appendix 1. Oligonucleotide primers. 79
Appendix 2. Sources of DNA probes. 81
Appendix 3. Sources of human-hamster somatic cell hybrids. 82

3. Repetitive sequence polymerase chain reaction. 84

Part 1. Rapid generation of human chromosome-specific DNA probes from a somatic cell hybrid. 84

3.1. Introduction. 84

3.2. Results. 85
3.2.1. *Alu*-PCR from a somatic cell hybrid. 85
3.2.2. Cloning of inter-*alu* PCR products. 86
3.2.3. Analysis of the 'inter-*alu*' PCR libraries. 90
3.2.4. Hybridisation analysis of 'inter-*alu*' PCR products. 90

3.3. Conclusions. 93

Part 2. Applications of the *alu*-PCR method: comparative analysis of somatic cell hybrids. 95

3.4. Introduction. 95

3.5. Results. 96
3.5.1. IRS-PCR of somatic cell hybrids: Primer IV. 96
3.5.2. IRS-PCR of somatic cell hybrids: Primers 517, 559 and LIHs. 96
3.5.3. Isolation and mapping of IRS-PCR 'difference' products. 99
3.5.4. Cloning of L5.4. 100
3.5.5. Isolation of a cosmid containing the L5.4 sequence. 100

3.6. Conclusions. 103

4. Mapping of microdissected genomic DNA probes in the *APC* gene-region. 107

4.1. Introduction. 107

4.2. Results. 107
4.2.1. Characterisation of Library I. 107
4.2.2. Characterisation of Library II. 112
4.2.3. Chromosomal *in situ* suppression (CISS)-hybridisation. 112
4.2.4. Conclusions from the initial characterisations of the microdissection libraries. 112
4.2.5. Large scale mapping of inserts from Library II. 114
4.2.6. Definition of a new minimal region encompassing the *APC* gene. 115

4.3. Conclusions. 119

5. Isolation of yeast artificial chromosomes. 125

5.1. Introduction. 125

5.2. Results 1. Isolation of yeast artificial chromosomes. 127
5.2.1. A feasibility study for YAC isolation by microclone hybridisation. 127
5.2.2. Isolation of additional YACs by microclone hybridisation. 128

5.2.3. Conclusions from the microclone hybridisation experiments. 138
5.2.4. Isolation of a YAC for probe cL5.4.s-1. 132
5.2.5. Isolation of YACs for the *MCC* gene. 132
5.2.6. Probes coincident to the same YAC(s). 132
5.2.7. Summary of candidate YAC clone detection. 133

5.3. Results 2. Assessment of candidate YAC clones. 133
5.3.1. Secondary screening. 133
5.3.2. Analysis by pulsed field gel electrophoresis. 133
5.3.3. Comparative human and yeast Southern blot hybridisation. 134
5.3.4. Fluorescence *in situ* hybridisation (FISH) of total YAC-containing yeast DNAs. 134

5.4. Conclusions. 141

6. <u>Ascertainment of YAC contigs 1.</u> Coincident hybridisation analysis and physical mapping. 145

6.1. Introduction. 145

6.2. Results. 146
6.2.1. Coincident mapping. 146
6.2.2. YAC physical mapping. 146
6.2.3. YAC contigs. 153

6.3. Conclusions. 159

7. <u>Ascertainment of YAC contigs 2.</u> Isolation of YAC-insert terminal sequences. 161

7.1. Introduction. 161

7.2. Results 1. Isolation of YAC insert termini. 163
7.2.1. *Alu*-vector polymerase chain reaction (PCR). 163
7.2.2. Conclusions on the *alu*-vector method. 166
7.2.3. 'Vectorette'-polymerase chain reaction (PCR). 166
7.2.4. Conclusions on the 'vectorette' method. 172

7.3. 'Junction-trapping': a new approach to the isolation of YAC-insert termini. 172

7.3.1. Introduction and strategy. 172

7.4. Results 2. 'Junction-trapping' of YAC termini in yeast B0624. 174
7.4.1. Preparation of insert DNA. 174
7.4.2. Ligations. 179
7.4.3. Generation of candidate termini PCR fragments. 179
7.4.4. Assessment of candidate termini fragments. 179
7.4.5. Mapping of the candidate insert termini fragments. 183
7.4.6. 'Junction-trapping' as a useful strategy. 183

7.5. Results 3. Coincidence mapping with YAC insert termini. 186

7.6. Conclusions. 186

A7. Appendix to Chapter 7: Physical mapping summaries. 191

8. Identification and characterisation of the *Adenomatous Polyposis Coli* gene-region. 197

8.1. Introduction. 197

8.2. Results. 199
8.2.1. Direct YAC screening on chromosome 5-specific cosmid libraries. 199
8.2.2. Direct cosmid screening. 200
8.2.3. Identification of an interstitial deletion. 201
8.2.4. Identification of the *APC* gene. 205
8.2.5. A detailed YAC-based physical map of the *APC* gene region. 206
8.2.6. Relative transcriptional orientation of the *MCC* and *APC* genes. 211
8.2.7. TB2 is associated with a CpG island. 211

8.2.8. Deletion breakpoints of 3214. 211
8.2.9. The maximal overlap of YACs A1010 and B0624 is 110kb. 212
8.2.10. Cosmids ym72 and ym75 contain portions of the *APC* gene. 212

8.3. Conclusions. 212

9. Discussions. 216

9.1. Isolation of DNA probes in the *APC* gene-region. 216
9.1.1. Interspersed repetitive sequence (IRS)-PCR. 216
9.1.2. Microdissection and Microcloning. 219

9.2. Identification and characterisation of YACs in Interval II. 222
9.2.1. YAC isolation. 222
9.2.2. YAC overlap detection. 223
9.2.3. YAC chimerism. 225

9.3. Identification of genes at 5q21-q22. 226
9.3.1. The *MCC* gene. 226
9.3.2. YACs for the *MCC* gene. 227
9.3.3. The *APC* gene. 227
9.3.4. A possible role for *MCC* in colorectal tumorigenesis ? 230

9.4. Future studies. 231

10. Bibliography. 234

11. Publications arising from this thesis. 268

List of figures.

1.1.	Surgical specimen of an affected colon from an FAP patient.	29
3.1.	Position and sequence of *alu*-based oligonucleotides.	87
3.2.	*Alu*-directed amplification of hybrid PN/Ts-1.	89
3.3.	Analysis of an inter-*alu* PCR product library from hybrid PN/Ts-1.	92
3.4.	Comparative repetitive sequence PCR analysis of normal and deleted chromosomes 5.	98
3.5.	Southern blot analysis of probes L5.4 and cL5.4.s-1.	102
4.1.	Microdissection of chromosome region 5q21-q22.	110
4.2.	Southern blot analysis of microclone inserts.	118
4.3.	Regional localisation of microclone inserts.	120
5.1.	Identification of positive YAC signals with probe mc534.	131
5.2.	Pulsed field Gel electrophoresis (PFGE) and comparative Southern blotting of potentially positive YAC clones.	136
5.3.	Fluorescence *in situ* hybridisation (FISH) of total yeast DNAs on human chromosome metaphase spreads.	138
6.1.	Partial restriction enzyme digest mapping of the YAC in yeast ICRFy900D073.	152
6.2.	Inferred restriction maps and contig structures.	156
7.1.	Isolation and characterisation of YAC terminus '714LHE'.	165
7.2.	Principal of the 'Vectorette-PCR' method.	168
7.3.	Generation and characterisation of 'vectorette' PCR products.	171
7.4.	Positions of pBluescript- and pYAC4-specific amplification primers.	176
7.5.	Schematic illustration of the 'junction-trapping' method.	178
7.6.	Generation and primary characterisation of YAC insert termini from two YACs in yeast ICRFy900B0624.	181
7.7.	'Fragment-5' maps to YAC 624-575.	185
A7.1.	Summary of physical mapping in Interval II.	194
8.1.	Identification of an interstitial deletion in patient 3214.	204
8.2.	Construction of a detailed physical map of YAC A1010.	208
8.3.	A detailed physical map of the *APC* gene-region.	210

List of tables

4.1a. Characteristics of microdissection libraries. 108

4.1b. Preliminary analysis of microclone inserts from microdissection libraries I and II. 108

4.2. Characteristics of deleted clones from preliminary library analyses. 113

4.3. Summaries of large scale mapping analysis. 116

5.1. Identification of YACs by probe hybridisation. 129

5.2. Characteristics of YACs isolated in this study. 142

5.3. Characteristics of additional independent YACs localised to Interval II. 143

6.1. Detection of coincident hybridisation on Interval II YACs. 148

7.1. Isolation and characterisation of YAC insert termini. 187

8.1a. Hybridisation of YACs on chromosome 5-specific cosmid libraries. 202

8.1b. Direct screening of YAC-identified cosmids on hybrid panels. 202

Abbreviations used in this thesis.

A	adenine residue in a DNA sequence
APC	Adenomatous Polyposis Coli (gene designation; McAlpine et al., (1987))
ATP	Adenosine 5'-triphosphate
BCIG	5-bromo-4-chloro-3-indolyl β-D-galactopyranosidase
bp	base-pairs
BSA	bovine serum albumin
C	Cytosine residue in a DNA sequence
Cot	the product of concentration (moles nucleotide per litre) and time (seconds) in a DNA reassociation procedure
dATP	2'-deoxyadenosine 5'-triphosphate
DCC	Deleted in Colorectal Carcinoma (gene designation; Fearon et al., (1990)
dCTP	2'-deoxycytosine 5'-triphosphate
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	ethylenediamineteraacetic acid
FAP	Familial Adenomatous Polyposis
G	guanine residue in a DNA sequence
g	grams
GS	Gardner's Syndrome
β-gal	β-galactosidase
IPTG	isopropyl β-D-thiogalactopyranosidase
kb	kilobase-pairs
L	litre
λ	Lambda
M	moles
Mb	megabase-pairs
MCC	Mutated in Colorectal Carcinoma (gene designation; Kinzler et al., (1991a)
μg	micrograms
μl	microlitre
μM	micromolar
mg	milligrams
ml	millilitre
mM	millimolar

mRNA	messenger ribonuleic acid
PCR	polymerase chain reaction
PEG	polyethyleneglycol
PERT	phenol emulsion reassociation technique
RFLP	restriction fragment length polymorphism
RNAse	ribonuclease
SSC	saline sodium citrate
T	thymine residue in a DNA sequence
Tris	tris(hydroxymethyl)aminomethane
U	units
UV	ultraviolet
v/v	volume for volume
w/v	weight for volume
YAC	yeast artificial chromosome

Additional abbreviations are explained in the appropriate text. See also Chapter 2, section 2.1.4.; solutions and buffers.

Chapter 1.

Introduction.

Chapter 1

Introduction.

Part I. Cancer genetics.

1.1. A genetic basis for cancer.

Extensive studies suggest that cancer is, fundamentally, a genetic disease at the cellular level. This is not meant to imply that cancer is a genetic disease in the same sense as classical human genetic diseases caused by single gene defects (McKusick, 1991), but rather that cancer in the general population is initiated and progresses by the acquisition of somatic mutation at the level of DNA. Although inherited cancer predisposition syndromes do exist, and, in their own right clearly demonstrate a genetic basis for cancer, epidemiological studies, which have demonstrated significant variation in the incidence of common cancers amongst different populations (e.g. Peto and Doll, 1981), in addition to extensive twin-studies (e.g. Holm et al., 1982), tentatively suggest that most human cancers are not caused by major inherited components.

Some of the principal lines of evidence for a genetic basis of cancer include: (1) the fact that carcinogens (chemical and radiation) are also often mutagens and have been shown to interact with DNA, (2) the observation of consistent chromosomal aberrations in certain cancers, particularly in the leukemias and lymphomas (e.g. Solomon et al., 1992). Chromosomal aberrations exist also in human solid tumours, but these are not always consistent. Key examples, however, include deletion of genetic material from chromosome 3p in renal cell and small cell lung carcinomas, deletions around 11p in Wilm's tumour of the kidney (and associated phenotypes) and deletion around 13q in retinoblastomas (Mitelman, 1988). These findings suggest the occurrence of specific somatic mutations in the neoplastic cell clone, (3) The recessive inheritance of defects in DNA repair, which lead to increased somatic mutation rates and an associated increase in the rate of tumour formation (reviewed in Friedberg, 1985).

Cancer is thought to be the end result of a multi-stage process. For most tumour types, the incidence of cancer increases as a function of age. This has been used as evidence to suggest that the formation of a cancer is

dependent on the acquisition of a number of discrete somatic mutations (e.g. Nordling, 1953; Armitage and Doll, 1954; Ashley, 1969a). If only a single somatic mutation was required for the formation of a cancer, then one would expect that most cancers would appear as a linear function of age, which, on the basis of epidemiological evidence, is not the case. A double logarithmic curve measuring the the death rate from a particular cancer versus the incident age of the cancer, for most cases, gives a straight line, the slope of which is a crude representation of the numbers of mutations required for cancer formation. It has been suggested that, in general, some five to six mutations are required for most common cancers (Ashley, 1969a). However, these values, which are quite variable amongst human cancers (see Ashley, 1969a), are entirely dependent on the nature of the cell types in which the mutations occur, and, to some extent, on whether these mutations increase the chance or potential for the occurrence of further genetic change (Fisher, 1958).

At the level of the cell, the multi-stage process is probably best exemplified by the common observation of a steady increase in the numbers and types of chromosomal aberrations during tumour progression, particularly in common solid adult tumours (reviewed in Sager, 1986). At the level of the individual, the process is revealed, in part, by the observation of pre-cancerous lesions, such as adenomas of the large bowel in Familial Adenomatous Polyposis (FAP) patients, which display some, but not all of the features of the fully malignant phenotype.

1.2. Congenital predisposition to cancer.

Predisposition to cancer with a marked familial tendency has long been recognised, particularly by observing individuals within families that exhibit pre-cancerous phenotypes not commonly found in the general population; for example, multiple colonic adenomas in Familial Adenomatous Polyposis (Bussey, 1975) or, more strikingly perhaps, the appearance of breast cancer in men (see King, 1990). Familial predisposition has also been recognised on the basis of groups within the general population that show a shifted age-incident distribution in tumours of common sites (i.e., an earlier age of onset), such as those of the colon and breast. Although inherited cancers account for probably no more than 1% of all cancers, they have led to a considerable insight into the genetic mechanisms underlying tumour initiation and formation.

Congenital predisposition does not result in the inheritance of cancer, *per se*, but rather in an increased susceptibility to cancer. For example, in families that show a clear pattern of inheritance to retinoblastoma, a rare childhood cancer of the retinal cells (discussed below), individuals have been described that do not manifest this cancer, but who are the sibs of an affected parent and, who, in turn, pass this defect to their offspring (see Knudson, 1985 and references therein). This strongly suggests that further steps must be required for the appearance of the tumour.

For virtually every form of human cancer, families have been reported in which the cancer appears to be inherited (Knudson, 1973; see also King, (1990), for a recent review). In the vast majority of cases, the susceptibility to cancer is dominantly inherited, although recessive predispositions to cancer also exist (Friedberg, 1985). For example, in Xeroderma Pigmentosa (XP), which is characterised by extreme sensitivity to sunlight, resulting in freckling and atrophy followed by benign growths and subsequently malignant skin tumours, the cells are unable to repair DNA damaged by ultraviolet radiation. Thus, the mutation causes a systemic effect in increasing the relative somatic mutation rate and therefore leading to a high incidence of skin and other cancers (Cleaver, 1968).

Of the dominantly inherited cases, two broad classes can be tentatively assigned. In the first, families have been recognised exhibiting a constellation of tumour types, such as in the Li Fraumeni syndrome where an index individual often exhibits a rare tumour-type (e.g. rhabdomyosarcoma; Li and Fraumeni, (1975)) and in patients with multiple endocrine neoplasia (DeLellis et al., 1986). In addition, individuals with certain developmental syndromes, such as Neurofibromatosis types 1 and 2 and the Beckwith-Weidemann syndrome, are at an apparently increased risk of cancer, exhibiting tumour-types that are also generally rare in the population as a whole.

The second, and perhaps most important class from a genetic point of view, are those families with a very high risk of developing a single tumour-type (rare or common), that arise sporadically in the general population. These syndromes, which are for the most part highly penetrant, are extremely valuable for linkage analysis, facilitating the localisation and eventual isolation of the genetic locus responsible for the

defect. In this respect, the pathology and histology of the sporadic and inherited cancers of common sites are generally indistinguishable and, although individuals with an inherited susceptibility are likely to have an earlier age of onset, the likelihood of recurrence and survival (after adjustment for age) are essentially the same. Thus, from a biological point of view, the inherited and sporadic cancers may be genetically similar, if not identical (see below), which clearly emphasises the importance of isolating and characterising these genes.

1.3. Germline and somatic mutations in cancer.

There is considerable evidence to suggest that mammalian cancer can arise and progress, for the most part, by the recessive abrogation of 'tumour-suppressor' genes that serve in the normal cell to control growth. This class of genes contrasts with the 'oncogenes' which confer a dominant phenotype in cell culture (see Bishop, 1987) and which are generally involved in the progression of human cancers by the acquisition of somatic mutation. The suppressor genes, which provide an explanation for the elusive genetic relationship of inherited and sporadic cancers, were originally inferred from two entirely separate strands of evidence.

In the first, somatic cell fusion of malignant cells with normal diploid cells, in general, led to a cessation of growth in immune deficient animals (e.g. Harris et al., 1969; Harris and Klein, 1969) suggesting the 'donation' of a dominant suppressive 'factor' from the normal cell. By extension, these results implied that the malignant cells did not contain this suppressive ability, having lost it during the progression to malignancy.

Critically, the reappearance of malignancy in rare hybrid cell segregants was correlated, in most cases, with the loss of specific chromosomes. For example, the reappearance of malignancy in a number of hybrid mouse cells was correlated to a loss of chromosome 4 (Jonasson et al., 1977; Evans et al., 1982) and, in humans, the reappearance of malignancy in hybrid cells derived from the fusion of a cervical tumour-derived cell line (HeLa) and normal diploid fibroblasts, was specifically correlated to loss of chromosomes 14 and 11 (Stanbridge et al., 1981). In support of these observations, the introduction of a single normal chromosome 11 into HeLa cells similarly shows a suppressive effect (Saxon et al., 1987). Reappearance of malignancy in these cases has been correlated with a loss

of the normal chromosome. A large body of similar studies exist, and suggest that the genetic suppression of malignancy, or at least the suppression of various features of the malignant phenotype, in a wide variety of human tumours is a general phenomenon (reviewed in Stanbridge, 1990).

The second line of evidence came from studies on rare childhood cancers with heritable tendencies, particularly retinoblastoma, but also Wilm's tumour of the kidney, neuroblastoma and phaeochromocytoma (Knudson, 1971; Knudson and Strong, 1972a; Knudson and Strong, 1972b).

Retinoblastoma is a rare childhood cancer of the eye, occurring at a frequency of about 1 in 14,000 live births (see Goodrich and Lee, 1990). Tumours can occur in both unilateral and bilateral forms (that is, in one or both eyes), and, in total about 40% of all retinoblastoma cases are inherited, the remainder occurring sporadically. Familial cases, in general, develop multiple bilateral tumours, but it has been estimated that some 10-15% of unilateral cases may also be inherited (Knudson, 1971; Knudson, 1985). In contrast, sporadic retinoblastoma cases are almost totally unilateral and, in addition, are distinguishable in that rarely more than one retinal tumour is found.

Knudson (1971) found that the number of tumours appearing in familial cases (that is individuals with no tumours, with unilateral tumours or bilateral tumours) appeared to fit with a poisson distribution, suggesting that the manifestation of a retinal tumour was the result of a random, independent event with a finite probability - roughly reflecting somatic mutation rates. The age incidence of hereditary tumours, in addition, fitted a model assuming linear tumour kinetics, suggesting a requirement for only a <u>single</u> 'hit' in tumour formation. Thus, in addition to the segregation of a germline defect, which has no visible effect on the phenotype of affected cells, a further somatic event must occur. In contrast, for cases without any familial predisposition, the incidence of a retinal tumour was in accord with an exponential function of age, that is, a requirement for <u>two</u> discrete 'hits'. This predicted that in sporadic cases two somatic events would have to occur in a single retinal cell, one of which would be equivalent to the germline defect in hereditary cases, and would, therefore, be rare (i.e., the square of the somatic mutation rate). In

both hereditary and sporadic cases, the formation of a tumour would be dependent on the acquisition of a somatic mutation in the same gene.

Taking these two disparate lines of evidence into account, Comings (1973) proposed an incisive model to suggest that human cancers could result from the homozygous abrogation of diploid regulatory genes which serve, in the normal cell, to suppress the effect of dominant transforming genes. In this model, it was suggested that congenital predisposition to cancer could arise by the inheritance of a mutated regulatory gene which would not confer a phenotypic effect. The appearance of the tumour would then be dependent on mutation of the remaining wild-type gene on the homologous chromosome. The presence of a normal allele in obligate mutants would be sufficient to suppress tumour formation, in line with the suppressive effect demonstrated by the introduction of a single normal chromosome into a tumour derived cell line (Saxon et al., 1987). The dominant phenotype of an inherited cancer would then be dependent on the second event occurring with sufficient probability. In sporadic cases, Knudson's two 'hits' would represent the somatic abrogation of both copies of the regulatory gene.

Experimental evidence for the functional inactivation of a susceptibility locus in retinoblastoma (referred to as Rb-1), came from a number of cytogenetic studies reporting deletions in familial cases centred around 13q14 (e.g. Francke and Kung, 1976; see Knudson et al., 1976). In addition, the co-segregation of the Rb phenotype with alleles of the esterase D (EsD) gene located on chromosome 13 (Sparkes et al., 1980; Sparkes et al., 1983) suggested that an Rb-1 gene on chromosome 13q was the target of germline mutation, in most, if not all cases. The observation of 13q deletions in the tumour cells from sporadic patients (Balaban et al., 1982) confirmed the genetic relationship between inherited and sporadic Rb cases, in that the Rb-1 locus was also the target of somatic mutation. Subsequently, one familial Rb case with a syntenic deletion of Rb-1 and EsD was reported and found to be homozygous for the null EsD allele (Godbout et al., 1983), strongly implicating the Rb-1 locus as the target for Knudson's second 'hit'.

More formal evidence for the homozygous mutation of Rb-1 came from a molecular genetic study using localised restriction fragment length polymorphisms (RFLPs) to follow the segregation of chromosome 13

alleles in normal and tumour cells (Cavenee et al., 1983). They were able to demonstrate a number of mechanisms by which somatic homozygosity could arise, with chromosome non-disjunction (with or without somatic reduplication) and mitotic recombination being particularly frequent events (see Cavenee, 1991, for a detailed discussion). Cavenee et al., were also able to demonstrate that it was the chromosome from the normal parent that was lost in familial cases, cementing the recessive hypothesis.

The observation of mechanisms occurring at sufficiently high frequencies that could reveal the existence of suppressor genes has had profound consequences for examining the relationship between inherited and sporadic cancers (e.g. reviewed in Lasko et al., 1992; see Seizinger et al., 1992). In most cases to date, linkage of a cancer syndrome to a particular region in the genome, has been confirmed by the observation of allele-loss affecting the same region in tumour cells. Indeed, positional cloning strategies (see part III of this introduction) have been greatly complemented by the observation of interstitial deletions and mitotic recombination events which serve to sub-localise, sometimes considerably, the position of the gene. Such approaches have allowed the cloning of strong candidates for genes that predispose to Wilm's tumour of the Kidney (e.g. Gessler et al., 1990) and retinoblastoma (Friend et al., 1986; Fung et al., 1987; Lee et al., 1987).

Some exceptions to this generalisation exist, and warrant caution in assuming that there is always a simple genetic relationship between inherited and sporadic cancers. For example multiple endocrine neoplasia type-2 syndrome (MEN2A), which is characterised by the appearance of medullary thyroid carcinomas and phaeochromocytomas, has been linked to chromosome 10. However, tumour cells do not exhibit loss of alleles in the same region. Indeed, allele-loss has only been consistently observed on chromosome 1 (Landsvater et al., 1989; Nelkin et al., 1989). Also, in Neurofibromatosis type-1 (NF-1), for which a strong candidate has been isolated on 17p11.2 (Wallace et al., 1990; Viskochil et al., 1990), there is considerable doubt as to whether allele-loss occurs in NF-1 tumours at this site (Menon et al., 1990). Conversely, in some kindreds exhibiting Wilm's tumour of the kidney, linkage has not been demonstrated on 11p (Grundy et al., 1988; Huff et al., 1988). This is in contrast to the vast majority of cases where the appearance of childhood Wilm's tumours is associated with loss of alleles on 11p.

The observation of allele-loss has, in addition, provided strong evidence to suggest that most common human cancers, such as lung, breast and colon, in addition to many others, progress though the consistent (and sometimes extensive) loss of suppressor genes, some of which are shared between disparate tumour-types (reviewed in Seizinger et al., 1992). Allele-loss in these cancers, in genomic regions where no inherited cancer predisposition appears to map, has, nonetheless, led to the identification of a number of critical suppressor genes, such as p53 on 17p (Baker et al., 1989) and DCC (Deleted in Colorectal Carcinoma) on 18q (Fearon et al., 1990). Mutations in the p53 gene (discussed in part II), have been implicated in a large variety of tumours, such as colon, breast, lung, brain and many others. In fact few cancers have been described that do not show mutations in this gene (Lasko et al., 1992; de Fromentel and Soussi, 1992). As a corollary to identifying genes by virtue of linkage with a cancer syndrome, these genes represent strong candidates for being inherited as mutants in syndromes exhibiting cancers for which such mutations have been described. Germ-line mutations of the p53 gene have subsequently been found in Li Fraumeni individuals (Malkin et al., 1990), who are predisposed to a sub-set of the tumour types in which p53 mutations have so far been described (de Fromentel and Soussi, 1992).

For some of the suppressor genes that have been cloned (e.g. Rb-1 and p53), in addition to some of the chromosomes on which these genes are inferred to lie (see Stanbridge, 1990), there is now considerable evidence to suggest that these genes, albeit in diverse ways (Weinberg, 1992), do serve to regulate normal cell growth control (see Lasko et al., 1992; Stanbridge et al., 1990). Transfer of whole human chromosomes into tumour-derived cell lines of different types, although largely unable to precisely pinpoint the genes or numbers of genes involved, appears to be a generally valid model in dissecting the role of many suppressor genes in a single tumour (e.g. Goyette et al., 1992), in addition to confirming the suspicions of suppressor gene localisation inferred by allele-loss studies (Stanbridge, 1990).

Part II. Colorectal cancer.

1.4. Common colorectal cancer.

Common or 'sporadic' colorectal cancer represents a very significant contribution to mortality in industrialised western countries. The annual mortality rate is second only to lung cancer in Britain and the United States. For example, in 1985, some 24000 new cases, and over 17000 deaths were registered in England and Wales (Office of Population Censuses and Surveys, 1985a; 1985b).

Epidemiological studies, which show a considerable geographic variability in the incidence of colon cancer (e.g. Boyle et al., 1985), and which, in addition, have shown that migrants commonly acquire the host country incidence rates (e.g. Washauer et al., 1986), suggest a major role for environmental factors, particularly diet, in the aetiology of colorectal cancer (Armstrong and Doll., 1975; Willett, 1989).

A combination of studies investigating mortality, incidence amongst relatives and, more recently, large pedigree analysis have, however, provided evidence to suggest a potential contribution of inherited factors in 'common' colorectal cancer aetiology (reviewed in Bishop and Thomas, 1991). Genetic models suggest the existence of a major gene, albeit with a very low frequency and a generally low life-time penetrance in the population as a whole (Burt et al., 1985; Cannon-Albright et al., 1988).

It is generally accepted that carcinoma of the colon, in the vast majority of cases at least, arises by clonal expansion from an adenomatous polyp, first suggested by Lockhart-Mummery in 1925 (cited in Bulow, 1987). This progression along a well defined histopathological sequence is commonly referred to as the 'adenoma-carcinoma' sequence. A large body of evidence exists, although largely indirect and circumstantial, to support this idea (reviewed in Morson, 1974a; Morson 1974b; Morson et al., 1983; Burt and Samowitz, 1989). Perhaps the most salient arguments are (1) the observation of a morphological associations between adenomas and adenocarcinomas, where the latter eventually becomes the predominant tissue aberration (Morson, 1966) and, (2) the fact that persistent removal of adenomas from patients with Familial Adenomatous Polyposis (see below) can reduce the expected incidence of rectosigmoid cancer by as

much as 85% (Fenoglio and Pascal, 1982). More recently, support for this idea has come from the observation of consistent somatic genetic alterations that appear to accumulate in the progression from adenoma to carcinoma (Vogelstein et al., 1988).

The idea of a progressive adenoma-carcinoma sequence correlates well with the statistical notion of cancer incidence as a function of age. It has been suggested that cancer of the colon may be the end result of as many as five to six discrete steps (Ashley, 1969b). Colonic crypts developmentally arise from single progenitor cells (Ponder et al., 1985) which maintain their status throughout adult life (Griffiths et al., 1988). The pathway from immature progenitor cells to differentiated colonic epithelial cells occurs over a period of four to five days. Thus, the colon in man, effectively undergoes complete cell turn-over (in the sense of a balanced equilibrium) in one week (Fenoglio-Preiser and Hutter, 1985). This has been used to suggest that somatic mutation rates in man are probably sufficient to account for the appearance of colorectal cancer during the life-time of an individual.

1.5. Familial Adenomatous Polyposis (FAP).

The existence of a highly penetrant and dominantly inherited syndrome in which individuals are specifically predisposed to colorectal cancer has long been recognised (Cripps, 1882; Lockhart-Mummery, 1925; cited in Bulow, 1987). This syndrome, most commonly referred to as Familial Adenomatous Polyposis (FAP), is of particular interest in that the development of colorectal tumours appears to progress along the same histopathological continuum described for common or 'sporadic' colorectal carcinoma (i.e., the adenoma-carcinoma sequence). Because of the highly penetrant autosomal dominant character of FAP (see below), this syndrome has provided an ideal model for identifying the genetic basis of the initiating events involved in colorectal cancer in the general population.

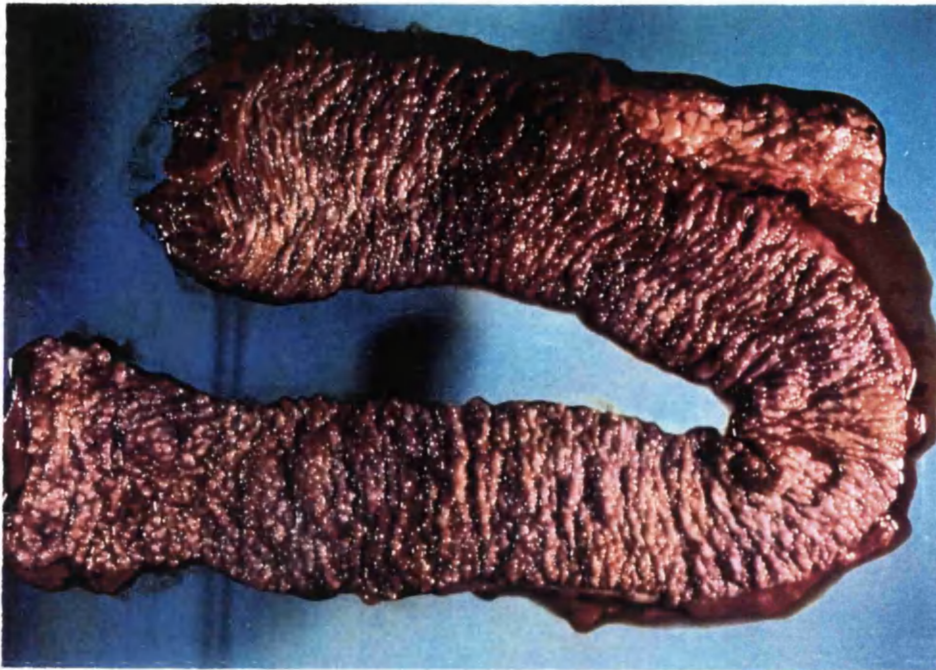
FAP is characterised by the appearance of multiple adenomatous polyps in the colon and rectum at a median age somewhere between 15 to 16 years (Murday and Slack, 1989; Bulow 1986, respectively). Although there are no defined upper and lower limits, the number of polyps can range from between about 150 to over 5000 (Bussey, 1975). Indeed, cases have been

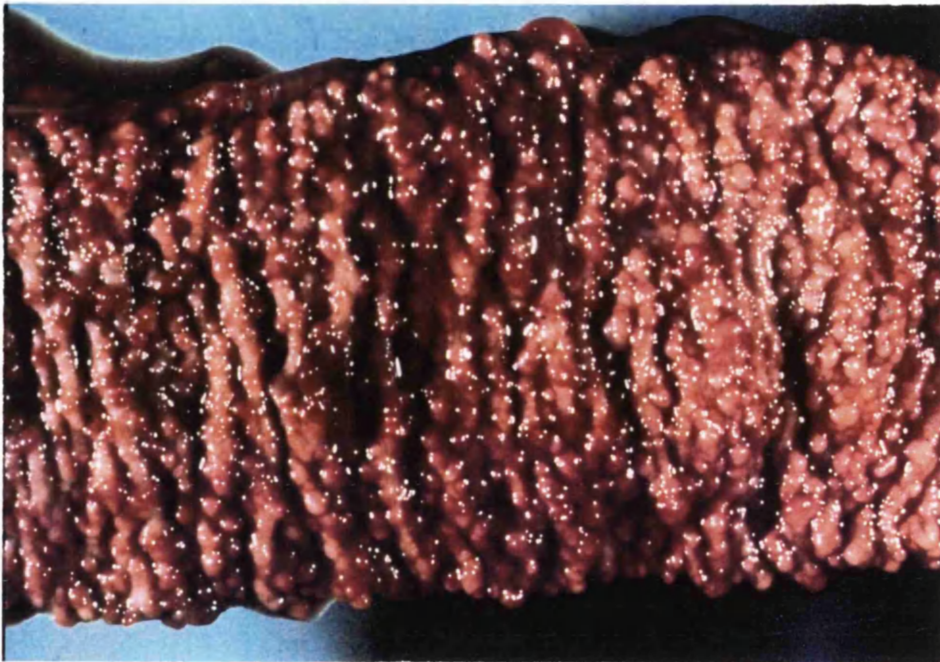
found in the St. Mark's Hospital series (London), where the colon is so carpeted with polyps that the normal colonic mucosal lining can hardly be seen (Bussey, 1982). Clinically, the appearance of at least 100 adenomatous polyps has been suggested as the guideline for diagnosis of polyposis coli to differentiate those cases exhibiting multiple 'sporadic' adenomatous polyps (Bussey, 1975). An example of a colon surgically removed from an individual with FAP is shown in figure 1.1.

In general, for individuals without a previously identified risk of FAP, the first symptoms, such as increasing bowel movements, mucous discharge and occasional rectal bleeding, are only realised at a median age of 30 years (Bulow, 1986) and many polyps may then be in the one centimetre size range (Bussey, 1982). At the time of diagnosis, usually prompted by severe rectal bleeding, as many as 60% of patients will already have colorectal cancers at one or more sites (Bussey, 1982). This predisposition to colorectal cancer in untreated individuals is very aggressive, resulting in premature death at around 40 years (Dukes, 1952; Bulow, 1986). This sharply contrasts with the generally late age of diagnosis (60 to 65 years) in patients diagnosed with common colorectal carcinoma (Bussey, 1975).

Incidence calculations, or risk estimates for FAP, have come from polyposis registries, the most reliable of which is probably the Danish register which is essentially complete (Bulow, 1986). The cumulative incidence in the Danish population has been estimated at 0.97×10^{-4} (i.e., approximately 1 in 10000 live births), over the period 1976-82 (Bulow et al., 1986; Bulow 1987). Although this figure is in approximate agreement with estimates from Sweden, Finland, Japan and Britain, most of the risk estimates in the latter groups are based on the comparison of death from colon cancer versus normal matched controls. These results are therefore biased by the improved prognosis of identified FAP individuals in recent times (Bulow, 1987). There appears to be no sex bias in the populations studied to date, consistent with autosomal inheritance (Bulow, 1986; Alm and Licznerski, 1973; Bussey, 1975).

Germline mutation rates for the gene that predisposes to FAP (*APC*; McAlpine et al., 1987) have been calculated from essentially complete polyposis registries. The rate is of the order of about 1×10^{-5} per generation (e.g. Reed and Neel, 1955) with almost complete penetrance (Veale, 1965; Utsunomiya, 1989). This rate appears to be essentially similar in many





<u>Figure 1.1.</u> Surgical specimen of an affected colon from an FAP patient.

In the top part of the figure, the bowel has been opened up longitudinally to reveal the mucosal surface carpetted with small polyps. The terminal ileum is at top left and the rectum at bottom left. The lower part of the figure shows the sigmoid colon in more detail where the polyps, in general, can be seen as small protrusions above the mucosal surface.

different populations suggesting that FAP is maintained by a mutation-selection balance (see Bodmer, 1989). Such a high mutation rate implies that most families will segregate independent mutations. Indeed, isolated cases of FAP without any obviously affected family members, occur in the Swedish register at a frequency of 30% (Alm and Lecznerski, 1973) and ranges up to as high as 47% in the Danish register (Bulow, 1987). Some of these apparently 'new mutations', however, may be due to incomplete ascertainment of family history.

A wide variety of extra-colonic manifestations have been described in patients with polyposis coli. Principally, these include mandibular osteomas and epidermoid cysts (Gardner, 1951; Gardner and Plenk, 1952; Gardner and Richards, 1953; Gardner, 1962) which, in combination with colonic polyposis, are referred to as Gardner's Syndrome (GS). GS was originally thought to be clinically distinct to FAP because of the apparent familial clustering of these manifestations. However, the finding of one or more of these manifestations in a significant proportion of 'simple' polyposis patients, particularly in Denmark (Bulow et al., 1984; Bulow, 1986) and Japan (Utsunomiya and Nakamura, 1975), suggest that they should, in fact, be considered as traits of FAP syndrome (see Utsunomiya, 1989; Cohen, 1982, for discussions). One other interesting anomaly, congenital hypertrophy of the retinal pigment epithelium (or CHRPE; Blair et al., (1980)) has, in addition, been increasingly associated with colonic polyposis. This condition, which is a benign ocular lesion and most often bilateral, has become a strong candidate for effective presymptomatic diagnosis of FAP as it appears to be highly penetrant in polyposis families (Polkinghorne et al., 1990; Chapman et al., 1989).

Polyposis patients, in addition, appear to be at a significantly increased risk of other tumour-types, particularly in the upper gastrointestinal tract, such as gastric and duodenal polyps, in addition to desmoid tumours (Bulow, 1987; Domizo et al., 1990). There also appears to be a significantly high risk of brain (Leeds Castle Meeting Report, 1988) and other tumours, such as hepatoblastoma and neurofibromatosis, which are generally rare in the population as a whole (Garber et al., 1988; Pratt et al., 1988).

These extra-colonic manifestations, which are in some cases at least, inextricably associated with polyposis, are likely to be a consequence of

mutations at the *APC* locus, which, in contrast to sporadic cases of colorectal cancer, are present in each cell of polyposis patient.

1.6. Localisation of the *APC* gene.

The highly penetrant nature of the FAP phenotype, in addition to the availability of well characterised families through polyposis registries, make this syndrome eminently suitable for the localisation of the causative gene by standard linkage analysis, although cytogenetic clues have considerably expedited these analyses. Veale at one time reported linkage for the polyposis disease locus with the MNSs blood group determinants. However, this was later retracted (Veale, 1965). Gardner et al., (1982) then reported the apparently consistent finding of an heteromorphism, tentatively defined as a deletion of q14.3-q21.3 on chromosome 2 in seventeen patients. These findings, however, could not be corroborated.

A further report by Herrera and colleagues at Roswell Park in Buffalo, U.S.A., was to prove correct. Herrera et al., (1986) reported a patient with multiple polyposis, mental impairment, developmental anomalies and a heterozygous deletion on chromosome 5q, around regions q13-q15 or q21-q22. Linkage analysis with an anonymous DNA probe, c11p11 (D5S71), recognising a polymorphic locus on chromosome 5, led Bodmer et al., (1987) and Leppert et al., (1987) to demonstrate linkage with polyposis. The LOD scores (log of the odds) were in excess of 3 in each case, suggesting combined odds of at least 1 in 10^6 against such a linkage association by chance alone. This marker was then localised to 5q21-q22 by *in situ* hybridisation by Bodmer and colleagues (Bodmer et al., 1987), a region entirely consistent with the deletion described by Herrera et al., (1986). These results were almost immediately confirmed by Meera Khan et al., (1987).

At least one definitive recombination between the disease locus and this marker was subsequently reported by Aldred et al., (1988). The identification of further recombinations has demonstrated that the probe is, in fact, located some 9 to 10 centiMorgans (cM) centromeric to the disease locus (Dunlop et al., 1989; Dunlop et al., 1990). Nakamura et al., (1988) further sub-divided this region by genetic analysis with 6 newly identified polymorphic loci. The minimum genetic distance between two

markers, YN5.64 and YN5.48, thought to flank the *APC* gene, was calculated to be about 10cM.

1.7. Loss of chromosome 5 alleles and the relationship between inherited and sporadic colorectal cancer.

Concomitant with the demonstration of linkage with FAP on chromosome 5q (Bodmer et al., 1987; Leppert et al., 1987), Solomon and colleagues (Solomon et al., 1987) investigated the possibility of recessive change involving the *APC* gene by analogy with the common mechanisms associated, in most cases, in familial and sporadic tumorigenesis (Knudson, 1971; Cavenee et al., 1983). They were able to demonstrate loss of heterozygous chromosome 5 alleles in between 20 to 40% of carcinomas derived from sporadic cases assumed to encompass the *APC* gene. In addition, they were able to infer a significant reduction to homozygosity in a number of tumour-derived cell lines (50% observed versus 8% expected). These figures are likely to be underestimated since the most informative locus-specific marker, λMS8, is located at 5qter.

These studies were extended and largely confirmed, with frequencies of chromosome 5 allele-loss in sporadic carcinomas generally in the range of about 30% (Okamoto et al., 1988; Vogelstein et al., 1988; Rees et al., 1989). More recent investigations with polymorphic markers closely flanking the *APC* gene, have suggested a frequency as high as 60% (Ashton-Rickardt et al., 1989). These studies demonstrated the long suspected relationship between familial and sporadic colorectal cancer in that allele-loss, which is usually the second event, implied the homozygous somatic inactivation of the *APC* gene in sporadic colorectal tumours. Colorectal carcinoma in FAP patients is, however, a secondary consequence of mutations in the *APC* gene since it is the appearance of adenomatous polyps that characterises the polyposis phenotype. Thus, one would expect to observe loss of heterozygous alleles on chromosome 5 in adenomas, in line with the recessive hypothesis (Knudson, 1985). Initial studies suggested that this was not the case (Solomon et al., 1987; Okamoto et al., 1988; Law et al., 1988).

However, a large and incisive study by Vogelstein and colleagues, using polyp tissue highly enriched for neoplastic cells, subsequently demonstrated loss of heterozygosity at a similar frequency (26-39%) in

medium and large sized adenomas as well as carcinomas from sporadic cases (Vogelstein et al., 1988). This clearly implicated loss of the *APC* gene as an early step in colorectal neoplasia. Vogelstein et al., (1988) were, however, unable to demonstrate loss of heterozygous alleles in any of 34 small adenomas from FAP patients; an observation that has been consistently reported (Solomon et al., 1987; Law et al., 1988).

This would suggest a fundamental difference in neoplastic progression in inherited and sporadic cases. However, allele-loss has since been demonstrated at least in large adenomas (Rees et al., 1989) and carcinomas (Sasaki et al., 1989; Rees et al., 1989) from FAP patients suggesting that this is probably not the case. Because of an inherent early ascertainment bias that will exist in FAP patients with a familial history, no studies have yet been carried out to compare, directly, allele-loss in small adenomas from both inherited and sporadic cases. In sum, the results imply that colorectal neoplasia, in contrast to other inherited cancer syndromes, is not necessarily initiated by homozygous inactivation of the *APC* gene.

This presents an inherent difficulty in formulating a genetic model for the expression of recessive alleles in familial colorectal tumour progression. Although each adenoma is probably derived from a clonal expansion of cells (Fearon et al., 1987), this expansion, in the initial stages at least, does not appear to be associated with the loss of the remaining wild-type alleles flanking the *APC* gene (Solomon et al., 1987; Vogelstein et al., 1988).

The situation can, however, be reconciled if one modifies the recessive hypothesis by assuming that being heterozygous for a mutant allele is sufficient to allow a phenotypic effect even in the presence of a wild-type allele. In this respect, Bodmer et al., (1987), have suggested that heterozygous mutations at the *APC* locus are sufficient, in ***totatum***, for the appearance of adenomatous polyps. They have proposed that adenoma initiation may be epigenetic in nature if the *APC* gene product were to have a negative effect on the production or steady state concentrations of epithelial cell growth-factors.

If the normal homozygote produces sufficient gene product to ensure maintenance of a critical threshold, it may be that the deficient heterozygote could allow random fluctuations below this critical level resulting in localised hyper-proliferation of colonic epithelial cells.

Vogelstein et al., (1988) suggest that inactivation of the remaining wild-type allele may magnify this hyper-proliferation, but would not be necessary for the effect. They also point out that other epigenetic effects, such as hypo-methylation, which is the only consistent difference observed between normal and dysplastic colonic epithelial cells (Golez et al., 1988; Feinberg et al., 1988) may promote a proliferative effect, viz-a-viz abnormal mitotic segregation of chromosomes (Schmidt et al., 1985), in revealing other partial-dominant loci.

In sum, these results suggest a critical role for mutations in the *APC* gene, in both inherited and sporadic colorectal cancer cases. Moreover, Miyaki et al., (1990) have recently demonstrated that for inherited cases, the chromosome from the normal parent is lost in the tumour, strongly supporting the analogy, in terms of formal genetics at least, with other inherited cancer syndromes (Knudson et al., 1985).

1.8. Other genetic alterations in the progression to colorectal neoplasia.

A number of other consistent somatic mutations involved in colorectal neoplastic progression have been described (Fearon and Vogelstein, 1990), the number of which is largely consistent with the incidence of colorectal cancer increasing as a function of a fifth to sixth power of age (Ashley, 1969b).

The acquisition of somatic mutations confering a dominant effect in colorectal neoplasia was originally inferred from cell transformation assays using DNA from tumour samples (Needleman et al., 1983). Although a number of alterations in oncogenes have been described in colorectal tumours, such as amplification of *c-myb* (Alitalo et al., 1984) and the *neu* oncogene (D'Emillia et al., 1989), the only consistent somatic mutation so far described has been found in the cellular homolog of the murine Kirsten-*ras* oncogene (c-Ki-*ras*). Activating mutations, most commonly found in codon 12, have been observed in up to 50% of colorectal carcinomas (Forrester et al., 1988; Bos et al., 1988) and at a similar frequency in more dysplastic adenomas greater than one centimetre in size. In contrast, *ras* gene mutation in adenomas of one centimetre or less in diameter occurs in only about 10% of cases (Vogelstein et al., 1988; Farr et al., 1988). These results, in addition to the fact that not all patients

exhibit *ras* gene mutations (Vogelstein et al., 1988), suggests that the acquisition of somatic mutations in the *ras* gene is not an initiating event, but more likely a later event, perhaps giving rise to a more dysplastic polyp.

Following provocative cytogenetic reports of clonal alterations involving chromosomes 17 and 18 in eleven patients with colorectal tumours (Muleris et al., 1985), loss of heterozygous alleles on chromosome 17, particularly on the short arm, was demonstrated in 76% of cases (Fearon et al., 1987). This was subsequently confirmed in similar studies (Law et al., 1988; Monpezat et al., 1988; Vogelstein et al., 1988). Examination of adenomas highly enriched for neoplastic cells suggested that loss of 17p alleles is a generally late event in tumour progression, largely coinciding with the transition from adenoma to adenocarcinoma (Vogelstein et al., 1988).

Loss of heterozygosity on chromosome 17 was shown to be centred on 17p12-p13.3, a region known to contain the gene encoding the tumour associated protein, p53 (Baker et al., 1989). Sequencing of p53 alleles in two tumours in which the other allele had been lost, revealed two independent somatic mutations, strongly implicating this gene as the target for somatic alteration on 17p in colorectal tumours (Baker et al., 1989). Further studies have revealed that mutation in p53 is a common event in colorectal tumours, in addition to a wide variety of tumour-types frequently exhibiting 17p allele-loss (Nigro et al., 1989; reviewed in de Fromentel and Soussi, 1992).

Initial studies of mutant copies of the p53 gene in cell transformation assays suggested that they behave essentially as activated oncogenes, particularly in co-operation with other activated genes, such as mutant *ras* (e.g. Parada et al., 1984). On the surface, this was inconsistent with the idea of p53 as a tumour suppressor gene. However, Baker et al., (1989) proposed that mutant copies of the p53 gene may act in a partial-dominant manner where the mutant gene not only losses normal growth suppressive function, but also interferes with the function of the remaining wild-type gene product. Consistent with this proposal, Nigro et al., (1989), have demonstrated the existence of at least one tumour in an 'intermediate' state, that is with one mutant and one wild-type allele. In the absence of any wild-type p53 alleles in colorectal tumour-derived cell

lines, re-introduction of a fully functional gene under a strong promoter has a potent suppressive effect (e.g. Baker et al., 1990).

Loss of heterozygous alleles on the q arm of chromosome 18 was also demonstrated in between 52% (Law et al., 1988) and 73% (Vogelstein et al., 1988) of tumours studied. Inactivation of a further suppressor gene on 18q, like the p53 gene on 17p, appears to occur during late adenoma formation (Vogelstein et al., 1988). Further allele-loss studies with a polymorphic locus at 18q23.1 revealed a homozygous deletion in one tumour and a suggestive alteration in one other. A gene spanning this region, termed DCC (Deleted in Colorectal Carcinoma) was subsequently identified (Fearon et al., 1990). Intriguingly, DCC has significant homology to a class of cell adhesion molecules, most closely resembling the neural cell adhesion molecule (N-CAM). This is a particularly provocative finding in that abrogation of cell adhesion or cellular communication would be consistent with DCC as a tumour suppressor gene.

Initial reports from Okamoto et al., (1988) suggested a major role for loss of constitutional heterozygosity on chromosome 22 in a study of colorectal tumours from Japanese polyposis patients. Although this study was confirmed by further investigation of Japanese cases (Sasaki et al., 1989), attempts to reproduce these studies, in the U.S. at least, has not been successful (Law et al., 1988; Vogelstein et al., 1989). Additional observations of apparently significant allele-loss frequencies at a number of other sites, such as chromosomes 6, 12q, 14 and 15 (Okamoto et al., 1988; Sasaki et al., 1989), have also met with little or no confirmation, except perhaps for loss of 6q alleles (Vogelstein et al., 1989). The significance of these observations, if any, may lie in alternative but perhaps equivalent accumulation of genetic changes in the progression to malignancy in the Japanese population.

An extensive chromosomal study of colorectal tumours, referred to as an 'allelotype', have been carried out (Vogelstein et al., 1989; Kern et al., 1989). The key feature of these studies is the large amount of chromosomal loss, particularly in patients with a generally poor prognosis (see also Kern et al., 1989). A number of chromosomal regions, particularly 8p, appear to be lost in up to 50% of cases, with the majority of other acrocentric arms exhibiting losses in the range of about 7-30%.

Although consistent somatic genetic alterations do occur, the salient feature of colorectal tumorigenesis appears to the accumulation of genetic change, rather that a rigorous order in which the individual changes must occur (Vogelstein et al., 1988). It is generally accepted that mutations in the *APC* gene are probably the earliest events, inferred from allele-loss studies, but also on the basis that inherited mutations in the *APC* gene give rise to multiple polyposis in individuals with FAP (Vogelstein et al., 1988). After localised colonic epithelial proliferation and polyp formation, somatic mutation in the Ki-*ras* gene, or mutations in other genes conferring an equivalent genetic effect, may occur to magnify the proliferative effect, probably after the initial formation of the polyp. Loss of 18q alleles (i.e. the DCC gene) appear to be the next most common event, followed by loss of chromosome 17p alleles (i.e., the p53 gene). Many tumours have, however, been observed that do not follow the preferred order, and, in addition, none of the genetic changes so far described appear to be mutually exclusive.

Loss of alleles (for example on 17p and 18q), are usually second events and may, in fact mask the existence of much earlier initiating mutations in these genes. In the case of p53, this does not appear to be the case, as antibody-staining against the mutant form of p53 is only detected late in colorectal tumour progression (Rodrigues et al., 1990).

Whether any of the clonally observed chromosomal losses at about 30% or above (Vogelstein et al., 1989) are significant or not is unclear. They may represent chromosomal regions with other as yet unidentified tumour suppressor genes, or may simply reflect an increase of mitotic instability in the tumour cell. In either case, they will inevitably appear clonal, whether they result from 'irreversible' non-disjunction or whether they confer a selective growth potential on the expanding tumour cell population.

Part III. Cloning human genes.

Considerable progress has been made in molecular genetic technology over the last decade. The present-day capacity to clone genes largely stems from the development of a number of key technologies, principally: the discovery of enzymes that can cleave DNA at specific nucleotide sequences (Smith and Wilcox 1970, Kelly and Smith, 1970); the use of self-replicating episomal vectors that allow propagation of specific DNA fragments in

bacterial hosts (see Berg et al., 1974, and references therein); the use of elegant chemical methods to 'read' the base sequence of DNA (Maxam and Gilbert, 1977; Sanger et al., 1977) and the ability to detect specific DNA sequences immobilised on solid supports by radiolabelled probe hybridisation (Southern, 1975). This basic collection of fundamental technology has, more recently, enormously benefitted from the ability to amplify specific DNA fragments by the polymerase chain reaction (PCR; see Saiki et al., 1988 and references therein). PCR has had dramatic consequences for human genetic studies, particularly in expediting previously tedious laboratory techniques, but also in providing entirely new strategies; from the generation of sub-chromosomal DNA probes (e.g. Nelson et al., 1989; Cotter et al., 1990; Ludecke et al., 1989) to the rapid detection of DNA sequence variation (e.g. see Hayashi, 1991).

1.9. Functional cloning.

The ability to clone specific genes responsible for human genetic disorders has, in the past at least, often relied on information concerning the part or step of a biochemical that has been affected. Where the biochemical defect is known at the protein level; for example in the haemoglobin molecules which give rise to the vast array of 'haemoglobinopathy' phenotypes (e.g. Weatherall, 1990), the genes can be cloned by 're-constructing' likely parts of the gene from partial amino acid sequence information, and selection of candidate DNA fragments from genomic libraries.

Previously unidentified genes, however, can also be identified without this prior biochemical knowledge, if, for example, the gene confers a dominant phenotype in cell culture. In the case of aggressively transforming oncogenes, the DNA fragment(s) from human tumours that give rise to the effect (i.e., the oncogene) can be recovered from transformed mouse cell foci (e.g. Tabin et al., 1982). Similarly, genes for novel products, such as cell-surface determinants, can be selected for by antibodies after transfection of whole cDNA libraries into cells that do not express these products (e.g. Seed and Aruffo, 1987). Adaptations of this basic technique can also be modified to complement for the loss or absence of biological function. For example, the loss of cell-matrix attachment ability in metastatic tumour derived cell lines can be complemented by the introduction of a cDNA library constructed from similar cells derived from tumours that have not undergone metastasis (e.g. Pullman and

Bodmer, 1992). Even though mutations in tumour suppressor genes represent a biological loss of function, the function of these genes (only one of which has, for example, been implicated in cell surface activity (DCC; Fearon et al., 1990)), is still largely obscure (Weinberg, 1992). Suppressor genes, in addition, are unlikely to display highly selectable phenotypes in cell culture. Some of these genes, however, clearly suppress the formation of tumours in immune deficient mice (see Stanbridge, 1990) but, the idea of selecting one or more cDNA clones that confer this effect at the level of the organism is practically and, to some extent, ethically untenable.

'Loss of function' mutations may also be amenable to selection by subtractive techniques at the level of gene expression. In such an approach, cDNA libraries from mutant cells would be compared to cDNA libraries from 'wild-type' cells by hybridisation of both with the wild-type library as a probe. At the outset of these studies, such an approach was seriously considered for identification of the *APC* gene. However, the mutation status in available cell lines derived from colorectal tumours was unknown; specifically, it was not clear at what level the gene would be expressed. For example, it the mRNA was transcribed at entirely normal levels, or even at sub-normal levels, the strategy would dictate that all differences in expression (which would invariably include those apparently different due to technical artifacts) should be examined.

Independent attempts using this approach by two groups resulted in the selection of a few hundred sequences either not expressed or more lowly expressed in carcinoma biopsies versus normal colonic epithelium (Augenlicht et al., 1987; Elvin et al., 1988). Although some of these cDNAs may well represent targets of mutation in the progression to malignancy, the numbers involved suggest a laborious and generally untenable strategy where no information on gene expression is available.

In sum, these methodologies, referred to as 'functional cloning' (Collins, 1992) have been considerably successful in identifying a few hundred genes. However, they rely on a knowledge of the gene product (or indications about the level of gene expression) or depend on the selection of a well defined phenotype. For the majority of congenital or somatic human genetic defects (McKusick, 1991), there are generally few if any clues as to the nature of the altered DNA sequences responsible, in

addition to phenotypes that are not, in most cases at least, amenable to cell biological techniques.

An alternative approach, however, best termed 'positional cloning', aims to identify a gene on the basis of genomic location, rather than a knowledge of its function. This approach has been considerably successful, having resulted in the cloning of thirteen genes previously undescribed, in addition to nine known genes whose involvement in specific disorders could be inferred purely on the basis of genomic location (see Collins, 1992). Positional cloning follows a strictly genetic approach, typically hierarchical in nature. The key elements of the general strategy are outlined below.

1.10. Positional cloning.

1.10.1. Localisation of the gene: Linkage analysis.

Natural variation that exists at the DNA sequence level in complex genomes has led to the identification of both protein and DNA polymorphisms. Polymorphic variants can allow the differentiation of homologous chromosomes since each individual inherits one chromosome of each autosomal pair from, in the general population, unrelated parents. The first human sequence variants uncovered at the DNA level were noted in the β-globin gene (Kan and Dozy, 1979) through differences in restriction endonuclease sites giving rise to differing fragment lengths (referred to as RFLPs). RFLPs, which were found to be quite common (see White, 1988 for an historical review) allow one to follow, in a genotypic manner, the segregation of specific alleles in individuals of a family, thus providing the basis for extensive linkage analysis in humans. This potential led Solomon and Bodmer (1979) to propose the construction of a linkage map of the entire human genome, which was more practically enlarged upon by Botstein et al., (1980).

Linkage, stated simply, assesses the likelihood of two or more loci being co-inherited. The further apart two markers are, the less likely that they will be co-inherited (and therefore 'linked') due to separation by meiotic recombination. Mathematical models (Morton 1955, Morton, 1957) allow the calculation of this likelihood as a LOD score or a logarithmic function of the odds of linkage versus non-linkage. The LOD score can be

maximised with reference to a recombination fraction, giving an estimate of the most probable distance between two loci. This genetic distance is expressed in centiMorgans (cM), where one cM is a notional distance over which one recombination would be expected during one hundred meioses (Morgan, 1911). The comparison of physical distance (i.e., the length of DNA) to genetic distance has suggested that one cM is approximately equivalent to about one million base-pairs (1Mb). However, the difference that exists in recombination rates between males and females and, in specific regions of the genome (for example on 21q, Cox and Shimizu, (1992)), warrants caution in this assumption.

For families exhibiting a near fully penetrant phenotype, such as familial adenomatous polyposis (FAP), a battery of RFLPs can be used to assess the likelihood of a polymorphic variant segregating with the phenotype, and therefore with the disease locus. For other disorders, the phenotype may be subject to misclassification (genetically heterogeneous), which, in addition to a social pressure for fewer offspring in western nation families, makes the general method more difficult. In the most ideal situation, two or more loci may be linked to the disease locus and flank either side of the gene which then defines boundaries in which the gene must lie. Subsequent multi-locus linkage analysis (e.g. Lathrop et al., 1984) with additional markers can further sub-define the gene's position to within a few cM. Linkage analysis below this level, however, becomes difficult since a very large number of meioses are required to increase the probability of recombination between a closely linked marker and the disease locus.

1.10.2. Localisation of the gene: Chromosomal aberration.

The identification of chromosomal aberration, which largely depends on the spectrum of mutation exhibited by the gene in question, can greatly expedite both the location of a gene of and, in addition, provide a significant resource for subsequent gene identification. The cloning of most genes approached through genomic location have depended, critically, on such aberrations. For some genes, however, such as cystic fibrosis (CF), the spectrum of mutation is severely limited, having arisen only a few times in the general population and does not appear to include structural aberration (NIH workshop on population screening for the cystic fibrosis gene, 1990). A similar situation is also apparent in

Huntington's disease, for which the defective gene (if it is, indeed, a gene; Laird, (1990)) remains to be cloned. Although chromosomal deletion and translocation are by far the most common structural abnormalities, special cases of 'rearrangement' do exist: for example, the appearance of multiple repeat elements in Myotonic dystrophy, X-linked and spinal Bulbar muscular atrophy and Fragile-X, which are detectable at the level of the Southern blot (reviewed in Richards and Sutherland, 1992).

The critical assumption, particularly in the case of large deletions and translocations, is that these aberrations affect the gene in question. It is of course possible - although unlikely - that the appearance or even segregation of such anomalies is peripheral to the disease. In the case of the *APC* gene, cytogenetically visible deletions have been observed in a number of patients, usually associated with other phenotypic anomalies, such as mental impairment and developmental dysmorphologies. One of these, identified by Herrera et al., (1986), proved pivotal in suggesting the genomic location of the *APC* gene on chromosome 5, subsequently confirmed by linkage analysis (Bodmer et al., 1987; Leppert et al., 1987). Since the mapping of the gene, a number of further deletions have been identified (Hockey et al., 1989; Varesco et al., 1989; Cross et al., 1992), including one deletion type present in two brothers (Hockey et al., 1989).

Apart from the obvious contribution of chromosomal anomalies in suggesting gene location, the altered chromosome can, in addition, be segregated in the absence of the normal homolog by the construction of inter-species chromosomal hybrids. This relies on the fusion of two cells from different species (e.g., Harris and Watkins, 1965; Pontecorvo, 1975), with the subsequent loss of human chromosomes in an apparently random manner (e.g. Nabholz et al., 1969). The chromosome of interest can be selected for either genetically in cell culture by virtue of genes on that chromosome, as first shown by Harris and Watkins, (1965), or, as more recently demonstrated, by PCR sib-selection (Markie et al., in press). For mapping studies, a deletion provides a 'window' in which the gene must lie. Segregation of the deleted chromosome in the absence of the normal homolog allows the rapid localisation of probes within this window, and suggests that the probes will necessarily be close to the gene. A number of somatic cell hybrids segregating chromosomes 5 with interstitial deletions have been constructed (Varesco et al., 1989; Thomas, 1991) and are described in the results chapters of this thesis.

In parallel with the finding of constitutional alterations in the germ-line of cancer-syndrome patients, the observation of allele-loss in tumours provides a wealth of information for the further sub-localisation of the gene-region. Due to the natural variation of breakpoints created by somatic deletion or mitotic recombination which, for the most part, occur at high frequency to select against a tumour suppressor gene, the gene can be narrowed down, in some cases quite considerably, by the overlap of the deletions or the sites of recombination (e.g. Ashton-Rickhardt et al., 1989; see also Baker et al., 1989). Moreover, the observation of one or more homozygous deletions in tumour cells would further indicate a very small region in which to assess candidate genes. Although homozygous deletion should be a very rare event, the use of a large number of tumours or tumour-derived cell lines should increase the probability of their identification (see for e.g., Fearon et al., 1990; Gessler et al., 1990). Allele-loss, in some cases used as the only tool for gene-localisation, has been disproportionately successful in identifying genes by positional cloning because of its inherent power in narrowing the region of interest (e.g. Baker et al., 1989; Fearon et al., 1990).

1.10.3. Isolation of DNA probes.

Once a gene of interest has been localised to within boundaries defined either by the smallest region of deletion breakpoint overlap and/or by linkage analysis, a framework needs to be constructed within this region. In the absence of other information, this is essential in understanding the nature of the region in question, that is, the physical genomic distance involved and the relative location of landmark probes best suited to genomic expansion by the selection of larger insert clones (Monaco et al., 1986; Rommens et al., 1990; Fearon et al., 1990).

A large number of methods have been devised to enrich for DNA sequences within specific chromosomal regions which depend, critically, on the resources available, i.e., whether deletions have been identified and so forth. These include phenol reassociation cloning (PERT), which can allow the enrichment of sequences from a normal chromosome that are absent from a deleted chromosomal region (Kunkel et al., 1985); selection and regional mapping of cosmids or bacteriophage clones from whole (e.g. Nakamura et al., 1988) or fragmented human chromosomes isolated in

somatic cell hybrids (Thomas, 1991) or obtained by flow-sorting (Krimlauf et al., 1982; Nizetic et al., 1991); cloning of DNA sequences from regionally localised large restriction fragments isolated from somatic cell hybrids (Michiels et al., 1987; Varesco et al., 1989; Pritchard et al., 1990) and cloning of DNA sequences bounding infrequently cutting restriction sites (e.g. Pohl et al., 1988; Borrow et al., 1990).

Other methods, more recently devised, include the selection of chromosomal or region-specific DNA sequences by repetitive sequence PCR (Nelson et al., 1989; Cotter et al., 1990; Brookes-Wilson et al., 1991). The development of this technique, colloquially termed '*alu*-PCR', and some of its applications, are presented in Chapter 3.

Perhaps the most direct strategy to enrich for DNA probes is by the physical microdissection of chromosomal sub-regions. In 1981, Edstrom and colleagues reported the first successful microdissection and microcloning experiments on segments of the *Drosophila* X chromosome (Scalenghe et al., 1981). Using microforged siliconised glass needles, a 1Mb segment of the X chromosome was dissected 6 times under phase microscopy with the aid of a crude micro-maniplulator. After purification and restriction digestion of the DNA in nanolitre volumes, the fragments were cloned into a bacteriophage vector and 80 clones recovered. *In situ* hybridisation demonstrated that the clones were derived from the region of physical dissection.

Although technically difficult, dissection of *Drosophila* chromosomes almost certainly represents an 'easier' model system since polytenisation of fruit-fly chromosomes in the salivary glands effectively increases the amount of DNA by some 2000-fold per dissection as compared to mammalian chromosomes. Nonetheless, this technique was successfully applied to the proximal half of mouse chromosome 17, known to contain the t-complex region (Rhome et al., 1984). Over 270 dissections of this region were performed and 212 microclones obtained. Many of these clones have since provided a valuable resource for the physical mapping of the t-complex region (Hermann et al., 1986). Subsequently, other mammalian chromosomal regions, such as proximal mouse X (Fisher et al., 1985) and distal human 2 (Bates et al., 1986) have been dissected and microclone libraries successfully generated.

The methods devised by Edstrom and colleagues, however, suffer from a number of limitations. Firstly, dissections have to be performed under oil using precision 'bent' microforged needles which limits the accuracy and the angles at which dissections can be performed. Thus, a large number of metaphase preparations have to be made for each set of dissection experiments. Secondly, the chromosomal regions to be dissected are difficult to identify since the metaphase chromosomes cannot be 'banded' by standard techniques if visualised under oil. Thirdly, due to the amount of DNA needed for standard bacteriophage cloning, a great number of dissections have to performed. For example, over 100 in the case of human chromosome 2 to generate approximately 2 picograms (pg) of DNA for cloning (Bates et al.,1986; Dr. G. Bates, personal communication).

More recently, two technical advances have led to the development of a highly efficient method for the microdissection and microcloning of 'band-specific' libraries which alleviate many of the difficulties described above (Ludecke et al., 1989). Firstly, a method was developed for rapid karyotyping in prenatal diagnosis ('pipette method'; Claussen et al., 1986). This involves the selection of a limited number of cells by micro-pipette extraction directly from a colchicine arrested culture. The cells are lysed *in situ* and rapidly fixed in the micro-pipette (about 10 seconds) before dropping onto coverslips. This method, although technically difficult, produces spreads of very high quality with only minimal exposure to acidic solutions. Further, advances in inverted microscope design enabled dissection of chromosomes in air at magnifications as high as x 1250, which contrasts to a magnification limit of ≈ x 650 by inverted oil microscopy. Chromosomes could therefore be pre-banded by giemsa-trypsin stains (GTG) prior to dissection resulting in a resolution down to specific chromosomal sub-bands.

Secondly, Horsthemke and colleagues developed a 'Universal Amplification' technique that allows sequence independent amplification. In this system, the dissected DNA is digested with a frequently cutting restriction enzyme, such as RsaI, and the fragments are ligated into a SmaI site of a modified plasmid. Amplification is carried out using primers designed to either side of the plasmid cloning site; the generally small size of the ligated DNA fragments ensuring that most will be represented after PCR. Ligation of RsaI fragments which are 'blunt-ended', to the SmaI site

in the plasmid (also 'blunt-ended'), allows for selection against non-recombinants by re-digestion with SmaI (see Ludecke et al., 1989).

These two developments, once combined, were successful in generating a microdissection library from around the Langer-Gideon syndrome region of chromosome 8 using chromosomal DNA from only 20 dissections (200 to 300 fentograms (fg)). This library contained some 2 x 10^4 clones, most of which were found to be independent (only 1 of 50 characterised clones represented twice). A number of 'band-specific' libraries have since been constructed by this method including the regions known to harbour the genes or sequences responsible for Prader-Willi syndrome (Buiting et al., 1990), Fragile-X syndrome (MacKinnon et al., 1990; Hirst et al., 1991a), Neurofibromatosis type-2 (Fiedler et al., 1991), Wilms Tmour Region I (Davis et al., 1990) and Adenomatous Polyposis Coli (Hampton et al., 1991a; 1991b; see Chapter 5). Over 99% of 105 randomly selected clones from five of these libraries examined to date appear to map back to the region of dissection (see Ludecke et al., 1990). The specificity of the dissections has also recently been elegantly demonstrated by the use of whole microdissection libraries as probes on metaphase spreads (Trautmann et al., 1991; Hampton et al., 1991a).

1.10.4. Physical characterisation of defined chromosomal regions.

Markers derived by any one or a combination of the methods above can be used in the construction of regionally ordered physical maps. Restriction digestion of mammalian DNA with endonucleases recognising nucleotide sequences rich in CpG dinucleotides results in the generation of DNA fragments many thousands of base pairs in length, partly because of the paucity of such CpG clusters in the genome, but also because many of these 'islands' are methylated *in vivo*. These fragments can be resolved by pulsed field gel electrophoresis (Schwartz and Cantor, 1984) and restriction maps constructed by hybridisation of Southern blots containing DNA fragments generated by combinations of 'rare-cutting' enzymes (Barlow and Lehrach, 1987). Unmethylated CpG islands have been shown to be frequently associated with the 5' ends of transcribed genes, more often house-keeping than cell-specific (reviewed in Bird, 1987).

More recently it has been possible to order markers on random human chromosomal fragments generated by ionising irradiation and retained in

somatic cell hybrids (Goodfellow and Pritchard, 1988; Benham et al., 1989; Cox et al., 1990). Recent work on chromosome 21 has shown that ordering of markers can be achieved down to a resolution of about 50-100kb (Cox et al., 1990). However, one of the more exciting advances in long range mapping has come from the ability to propagate extremely large DNA fragments as artificial chromosomes in yeast (Burke et al., 1987).

1.10.5. Yeast artificial chromosomes

The sequences essential for autonomous propagation of a linear chromosome in the yeast *Saccharomyces cerevisiae* have been cloned and characterised (see Murray and Szostak, 1983). Linear propagation requires three essential *cis*-acting sequences; (1) centromeric DNA (CEN) which allows appropriate meiotic and mitotic segregation, (2) autonomous replicating sequences (ARS) which function as the origin of chromosomal DNA replication and (3) telomeric repeat sequences (TEL) which stabilise and allow faithful replication of chromosomal ends. Experiments demonstrating increased stability with increased length of the chromosome in yeast (Murray and Szostak, 1983), led to the development of the original pYAC vector (Burke et al., 1987) which can allow the insertion and stable propagation of foreign DNA up to, or even greater than 1000kb in length (McCormick et al., 1989; Larin et al., 1991).

The vector, in isolation, is propagated in bacteria by inclusion of pBR322 sequences conferring antibiotic resistance (ampicillin) and an origin of replication (*ori*). The stable propagation of the artificial chromosome in yeast can be selected by vector complementation of defective uracil (URA3) and tryptophan (TRP1) genes in the yeast host (AB1380). Insertion of foreign DNA is indicated by disruption of an ochre tRNA suppressor gene at the cloning site giving rise to the red *ade2* phenotype in recombinants (Burke et al., 1987).

A large number of libraries, prepared either by total digestion of genomic DNA with infrequently cutting enzymes or, more commonly, by partial digestion with EcoRI, have now been cloned in both standard and modified pYAC vectors (reviewed in Schlessinger, 1990). These represent genomic libraries from plant (Guzman and Ecker, 1988), bacteria (Kuspa et al., 1989), *C. elegans* (Coulson et al., 1988), *Drosophila* (Garza et al., 1989), human (e.g. Brownstein et al., 1989; Little et al., 1989; Anand et al., 1989;

Larin et al., 1991) and mouse (e.g. Larin et al., 1991). Libraries from more specific sources; such as inter-species somatic cell hybrids (e.g. Abidi et al., 1990), have also been constructed.

In general, libraries have been constructed with insert sizes certainly larger than in prokaryotic systems. However, technical difficulties associated with the preparation and ligation of large DNA fragments (reviewed in Burke, 1990), has typically limited insert sizes to a range of 225kb to about 370kb (Brownstein et al., 1989; Anand et al., 1989). Recent developments in minimising sheer forces during DNA manipulation with synthetic polymers has led to libraries with average insert sizes of 620kb to 700kb from human and mouse genomes, respectively (Larin et al., 1991). With average insert sizes as large as these, complete genome coverage can be attained in some 5000 clones, in contrast to about 8×10^4 clones for a library propagated in cosmid vectors. These YAC libraries can be screened either by radiolabelled probe hybridisation (Brownstein et al., 1989) or by PCR-based protocols (Green and Olson, 1990a), the relative merits of which are discussed in detail in Green et al., (1991b) and Monaco et al., (1992).

Cloning in YACs offers several distinct advantages in positional cloning strategies. Firstly, it has become apparent that DNA sequences 'unclonable' in prokaryotic based vector systems, such as cosmids or bacteriophage, can be faithfully propagated in yeast (Coulson et al., 1988; Schlessinger et al., 1991; see also Chapter 10). Secondly, effective genomic coverage of sub-defined genomic regions can be accomplished in relatively few 'steps', either by selection of a large number of overlapping YACs (contig) from a region saturated with a large number of random DNA markers (e.g. Green and Olson 1990b), or by genomic 'walking' from a few key landmark YACs (e.g. Silverman et al., 1989; Silverman et al., 1990). Already, contiguous maps have been constructed for lower eukaryotic genomes like *C. elegans* (Coulson et al., 1988), large gene loci such as Duchenne Muscular Dystrophy (Monaco et al., 1992; Coffey et al., 1992), gene complexes (e.g. HLA; Ragoussis et al., 1991) and large chromosomal regions, such as Xq28 (Little et al., 1992).

The ability to 'walk' specific chromosomal regions in YACs relies on the isolation of the terminal sequences from the genomic inserts. These terminal sequences can be used, either by hybridisation or a PCR-based strategy, to re-select overlapping YACs from the original library. Although

walking is conceptually simple, and for prokaryotic based systems at least, can be readily achieved (Monaco et al., 1986; Fearon et al., 1990), the situation in YACs is certainly more difficult (Nelson, 1990). A number of methods exist, each with their own relative merits. Perhaps the most straight-forward approach is the construction of bacteriophage or cosmid libraries generally from whole YAC-containing yeast DNA (Coulson et al., 1988). These libraries can be hybridised to sequences specifically recognising the 'left' and 'right' arms of the pYAC4 vector. This method is, however, time consuming and extremely labour intensive for a large number of YACs, particularly where other sequences from the insert are not required.

Another related method, 'circularisation' requires the digestion of the YAC DNA with a restriction enzyme, such as XhoI, that cuts the 'right' vector arm distant to the *E. coli* ampicillin resistance gene (*amp*) and in the insert DNA close to the EcoRI pYAC4 cloning site. Re-circularisation of the digested DNA in dilute ligation reactions and transformation to *E. coli* allows the recovery of molecules containing the vector-insert XhoI fragment by genetic selection. The paucity of human genomic XhoI sites and the non-trivial re-circularistion method (see Traver et al., 1989) renders this method generally unsuitable. Moreover, in the pYAC4 vector, only the 'right' hand end fragments can be isolated.

With the advent of PCR, this method has been adapted to allow the generation of DNA fragments representing the vector-insert junctions by 'inverse-PCR'. In this protocol, the yeast DNA is digested with restriction enzymes that cut distant to the primer binding sites in the pYAC4 vector and close to the cloning site in genomic DNA. After re-circularisation, 'inverse' PCR is carried out using two vector-specific oligonucleotides priming in opposite orientations such that a linear PCR fragment is generated and bounded by vector sequences. This method has the advantage of being able to generate 'end'-fragments from both vector arms and has been used to significant effect in genome walking (e.g. Silverman et al., 1989; Silverman et al., 1990)

A novel method for insert termini isolation, termed 'vectorette-PCR' has recently been developed (Riley et al., 1990). Yeast DNA is digested with restriction enzymes, like those for the 'inverse-PCR' method, that cut distant to vector-specific primer binding sites and randomly in genomic

DNA. After digestion, a synthetic linker is ligated to the restriction fragments. This double stranded oligonucleotide linker is specifically designed with a central region of non-complementarity such that amplification can only occur between primers specific for one strand of the linker and the pYAC4 vector. Amplification cannot occur after the first round of DNA synthesis between the linkers ligated on either side of a random yeast or YAC DNA fragment. This method is powerful because of the inherent selectivity for the insert-vector junction fragments and the number of restriction enzymes that can be used (see Chapter 7).

Perhaps the simplest method, however, is the amplification of the vector-insert junction sequences by PCR using primers directed toward repetitive elements and specific sequences in the vector ('*alu*-vector' PCR; Nelson et a., 1989; Breukel et al., 1990). A product will be generated if a repetitive element in the insert is within the range and correct orientation for PCR (see Chapter 7).

One particular development in YAC-based technology offers considerable promise for studies aimed at understanding the biology of complex or large gene loci. Transfer of YACs to mammalian cells by transfection (Eliceiri et al., 1991), fusion (Huxley et al., 1990) or microinjection (Huxley and Gnirke, 1991), can allow analysis of genes in a biological context, that is, in association with flanking DNA likely to contain appropriate sequences controlling gene expression in a cell-specific fashion. Moreover, the ability to transfer such large pieces of DNA suggests a method in which 'loss of function' or recessive genes can be identified by complementation in cell or animal systems. Tumour suppressor genes may well be amenable to such approaches in animal systems. Only a limited number of YAC-containing cell lines would have to be assessed for tumour formation in nude mice in order to identify the gene where a contig already exists in the region of interest.

1.10.6. Identification of genes in cloned DNA.

A number of strategies exist for the identification of genes in a sub-defined region. The cloning of a number of major genes has been facilitated by the observation of cross-species homology, where the assumption is that genes of physiological importance will be conserved across species (Monaco et al., 1986; Rommens et al., 1990; Fearon et al., 1990). The success of this

method, which has generally been shown to hold true may, however, be limited if the location of the gene has not been sufficiently narrowed down. The observation that many genes are associated with undermethylated CpG islands (reviewed in Bird, 1987) has provided a methodology that facilitates the 'scanning' of very large regions of DNA. Moreover, probes that recognise restriction sites clustered in these regions provide a direct route to the identification of such genes (e.g. the Wilm's tumour locus; Gessler et al., (1990)). The approach is, however, limited by the fact that only a proportion of human genes will be associated with these undermethylated islands. By either of these methods, the probes of interest are then screened on cDNA libraries, the choice of tissue from which they are derived based largely on suspicions from the phenotype. A difficulty can arise here, however, if the gene of interest is expressed at very low levels. However, sensitive PCR-based approaches have been devised to counter this problem (e.g. Fearon et al., 1990).

More recently, function-based approaches have been developed to 'trap' genes by forced splicing of candidate exons in cloned DNA to known exons in a vector system, but appear to be at present limited to cosmid-sized inserts (Duyk et al., 1990; Buckler et al., 1990). Due to the large cloning capacity in YACs and the capability to clone extremely large genomic regions of interest, a major challenge must be to devise strategies for the direct selection of genes from this cloned DNA. Screening YAC inserts directly on cDNA libraries is one approach (Wallace et al., 1990; Elvin et al., 1990), but is generally difficult (see Collins, 1992). Other approaches, such as direct selection of cDNAs or conserved genomic DNA on immobilised YACs (Lovett et al., 1991; Parimoo et al., (1991), offers some promise to this eventual goal. Nonetheless, the large amount of insert DNA that can be cloned in YACs still provides a significant amount of DNA with which to pursue more standard approaches, such as 'scanning' for conserved DNA with YAC sub-clones or cosmids derived by direct YAC screening (Baxendale et al., 1991) or the identification of CpG islands (Wilkes et al., 1989).

1.11. Aims of this thesis.

The reports of Bodmer et al., (1987), Solomon et al., (1987) and others described in Chapter 1 (part II) suggested that mutation of the *APC* gene is a critical step in colorectal neoplastic progression, both in rare families with multiple colonic polyposis, but also in patients with 'sporadic' colorectal carcinoma.

This thesis was concerned with attempts to isolate the *APC* gene by a positional cloning strategy on the basis of the chromosomal localisation to 5q21-q22. At the outset of these studies, few available DNA markers had been mapped within this chromosomal region. Thus, this project had two principal aims:

(1) The investigation and assessment of methods that would allow the derivation of a large number of DNA markers specifically mapping into this region. This would enable the construction of a detailed physical framework map to define the *APC* gene-region. To expedite this process, a series of somatic cell hybrids had been independently constructed from FAP patients that segregated both normal and abnormal chromosomes 5 with cytogenetically visible deletions around this region (Varesco et al., 1989; Thomas, 1991). Thus, methods were considered, such as physical chromosomal microdissection, that would allow the enrichment of markers within the boundaries defined by the deletions.

(2) The second aim of this project was to use these DNA markers in the most appropriate ways to sub-localise the position of the *APC* gene. It was anticipated that these might include searching for polymorphisms for further genetic definition of the region, in addition to studying, perhaps, the patterns of allele-loss around the *APC* gene. In particular, it was thought that a high density of markers in this region could facilitate the identification of smaller, sub-microscopic germline chromosomal aberrations, such as deletions in FAP patients, or homozygous deletions in tumour-derived cell lines, that would considerably narrow down the position of the gene.

Chapter 2.

Materials and Methods.

Chapter 2.

Materials and Methods.

2.1. Materials.

2.1.1. Enzymes.

Restriction endonucleases were supplied by New England Biolabs and Boehringer Mannheim GmbH. (various concentrations - 4 to 400 units (U)/μl). T4 DNA ligase was supplied by New England Biolabs (400U/μl). Calf Intestinal Phosphatase (ammonium precipitate at 1U/μl) and DNA Polymerase (Klenow fragment; 1-9U/μl) were purchased from Boehringer Mannheim GmbH. Amplitaq™ DNA Polymerase (from *Thermus aquatus* was supplied by Perkin-Elmer Cetus (5U/μl). *Taq* polymerase was also supplied by Promega (5U/μl). Bovine pancreatic ribonuclease type A (RNAse A), Proteinase K (fungal) and lyticase (from *Arthrobacter luteus*) were supplied by the Sigma Chemical Company. Dessicated Proteinase K was made up to a final concentration of 50 to 100mg/ml in sterile H_20. Dessicated lyticase was prepared according to Burgers and Percival (1987) at a final concentration of 50U/μl and stored at 4^0C. DNAse-free RNAase was prepared from RNAse A by boiling a 10mg/ml solution (in distilled H_20) for 15 minutes, and then stored frozen at -20^0C. Dessicated agarase was purchased from BioLabs and prepared by the addition of sterile glycerol to give a final concentration of 50U/μl. Agarase was stored at 4^0C.

2.1.2. Chemicals.

Chemicals supplied by the manufacturer, unless already in solution, were prepared according to the supplier's recommendations. Unless otherwise stated, chemicals were dissolved in sterile distilled H_20. The final storage concentrations are indicated, where appropriate.

Deoxycytidine triphosphate ([α^{32}P]-dCTP) was supplied by Amersham International, Plc., at a specific activity of 10μCi/μl in aqueous solution. Ampicillin (sodium salt; 100mg/ml), kanamycin sulphate (50mg/ml), polyethylene glycol (PEG)-8000, sarkosyl (*N*-lauroylsarcosine; sodium salt), phenylmethylsulfonylflouride (PMSF), bovine serum albumin (BSA; molecular biology pure; DNAse-, RNAse- and proteinase-free; 10mg/ml),

ribose adenosine-triphosphate (sodium salt; 10mM), deoxyribose adenosine-, cytidine-, thymidine-, guanosine-triphosphates (sodium salts; 10mM), phenol (molecular biology grade, melted and equilibrated with T.E., pH8.0), chloroform, Isopropyl β-D-thio-galactopyranosidase (IPTG), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranosidase (X-gal), Trizma base, Trizma-chloride, Ficoll, spermidine (100mM), polyvinylpyrolidone and ethidium bromide (10mg/ml) were all purchased from the Sigma Chemical Company. Trinitroacetic acid (100mM) was supplied by Merck. GTG molecular biology grade *ultra*-PURE™ standard, low melting point and *Nu-sieve* ™ agarose were purchased from Bethesda Research Laboratories. β-mercaptoethanol was supplied by the Eastman Kodak Company. Bromophenol blue was obtained from BDH chemicals. Absolute ethanol was supplied by Fisons, and isopropanol (propan-2-ol) was supplied by Hayman Ltd. Sodium dodecyl sulphate (SDS) was purchased from SERVA, Feinbiochemicha GmbH & Co. Sephadex (G-50) was supplied by Pharmacia.

2.1.3. Nucleic acids

Human DNA samples were a gift from Dr. Nigel Spurr and prepared at the ICRF Clare Hall Laboratories. Plasmids pBR322 and pBluescript SK™ were supplied by New England Biolabs and Stratagene, respectively. Plasmid pYAC4 was a gift from Dr. Ketan Patel (Human Cytogenetics Laboratory, ICRF). HaeIII digestions of bacteriophage ΦX174 and HindIII digests of bacteriophage λ DNA were purchased from New England Biolabs. The 1kb ladder was purchased from Gibco-BRL. The 100bp ladder, multimers of bacteriophage λ DNA and yeast DNA PFGE markers were supplied by Pharmacia LKB Biotechnology AB. Sheared and sonicated human placental DNA was a gift from Ms. S. Cottrell (Cancer Genetics Laboratory, ICRF) at 4mg/ml, or purchased from the Sigma Chemical Company (10mg/ml). Dessicated salmon sperm DNA (deoxyribonucleic acid, sodium salt, type III from salmon testes) was also supplied by the Sigma Chemical Company and prepared by boiling in distilled sterile H_20 at a final concentration of 10mg/ml. The sources of nucleic acid probes (apart from those isolated, and therefore described in this thesis) are detailed in appendix 2 to this chapter.

2.1.4. Solutions and buffers.

Enzyme buffers.

The following buffers were prepared from autoclaved stock solutions and stored as 10 x concentrated solutions at -20^{0}C. Typically, for restriction endonucleases three buffers were used on the basis of salt concentration tolerance. This was calculated from information supplied by New England Biolabs (1990). For other restriction endonucleases requiring specialised components, buffers were prepared according to the suppliers recommendations or were purchased from the suppliers. RNAse A, proteinase K, lyticase and agarase were used in appropriate solutions detailed in the methods section. Phosphatase (Calf intestine) was typically used in medium salt buffer. The buffer components are written as 1 x concentrated.

High salt buffer:	10mM Tris-HCl (pH7.5), 10mM MgCl, 100mM NaCl, 1mM dithiothreitol (DTT).
Medium salt buffer:	10mM Tris-HCl (pH7.5), 10mM MgCl, 50mM NaCl, 1mM DTT.
Low salt buffer:	10mM Tris-HCl (pH7.5), 10mM MgCl, 1mM DTT.
Ligation buffer:	50mM Tris-HCl (pH8.0), 10mM MgCl, 20mM dithiothreitol, 1mM ATP, 50μg/ml BSA.
PCR buffer:	50mM KCl, 10mM Tris-HCl (pH 8.3), 1.5-3mM MgCl, 0.01% gelatin.

For less specialised PCR procedures (that is, for the preparation of probes), a PCR buffer supplied by Promega biochemicals was used in conjunction with a *Taq* polymerase purchased from the same distributor.

General Solutions and Buffers.

Buffer A: (mammalian cell lysis buffer)	10mM Tris-HCl (pH8.0), 10mM EDTA (pH8.0), 10mM NaCl, 0.5% SDS.

Chloroform/Iso-Amyl-Alcohol:	94 parts chloroform, 4 parts iso-amyl alcohol, stored at room temperature.
Denaturation buffer:	1.5M NaCl, 0.5M NaOH. This was usually made up as a 10L stock and stored at room temperature.
Denhardt's Solution:	2% w/v BSA, 2% w/v Ficoll, 2% w/v polyvinylpyrolidone.
ET:	50mM EDTA, pH8.0, 10mM Tris-HCl, pH8.0. Sterilised by autoclaving and stored at room temperature.
Ficoll Loading Solution:	15% w/v Ficoll 400, 0.25% w/v bromophenol blue. Sterilised by autoclaving and stored at room temperature.
Hybridisation Buffers:	
(1) Pre-hybridisation buffer: (Amersham Int., Plc.)	6 x SCC, 0.25% SDS, 5 x Denhardt's Solution, 25μg/ml Salmon sperm DNA.
(2) Hybridisation solution: (Amersham, Int., Plc.)	as for pre-hybridisation buffer with the addition of 10% Dextran Sulphate.
(3) Church and Gilbert: hybridisation buffer (modified from Church and Glibert, 1984)	0.5M Sodium Phosphate (pH7.2), 7% SDS, 1mM EDTA.
Neutralisation buffer:	1.5M NaCl, 10mM Tris-base, 400mM Tris-HCl. This was usually

made up as a 10L stock and stored at room temperature.

<u>Phenol/Chloroform:</u>	500g Phenol (molecular biology grade), dissolved in 500ml chloroform. 20mls of isoamyl alchohol and 0.5g 8-hydroxyquinone were added and 500mls of 10mM Tris-HCl (pH8.0) layered on top. Stored at room temperature in a darkened bottle.
<u>Proteinase-K-sarkosyl buffer:</u>	1% (w/v) Sodium Lauroyl Sarcosinate (SLS), 0.5M EDTA, 2mg/ml Proteinase K.
<u>SCE:</u>	1M Sorbitol, 100mM sodium citrate (pH5.8), 10mM EDTA.
<u>SCEB-lyticase:</u>	SCE supplemented with 14mM β-mercaptoethanol and lyticase (4U/μl final concentration).
<u>Solution I.</u> (Bacterial alkaline lysis):	50mM Glucose, 25mM Tris-HCl (pH8.0), 10mM EDTA (pH8.0) Autoclaved and stored at 4^0C.
<u>Solution II:</u> (Bacterial alkaline lysis):	0.2M NaOH, 1% SDS. Prepared fresh each time.
<u>Solution III:</u> (Bacterial alkaline lysis):	60mls 5M potassium acetate, 11.5mls glacial acetic acid, 28.5mls sterile distilled H_20. Stored at room temperature.
<u>SSC:</u>	150mM NaCl, 15mM sodium citrate, pH7.0 (adjusted with NaOH). This was made up as a concentrated solution (20 x SSC),

sterilised by autoclaving and stored at room temperature.

STE: 100mM NaCl, 10mM Tris-HCl, pH8.0, 0.1M EDTA. Sterilised by autoclaving and stored at room temperature.

TAE Electrophoresis Buffer: 40mM Tris-acetate, pH7.2, 40mM sodium acetate, 1mM EDTA. This was made up as a concentrated solution (40 x TAE) and stored at room temperature.

TBE Electrophoresis Buffer: 90mM Tris-HCl, pH7.2, 90mM Boric acid, 2.5mM EDTA, pH8.3. This was made up as a concentrated solution (5 x TBE) and stored at room temperature.

TE: 10mM Tris-HCl, pH8.0, 0.1M EDTA, pH8.0. Sterilised by autoclaving and stored at room temperature

2.1.5. Media

Difco supplied the following media components; Bacto-tryptone (casein, enzymatic hydrosylate), Bacto-yeast extract, Bacto-agar, Bacto-peptone, yeast nitrogen base (without amino acids), casaminoacids. Individual amino acids; tryptophan and tyrosine, in addition to adenine base, were supplied by the Sigma chemical Company. Glucose (20% w/v), glycerol (80% v/v) were supplied by ICRF central services.

Bacterial media and maintainnance of clones.

LB-broth: 10g Bacto-tryptone, 5g yeast extract, 5g NaCl, in 1 litre H_20. pH was adjusted to 7.2, autoclaved and

stored at room temperature until use.

LB-agar:	LB-broth with 10g Bacto-agar. Autoclaved and stored at room temperatre until use.
YT medium:	8g Bacto-tryptone (casein hydrosylate), 5g yeast extract, 5g NaCl, in 1 litre H_20. Autoclaved and stored at room temperature until use.
Stab agar:	10g Bacto-tryptone, 7g Bacto-agar, in 1 litre H_20. Heated to dissolve, autoclaved and aliquoted to small Nunc™ vials.

For antibiotic supplementation, ampicillin or kanamycin (both prepared in sterile H_20, filter sterilised and stored at -20^{0}C) were added at 100µg/ml and 50µg/ml, respectively, just prior to initiation of bacterial growth. For the detection of α-complementation, L-agar (Miller, 1987) was supplemented with 160µg/ml IPTG (prepared in distilled H_20) and 400µg/ml X-Gal solution (prepared in dimethylformamide and stored in a light-sealed tube) prior to casting of plates.

Bacteria containing plasmids or cosmids were stored (1) in Nunc™ freezing tubes containing 30% (v/v) glycerol-LB-broth cultures at -70^{0}C and (2) as anaerobic stab cultures in Nunc™ tubes (stab agar) in the dark at room temperature. These stabs were subcultured at approximately 6 month intervals.

Yeast media and maintainance of clones.

YPD medium:	10g Bacto-yeast extract, 20g Bacto-peptone, 20g glucose, in 1 litre H_20. Autoclaved and stored at room temperature until used.

YPD-agar:	YPD medium with 20g Bacto-agar. Autoclaved and stored at room temperature until used.
SD-Caa- medium:	6.7g Bacto-yeast nitrogen base (w/o amino-acids), 20g glucose, 14g casaminoacids plus appropriate amino acid supplements (see below), in 1 litre H_20. Autoclaved and stored at room temperature until used.
SD-Caa-agar:	SD-Caa- medium with 20g Bacto-agar. Autoclaved and stored at room temperature until used.
Amino acid supplements:	(a) L- tyrosine: 2mg/ml stock solution. Autoclaved and stored at room temperature, added to SD at a final concentration of 30μg/L.
	(b) L-tryptophan: 10mg/ml stock solution. Filter sterilised and stored at 4^0C in the dark, added to SD at a final concentration of 20mg/L.
Base supplements:	adenine sulphate: 2mg/ml stock solution. Autoclaved and stored at room temperature, added to SD at a final concentration of 20μg/L.

Yeast strains were processed in two ways for long term and short term storage. For indefinite storage, 1ml aliquots of yeast cultures prepared for DNA isolation were decanted into Nunc™ freezer tubes and the solution made up to a final concentration of 30% autoclaved glycerol. These tubes were stored at -70^0C. For short term storage with easy access, half yeast colonies were streaked on appropriately supplemented SD-Caa-agar slopes, grown at 30^0C for 3 days and sealed with parafilm. These slopes were kept at 4^0C and subcultured at approximately 3 month intervals.

2.2. General DNA methods.

2.2.1. Preparation of plasmid DNA.

Large scale preparations.

Large scale preparations were based on a method described by Miller (1987) which is similar to the alkaline-lysis method described by Maniatis et al., (1982). 500ml cultures of plasmid-bearing bacteria were grown to saturation overnight either in LB-broth (plasmids such as pBluescript) or YT media (cosmids) supplemented with the appropriate antibiotic(s). The cells were pelleted at 3000rpm in four 250ml centrifuge flasks for 15 minutes, resuspended in 20mls GTE buffer (Solution I) and left at room temperature for 5 minutes. 10mls of 0.2M NaOH-1% SDS (Solution II; freshly prepared) was then added and mixed gently until the walls of the flask were coated with the solution. This was then incubated on ice for 10 minutes. 7.5mls of a Sodium Acetate solution (Solution III) was then added and the mix put on ice for 15 minutes. The tubes were centrifuged for 15 minutes at 8000rpm and the supernatant collected through a small tea-strainer. The tubes were re-centrifuged and the remaining supernatant collected as before. To this supernatant an equal volume of isopropanol was added and the tubes put on ice for 60 minutes. The tubes were then spun at 8000g for 20 minutes and the supernatant poured off. The tubes were inverted and allowed to drain for a few minutes. The pellet was resuspended in 5mls of T.E. in falcon tubes and the solution extracted once with phenol and then 3-4 times with phenol-chloroform (until the interphase was clear) and finally extracted once with chloroform. The DNA solution was then precipitated with 2.1 volumes of ethanol with a final concentration of 0.3M Sodium Acetate. The solution was resuspended in 1ml of T.E. supplemented with RNAse at a final concentration of 1µg/ml and incubated at room temperature for 10 minutes. The solution was extracted once with phenol-chloroform, once with chloroform and then re-precipitated with ethanol. Finally, the pellet was washed with 70% ethanol, dried under vacuum and resuspended in 0.5-1ml T.E. The concentration of DNA in solution was calculated from optical density readings at 260nm. The relative purity of the DNA was calculated from a ratio of the optical densities at 260 and 280mn (i.e. 260nm/280nm = x, where x is usually >1.7 for DNA solutions containing little contaminating protein)

Small scale preparations of plasmid DNA.

Small scale plasmid DNA preparations were generally carried out by a scaled-down version of the alkaline lysis procedure (Miller, 1987). 2ml cultures were grown overnight in LB (plasmids) or YT (cosmids) and 1ml of the cells centrifuged at 13,000rpm in 1.5ml microfuge tubes (eppendorf). To eliminate extruded nucleases or enzyme inhibitors, the cells were resuspended in 100µl of STE buffer and re-centrifuged as before. 100µl of Solution I was added and the tubes put on ice for 5 minutes. 200µl of Solution II was then added and the tubes kept at room temperature for 5 minutes. Subsequently, 150µl of Solution III was then added and put on ice for 10 minutes. The solution was spun at 13,000rpm in a microfuge for 15 minutes at 4^0C and the supernatant collected in a new tube. The tube was filled up with cold ethanol and put on ice for 10 minutes. The DNA was pelleted for 1 minute at 13,000rpm in a microfuge and the supernatant aspirated. The pellet was washed with 0.5ml 70% ethanol and the pellet dried under vacuum. The pellet was resuspended in 50µl of T.E. and RNAse added to a final concentration of 10µg/ml. The pellet could be re-precipitated with ethanol if difficulty was encountered with restriction endonuclease digestion.

2.2.2. Extraction of genomic DNA.

Total genomic DNA, prepared from a number of 'normal' individuals, was kindly supplied by Dr. N. Spurr (Human Genetics Resources Laboratory, ICRF).

DNA from somatic cell hybrids (see appendix 3) was prepared from cultured cells, the seeding and maintainance of which was carried out exactly as described by Varesco et al., (1989). Cells were grown to approximately 70% confluence on 10cm culture dishes and washed two times with 10-20mls of PBSA. Usually, a total of about 5×10^7 to 1×10^8 cells were collected by a 'rubber-policeman' and deposited in 20mls PBSA in a 50ml Falcon tube. The cell suspension was centrifuged at 1000rpm for 10 minutes at room temperature. The cells were washed again in 20mls of PBSA and centrifuged as above. Supernatants were carefully drained and the pellets carefully resuspended in 40mls of buffer A for every 10^8 cells. Proteinase K (100µg/ml final concentration) was added and the solution incubated at 37^0C for 12

hours. Subsequently, RNAse A was added to a final concentration of 50μg/ml and the solution incubated at 37^{0}C for a further two hours.

An equal volume of equilibrated phenol was added and the solution mixed with gentle rocking for 1 hour at room temperature. The phenol was removed from the aqueous phase by suction using a wide-bore 10ml pipette. This was repeated 2 to 3 times until the interphase was clear. The solution was then extracted once with an equal volume of phenol/chloroform and once with an equal volume of chloroform. DNA was precipitated from the aqueous phase by the addition of two volumes of absolute ethanol and a 1/10 volume of 3M sodium acetate. The DNA was recovered by spooling on a bent pasteur pipette, washed with 70% ethanol and resuspended in 0.5 to 1ml of T.E., pH8.0. The concentration of DNA in solution was calculated from optical density readings at 260nm. The relative purity of the DNA was calculated from a ratio of the optical densities at 260 and 280mn (i.e. 260nm/280nm = x, where x is usually >1.7 for DNA solutions containing little contaminating protein).

2.2.3. Restriction endonuclease digestions

Endonuclease digestions were carried out using the appropriate buffers and at the temperatures recommended by the suppliers (see section 2.1.3). Typically, for genomic DNA, 5U of enzyme was used per μg of DNA; where one unit is defined as the amount of enzyme required to digest 1μg of wild-type λ DNA in 1 hour. For cloned DNA, an approximately two-fold excess of enzyme was used. Digestion reactions were generally incubated for 2-3 hours for cloned DNA and 6 to 12 hours for genomic DNA.

For multiple digestions, the DNA samples were incubated with one or more enzymes if the buffers were compatible in terms of salt concentrations. If this was not the case, digestions were carried out sequentially, either by increasing the salt concentrations for the second enzyme, or by precipitating the DNA with ethanol in between the two digestion reactions. In general, enzymes were inactivated either by heat (65^{0}C for 15 minutes) or by phenol/chloroform extractions and ethanol precipitation (Maniatis et al., 1982). Digested DNA fragments were resolved on the appropriate percentage agarose gel (see section 2.2.4). If incomplete digestion was observed, the DNA samples were precipitated with ethanol and re-digested with the appropriate enzyme.

2.2.4. Agarose gel electrophoresis

Electrophoresis gels were prepared by dissolving an appropriate amount of agarose in TAE or TBE buffer to give final concentrations of between 0.3-3%, depending on the size of DNA fragments to be resolved. Gels were cast in 10 x 8cm mini-gel trays (Scotlab) with varying numbers of gel-slot formers (8, 12, 16) or in 14 x 10cm midi-gel trays (Scotlab) with 14-well gel-slot formers, in addition to 24 x 20cm standard gel trays (Scotlab) with 20- or 36-well gel-slot formers. DNA samples were mixed with one-tenth volumes of loading solution and applied to the well-slots submerged under running buffer (TAE or TBE). Voltage gradients of between 0.5 to $10Vcm^{-1}$ were applied for 1 to 48 hours. Nucleic acids were visualised by ethidium bromide staining (0.5μg/ml) and ultraviolet trans-illumination.

2.2.5. Southern blotting.

Agarose gels containing resolved DNA fragments were transferred by capillary action to nylon membranes (Hybond N or Hybond N^+; Amersham International) according to the suppliers protocol. Most commonly the agarose gels were submerged for 40 minutes each in denaturation and neutralisation buffers and subsequently transferred to Hybond N^+ by capillary action. The membranes were then placed on 2 sheets of 3MM paper soaked with 0.4M Sodium Hydroxide (NaOH) to fix the DNA. The filters were then extensively neutralised by 4 x 15 minute washes in 2 x SCC and dried in air on 3MM paper (Whatman).

2.2.6. Radiolabelling of DNA probes.

Probes to be labelled were prepared either by digestion from plasmid DNA preparations or by PCR (see section 2.3.3 and 2.5). Relevant DNA fragments were electrophoresed in low melting point gels and excised. The intensities of the excised fragments were compared by UV trans-illumination directly to known amounts of undigested λ DNA resolved on the same gel. After estimation of DNA concentration, the gel fragments were diluted to 1-2ng/μl in distilled H_20.

Radiolabelling of probes was carried out by the random priming method (Feinberg and Vogelstein, 1984). Probes were denatured either by boiling in a water-bath or by incubation at 99^0C in a PCR thermocycler (Hybaid or Perkin-

Elmer) for 10 minutes. Between 20-50μCi of [α^{32}P]-dCTP was used and the reaction initiated with 1-2U of Klenow fragment DNA polymerase I. The reactions were allowed to proceed for 3 hours at 37^{0}C or for 16 hours at room temperature. Unincorporated nucleotides were removed by centrifugation of the reaction mix through a G-50 Sephadex column equilibrated in STE at 1000rpm for 3 minutes (Maniatis et al., 1982). Nucleotide incorporation was calculated by Cerenkov counting of samples before and after column centrifugation. Typically, 50 to 80% incorporation was observed. Specific activities were calculated from counts per minute (cpm) of an estimated amount of DNA. In most cases specific activities were, apparently, in the range of 1-5 x 10^9 per μg of DNA.

2.2.7. Hybridisation.

Hybridisation of nylon membranes was carried out at 65^{0}C either in sealed plastic bags or more commonly in rotary bottles (Hybaid). Denatured probe was incubated after 1-2 hours pre-hybridisation (buffers 1 or 3) in a volume not exceeding 10mls of hybridisation buffer (buffers 2 or 3) for a 22 x 22cm filter. In general, the entire labelled probe was incubated. However, in cases where probe signal-to-noise ratios were known to be high, the probes were added at a final concentration of 1 x 10^6 counts per ml of hybridisation buffer.

Non-specifically bound probe was removed by successive washing of the filters in 0.1% SDS at temperatures of between 30^{0}C and 65^{0}C. SDS was supplemented with varying amounts of SSC, depending on the relative stringency required. For small probes (≈80bp to ≈500bp), the filters were washed 3-4 times for 30 minutes each in a final concentration of 2 x SSC. For other larger probes or for probes containing repetitive sequences, filters were typically washed in final concentrations of 0.1-0.2 x SSC. Filters were exposed to XAR 5 autoradiographic film (Kodak) for 1 hour to 7 days with or without intensifying screens at -70^{0}C or at room temperature.

2.2.8. Competitive hybridisation.

For probes containing repetitive sequences, pre-annealing of the denatured DNA with sonicated and sheared human placental DNA was carried out. Competitor DNA was added to a final concentration of 0.4-1mg/ml in 0.12M Sodium Phosphate (pH 7.2) and incubated for an appropriate time at 65^{0}C to achieve an effective Cot value of 50-100 (Sealy et al., 1985). On occasion,

competitor DNA was added during pre-hybridisation (at between 50-200μg/ml final concentration; Litt and White, (1985); Nelson et al., (1989)). This was necessary in the case of whole cosmid hybridisations and some inter-IRS PCR products (see sections 2.3 and 2.4).

2.2.9. Removal of radioactive probes.

For subsequent hybridisations on the same filter, membranes were immersed in a solution of boiling 0.5% SDS and allowed to cool to room temperature with gently rocking. This was carried out two times. The membranes were then washed 3 times in 2 x SSC for 5 minutes each with gentle rocking and air dried on 3MM paper. Yeast artificial chromosome filters (see below) were 'stripped' by incubation in a solution containing 0.1% SDS, 0.1 x SSC and 1mM EDTA for 1 hour at 65^0C to avoid any removal of bound DNA from the filters. Filters were usually exposed to autoradiographic film after removal of probe to detect the presence of any residual signal.

2.2.10. Polymerase chain reaction (PCR).

Oligonucleotide synthesis.

Oligonucleotide primers were synthesised at the Human Genetic Resources Laboratory (ICRF, Clare Hall Laboratories). Oligonucleotides were supplied as ethanol precipitates. DNA pellets were dried in air, dissolved in re-autoclaved distilled H_20 and permanently stored as 200μM stocks at -20^0C. For working solutions, these were diluted to 20μM. Approximate melting temperatures were estimated by G+C content calculations ($T_m = 4(G+C) + 2(A+T)$); or from published reports. In general, primers were assayed at varying temperatures (typically 50^0C to 65^0C) to assess maximum specificity. The sequence and annealing temperatures of each of the primers used in this thesis are summarised in appendix 1 to this chapter.

Polymerase chain reactions (PCR).

Polymerase chain reactions (PCR) were generally set up as follows. PCR pre-mixes (containing all of the components except DNA polymerase and DNA templates) were assembled (for batches of 10 to 50 reactions) and each reaction was carried out in a final volume of between 50 to 100μl in a 0.5ml microfuge tube. These contained 200μM final concentration of each deoxy-nucleotide-

triphosphate (dATP, dTTP, dCTP, dGTP; taken from 10mM stocks), a 1/10 volume of 10 x PCR buffer (see section 2.1) and the appropriate concentration of one or more oligonucleotide primers (detailed in the appropriate sections of this chapter). Optimal Mg^{2+} concentrations were determined empirically (between 1.5 and 3mM MgCl final concentration). After the addition of template DNA (see relevant sections), the reaction mix was covered by a 20-100μl layer of mineral oil (light, Sigma).

Usually, the PCR reactions were denatured for 5 to 10 minutes in a thermocycler (Hybaid or Perkin-Elmer) before the addition of 0.5-2.5U of *Taq* polymerase (previously diluted in 1 x PCR buffer). Reactions were cycled 28-40 times with denaturation at 94^{0}C for 0.5 to 1 minute, annealing at the appropriate temperature for 1 to 2 minutes and extension at 72^{0}C for 1 to 4 minutes. A final extension was usually performed at 72^{0}C for 10 minutes. Details of each of the cycling reactions are presented in the relevant sections.

2.3. *Alu*-PCR methods.

2.3.1. *Alu*-PCR.

Each of the *alu*-based primers (see appendix 1 to this chapter) were used as follows: 559 (TC-65), (Nelson et al., 1989), IV (Cotter et al., 1990), LIHs (Ledbetter et al., 1990) and the PDJ series (i.e., 33, 34 and 66; P. de Jong, personal communication, see also Butler et al., 1992) were used at a final concentration of 1μM. Primer 517 (Nelson et al., 1989) was used at a final concentration of 0.2-0.5μM. These primers were used at the same relative concentrations when in combination with each other in the same PCR reaction. The optimal Mg^{2+} concentration was found to be 1.5mM for all of the primers except the PDJ series, where 3mM was found to be ideal (see also Butler et al., 1992). One μg of template DNA (either total human genomic or somatic cell cell hybrid) was added to the reaction. For primers IV, 517, 559 and LIHs, 35-40 cycles were found to be ideal, but, for the PDJ series, optimal results were obtained after 28 to 30 cycles. Generally, good results were obtained by using 2.5U of *Taq* polymerase, but, 1 to 1.5U was usually sufficient for the PDJ series.

2.3.2. Cloning of inter-*alu* PCR products: PN/Ts-1 library.

80μl of the preparative PCR reactions using primer IV on the PN/Ts-1 hybrid were extracted with an equal volume of phenol/chloroform and centrifuged

in a G-50 sephadex column (Maniatis et al., 1982) previously equilibrated in T.E. The eluate was then precipitated with ethanol and the DNA pellet resuspended in 10-20μl of T.E. (final concentration of ≈100-200ng/μl). pBluescript™ plasmid (at 1mg/ml) was digested with EcoRV for 3 hours at 37^{0}C. The digested DNA was electrophoresed and excised from a low melting point agarose gel. The plasmid was purified from the gel slice by Gene-clean (Bio101) and diluted to about 2.5ng/μl in T.E. A constant amount of PCR product (usually 100ng) was ligated to varying amounts of the digested pBluescript™ plasmid (that is, 0.25, 2.5, 5 and 10ng) to assess the final ratio of non-recombinant to recombinant colonies. The ligations were carried out in a final volume of 20μl containing a 1/10 volume of commercial 10 x ligation buffer, 5% polyethylene glycol (PEG-8000) and 200U of T4 DNA ligase for 12-14 hours at 16^{0}C.

2.3.3. Analysis of inter-*alu* product libraries.

Aliquots of the ligation mixes (usually 1-2μl) were transformed to competent DH5α *E. coli*. Competent cells for transformation were prepared according to Hanahan (1985) or were purchased from Stratagene. Transformed *E. coli* cultures (0.5-1ml) were spread onto 22 x 22cm LB-agar plates. Analysis of individual recombinant (white) colonies was carried out by direct bacterial cell PCR. Briefly, half of one colony was lifted with a sterile toothpick and dipped into a tube containing 50 to 100μl of PCR-premix containing *alu*-based primer IV. PCR was carried out as described for genomic DNA, except that before analysis of the products, the tubes were centrifuged at 6500rpm in a microfuge for 5 minutes to pellet the bacterial debris. PCR products used for hybridisation analyses were electrophoresed in low melting point agarose gels, excised and labelled as in section 2.2.6.

Identification of colonies containing highly repetitive plasmid insert sequences was carried out by filter colony hybridisation with radiolabelled total human genomic DNA. *In situ* filter lifts were performed as described by Buluwela et al., (1989). Individual recombinant colonies (white) were gridded out in ordered arrays on selective LB-agar plates and grown overnight at 37^{0}C. 8.2cm-diameter filters were placed on the colonies for 1 to 2 minutes to allow bacterial cell adherence. The filters were then placed, colony side up, on 3MM paper soaked in 5 x SSC, 0.1% SDS for 2 minutes in a heat-resistant plastic dish. The dish was transferred to a 650W microwave oven (with rotary base)

and baked for 2.5 minutes at full setting. The filters could then be used, directly, in hybridisation assays.

2.4. Identification and isolation of cosmids.

2.4.1. Gridded cosmid library screening.

Cosmids were identified from a chromosome 5-specific library prepared from flow-sorted chromosomes in the sCos-1 vector (Evans and Wahl, 1987). This library was constructed by Dr. Larry Devan and colleagues at Los Alamos National Laboratory, U.S.A. Cosmid-containing bacterial colonies were spotted in high density gridded arrays by robotic application (≈ 9000 clones on one 22 x 22cm nylon filter; Dr. Anna-Maria Frischauf, Analysis of Mammalian Mutation Laboratory, ICRF). This library represented approximately 2 genome equivalents of chromosome 5. Filters were screened by standard radiolabelled probe hybridisation (see section 2.2.7). To determine the grid co-ordinates of the positive signals, the filters were hybridised with radiolabelled pWE15 cosmid DNA. The position of colony-specific signals were calculated by superimposition of both autoradiographs according to predetermined markings on the filters. Bacterial cells from the positive grid coordinates were isolated from the appropriate microtitre plate wells, streaked on LB-agar supplemented with 50µg/ml Kanamycin and grown at 37^0C overnight. Secondary screening was carried out by colony filter hybridisation using the protocol of Buluwela et al., (1989). DNA was prepared by the alkaline-lysis procedure exactly as described in section 2.2.1.

2.4.2. Isolation of unique cosmid sub-fragments.

Cosmid DNA (section 2.2.1) was digested with restriction enzymes (see Chapter 3, part II) and the DNA fragments resolved in agarose gels. Southern blots of the fragments were probed with radiolabelled total human genomic DNA. Fragments visualised by UV transillumination, but exhibiting no signal for the genomic probe, were isolated from low melting point gels and processed as for standard probe preparations (section 2.2.6).

2.5. Microclone methods (large scale analysis of library II).

Microclones were received as either 1ml overnight bacterial cultures (in the early stages of analysis (see sections 4.2.1. to 4.2.3., Chapter 4) or as single

colonies gridded on LB-agar plates. Recombinant plasmid-containing bacteria (DH5α) were transferred by sterile toothpicks to 96-well plates (Falcon). Each of these wells contained 100μl of LB-broth supplemented with ampicillin at a final concentration of 100μg/ml. The plates were covered with parafilm and incubated with gentle shaking overnight at 37^0C. For permanent storage, bacterial cells were replicated into further 96-well dishes supplemented with 30% (v/v) sterile glycerol using a 96-pin hedge-hog device (Coulson et al., 1988). These plates were stored at -70^0C. For immediate analysis, 1μl of saturated bacterial culture from each of the independent wells were placed directly into tubes containing the appropriate PCR pre-mix and amplified using the M13 Universal sequencing and reverse primers (see appendix 1 to this chapter). Thirty cycles were performed using 1μM of each primer and 1U of *Taq* polymerase. The amplification products served as the source of DNA for microclone insert probes which were labelled as described in section 2.2.6. Microclone inserts were screened on digested human genomic DNA or on the somatic cell hybrid panels described in Chapter 4. Post hybridisation washes were typically carried out at 2 x SSC/0.1% SDS at 55^0C.

2.6. Yeast Artificial Chromosome (YAC) methods.

2.6.1. YAC library screening.

The YAC library used in these studies is described in Larin et al., (1991). Briefly, this library was prepared by partial EcoRI digestion of human lymphoblast DNA (containing four copies of chromosome X) ligated to pYAC4 vector and propagated in the AB1380 yeast host (Burke et al., 1987). Either yeast DNA or whole yeast colonies (≈ 9000 clones or ≈ 2 genome equivalents) were spotted in high density arrays on 22 x 22cm nylon filters by robotic application exactly as described for the isolation of cosmids (kindly provided by Drs. Mark Ross, Anthony Monaco and Hans Lehrach).

Test probes were typically labelled with 50μCi of [$\alpha^{32}P$]-dCTP. Post hybridisation washing was carried out at between 30^0C to 40^0C for extended periods of time (usually 1-2 hours). For larger sized probes or probes containing a limited amount of repetitive sequence, the temperature was increased to 65^0C for the final wash. Yeast cells from each of the positive signals (converted into grid co-ordinates) were isolated from the appropriate microtitre dish wells and streaked onto non-selective YPD-agar plates which were grown for 2-3 days at 30^0C.

2.6.2. Secondary screening.

Confirmation of positive YAC clones was performed by colony filter hybridisations as follows: Cells were collected from the most dense region of the YPD plates and streaked for single colonies on selective SD-Caa plates lacking uracil. The cells were allowed to grow at 30^{0}C until colonies were 1-2mm in size. Subsequently, 8.2cm diameter nylon filters were placed on these plates for 3 minutes for yeast cell adherence. These filters were placed, colony side up, on fresh selective agar plates and incubated for 3-4 hours to initiate log-phase growth. The nylon filters were placed on 3MM paper soaked in SCE for 3 minutes at room temperature and then on 3MM paper soaked with SCEB supplemented with lyticase (4U/μl) overnight at 37^{0}C to lyse the cell walls. On occasion, the success of cell-wall lysis was checked by observation of yeast cells in 5% SDS under the microscope. Filters were prepared for hybridisation by successive incubation on 3MM paper soaked in denaturation and neutralisation solutions. The DNA was fixed by UV illumination (2 minutes; 306nm). Hybridisations were carried out by standard methods (see section 2.2.7).

2.6.3. Yeast DNA preparations.

Half of one positive colony (determined by secondary screening) was used to initiate DNA isolation procedures; the other other half was used for storage (see section 2.1.5). DNA was isolated in agarose blocks by a modified proteinase K method (Rubock et al., 1990; Ragoussis et al., 1991). Between 10 and 50mls of autoclaved selective growth media (SD-Caa solution lacking uracil) in 250ml conical flasks was inoculated with the appropriate yeast cells and incubated with vigorous shaking and aeration for 1-2 days at 30^{0}C. Yeast cells were collected by centrifugation at 3000rpm in a 50ml falcon tube for 10 minutes at room temperature. The yeast cell pellet was resuspended in 10mls of ET buffer and re-centrifuged as above. The supernatant was discarded and the pellet resuspended in SCEB supplemented with lyticase (4U/μl). Usually, 40μl of SCEB-lyticase was added per ml of starting yeast culture (3-5μg of DNA/ml). This solution was aliquoted into eppendorf tubes (500μl each). An equal amount (v/v) of 1.5% low melting point agarose dissolved in SCEB (held at a constant temperature of 45^{0}C) was added and 80μl aliquots added to block formers over ice. After 15 minutes at 4^{0}C, the blocks were extruded into a solution of SCEB-lyticase (1ml per 80μl block) and incubated for 4 hours at

37^{0}C. The blocks were then added to a buffered proteinase K-sarkosyl solution for 36-48 hours at 50^{0}C with gentle rocking.

For subsequent manipulations, blocks were washed in T.E. two times for 30 minutes each at 50^{0}C and then twice for 30 minutes each at 50^{0}C in T.E. containing 0.04μg/ml phenylmethylsulfonyl-flouride (PMSF) to inactivate the proteinase K. Removal of PMSF was achieved by extensive washing in T.E. at room temperature (usually 3 x 30 minutes). For long term storage, blocks were added to 0.5M EDTA and kept at 4^{0}C.

For the isolation of small amounts of DNA, 1-2 YAC blocks (PMSF-treated) were incubated with an equal volume of T.E. NaCl was then added to a final concentration of 30mM. The blocks were melted at 65^{0}C for 15 minutes and then equilibrated at 37^{0}C for a further 10 minutes. 5U of agarase was added and the incubation carried on for a further 2 hours. Successful agar digestion was assessed by incubation of the solution on ice for 10 minutes. This solution could be used directly in this form for PCR reactions (for example inter-*alu* PCR and *alu*-vector PCR; see sections 2.3.1 and 2.7.1, respectively). However, for restriction enzyme digestion analysis (see section 2.6.4), the solution was purified by phenol/chloroform extraction and ethanol precipitation using 0.01μg/ml of dextran polymer as a DNA carrier.

2.6.4. Restriction digestions of yeast DNA.

Restriction enzyme digestions of yeast DNA were usually carried out on the agarose block preparations. Restriction digests of purified yeast DNA with frequently cutting restriction enzymes were carried out as for cloned or genomic DNAs.

For infrequently cutting restriction enzymes, approximately 1/3 of an 80μl block was used. After extensive washing in T.E., 1/3 block slices were equilibrated in 1ml of the appropriate restriction enzyme buffer at 4^{0}C with gentle rocking. Digests were then carried out in a final volume of 100μl with the appropriate restriction enzyme buffer, supplemented with bovine serum albumin (BSA) and spermidine (0.5mg/ml and 5mM final concentrations, respectively). The tubes were incubated on ice for 30-60 minutes to allow maximal enzyme diffusion into the agarose blocks. Subsequently, the tubes were incubated for 4-6 hours at the appropriate temperature for total digestions. For partial digestion, yeast DNA was either incubated with a

constant amount of enzyme and digested for varying amounts of time, or digested for a constant period of time with varying amounts of restriction enzymes. The times of digestion or the amounts of enzyme are described in detail in the relevant figure legends. After digestion, the blocks were immediately immersed in 0.5M EDTA for 15 minutes on ice and then washed two times in T.E. for 15 minutes each at 4^0C. For digestions with two restriction enzymes, the enzyme requiring the lowest salt concentration was used first. After initial digestion, the blocks were equilibrated in T.E., then equilibrated in the second restriction buffer, and digested with the second enzyme.

2.6.6. Pulsed field gel electrophoresis.

Yeast DNAs digested with infrequently cutting restriction enzymes were typically resolved by pulsed field gel electrophoresis on a Biorad CHEF DR-II apparatus. 1% agarose gels were cast in 0.5 x TBE buffer with appropriate gel combs (14 or 30-well). Yeast DNA blocks were washed in T.E. and placed at the edge of the well formers which were horizontally held. Excess solution was absorbed from the blocks by tissue paper to allow the blocks to remain on the gel combs by surface tension. Subsequently, the gel combs were lowered vertically and the gel cast around the comb. The gels were electrophoresed under 0.5cm running buffer (0.5 x TBE at a constant circulating temperature of 14^0C) at 5V/cm for varying times. The switching times were calculated from a software package supplied by Biorad Ltd. for that purpose. The conditions used for the appropriate resolution of YACs or restriction fragments are described in the relevant figure legends. For an initial analysis to size the YACs, yeast chromosomes were separated with a switching time of 24 to 80 seconds for 20 hours (resolves in the 100kb to 1Mb region).

Sizes of the YACs and/or restriction fragments of the YACs were determined by comparison to concatamers of bacteriophage λ (steps of ≈ 50kb) and yeast DNA PFGE makers (*Saccharomyces cerevisiae*). For the sizing of small restriction fragments, bacteriophage λ, digested with HindIIII, was set in agarose blocks and co-electrophoresed. Pulsed field gels were stained in TBE buffer supplemented with ethidium bromide (at 0.8μg/ml final concentration). Subsequently, the DNA was processed *in situ* and transferred to nylon membranes as described.

2.6.7. Derivation of YAC physical maps.

Southern blots of partially digested YAC DNAs were sequentially hybridised with probes prepared from pBR322 which are specific for each of the pYAC4 vector arms. The right arm (defined as that arm containing the URA3 gene) can be recognised, specifically, with a 1.4kb SalI-PvuII fragment. The left arm (containing the TRP1 gene) can be recognised with a 750bp fragment derived from digestion of pBR322 with PstI and EcoRI. Confirmation of the terminal restriction fragments detected by the pBR322 probes in partially digested yeast DNA was achieved, in each case, by hybridisation to total digests.

For total digests, restriction fragments were also detected by a probe containing both an *alu* and LIHs element ligated in tandem (IE12) which was kindly supplied by Dr. J. Ragoussis (Ragoussis et al., 1991). Localisation of the probes which were used to isolate the YAC was usually carried out by hybridisation to both the partial and total (single and double) digestions at the same time. Physical mapping of a specific YAC from a co-transformed yeast host was achieved by the selective isolation and hybridisation of specific YAC insert termini sequences (see section 2.7).

2.6.8. Comparative yeast and human Southern blot hybridisations.

To assess the fidelity of cloning in the identified YACs, yeast DNA (purified from agarase treated blocks) was digested in parallel to human genomic DNA with a number of different restriction endonucleases. Approximately 50ng of whole yeast DNA and 10µg of human genomic DNA were loaded in parallel on 25cm, 0.8% agarose gels and electrophoresed at 2.5V/cm for 48 hours. Processing of the agarose gels, southern blot transfer and hybridisations were performed as described in section 2.2.

2.7. Isolation of YAC insert termini.

2.7.1. *Alu*-vector PCR.

Alu-vector PCR was carried out exactly as described for genomic inter-*alu* PCR analysis, except that an additional primer based on 'left' or 'right' arm of the pYAC4 vector (Riley et al., 1990; see appendix 1 to this chapter; see also figure 7.4, Chapter 7) was added, initially at a final concentration of 1µM. In addition, 200-fold less template DNA was added in line with the differences in genome size (usually 5-10ng of whole yeast DNA was sufficient). To

enhance the appearance of novel vector-insert junction fragments, the pYAC4 specific primer was added in a 10-fold excess (0.1μM *alu* primer; 1μM pYAC4-specific primer).

Restriction digestion of the PCR fragments was carried out as follows: a 20μl aliquot of PCR product was added to 15μl of water, 4μl of the appropriate restriction enzyme buffer and 1μl of the appropriate restriction enzyme (usually 10U). This mix was then incubated at 37^{0}C for 1 hour. The fragment representing the sequences from the *alu* element to the EcoRI cloning site in pYAC4 was excised from low melting point gels, labelled and hybridised as described in section 2.2.

2.7.2. Vectorette PCR.

Construction and use of the vectorette libraries was carried out essentially as described in Riley et al., (1990). Refer also to Chapter 7 for a schematic illustration of the method (in addition to the sequence of the vectorette cassettes). Three vectorette cassettes were used which were specific for BglII (or BamHI), HindIII and PstI restriction enzymes. Briefly, whole yeast DNA in 1/3 block slices was digested as described in section 2.6.4. Half of each of the digested blocks were examined by conventional gel electrophoresis in 0.8% agarose gels. The digested blocks were then equilibrated in 1ml of ligation buffer for 30 minutes at 4^{0}C. Subsequently, the blocks were incubated in ligation buffer in a final volume of 100μl and 1μg of vectorette cassette added. The tubes were incubated at 65^{0}C for 15 minutes to melt the agarose and then equilibrated at 37^{0}C for the ligation reaction. Ribose-ATP (rATP) was added to a final concentration of 1mM (from a 10mM stock, used once or twice and then discarded) and ligation carried out at 37^{0}C for 2 hours with 10U of T4 DNA ligase.

After ligation, the library was diluted to a final volume of 400μl with distilled H_20 and stored at -20^{0}C. PCR was carried out using 5μl of the library and amplified with the vectorette universal primer (224; see appendix 1 to this chapter) in combination with either primer 1089 (specific for the left-hand side of pYAC4 (TRP1 gene); see appendix 1) or primer 1091 (specific for the right hand side (URA3 gene)); see appendix 1) at final concentrations of 1μM each. The PCR reaction was cycled 38 times at 94^{0}C for 1 minute, 65^{0}C for 2 minutes and 72^{0}C for 4 minutes. A final extension was carried out at 72^{0}C for

10 minutes. The vectorette PCR products were digested and purified as described for *alu*-vector PCR.

2.7.3. 'Junction-trapping' of YAC insert termini.

This method, which is discussed in detail in Chapter 8 relies on the 'trapping' of YAC insert junction sequences by simple plasmid ligation. The positions of the PCR primers that are used in this method, in addition to a schematic illustration of the technique itself are depicted in figures 7.4 and 7.5, Chapter 7, respectively. The method was generally carried out as follows:

Two yeast DNA blocks (equilibrated in T.E.) were cut in half. All four slices were incubated in a final volume of 400µl containing the appropriate Sau3A1 restriction enzyme buffer and 20U of Sau3A1 enzyme. The mix was incubated on ice for 30 minutes to allow diffusion of the enzyme into the blocks and then at 37^0C for a total of 60 minutes. Just prior to restriction digestion, 1U of calf intestinal phosphatase was added to 5'-dephosphorylate the insert DNA. A half-block was removed every 15 minutes and immersed in 0.5M EDTA. After this, the half-blocks were re-equilibrated in T.E. and a small slice from each half-block was taken and electrophoresed in a 0.8% agarose gel to assess digestion. If any or all of the blocks contained partially digested DNA, they were pooled and the DNA purified as follows: A small hole was pierced in the bottom of a 0.5ml eppendorf tube with a hot needle. A small amount of glass wool (up to about 50µl depth) was packed inside. This tube was then placed inside a 1.5ml eppendorf tube with the cap cut off. The yeast blocks were placed on top of the glass wool inside the 0.5ml tube. Both tubes (i.e. one inside the other) were spun at 6500rpm for 10 minutes in a microfuge. The DNA solution, collected at the bottom of the 1.5ml eppendorf tube, was placed into a new tube. To inactivate the phosphatase, the DNA solution was incubated with trinitroacetic acid (10mM final concentration) at 65^0C for 15 minutes and then extracted two times with phenol/chloroform. The DNA was then precipitated. The DNA was electrophoresed, again, to estimate the concentration. Usually, all of the partially digested DNA (≈ 100 to 200ng) was ligated with 10U T4 DNA ligase to 2.5-5ng of BamHI-digested pBluescript plasmid (Stratagene) in a commercial ligation buffer at 16^0C overnight. The ligation mix (10µl) was diluted with an equal volume of distilled H_20 and stored at -20^0C.

The first round of PCR utilised the M13 Universal sequencing (M13F) and reverse sequencing (M13R) primers in combination with the pYAC4-specific vector primers described by Riley et al., (1991) in all possible combinations. For example M13F was used in combination with 1089 or 1091 and so forth for M13R (see appendix 1 and also figure 7.4, Chapter 7). The primers were used in a 10:1 ratio in favour of the pYAC4-specific primers (that is, 0.1µM M13 primer (F or R) and 1µM pYAC4 (1089 or 1091)). In addition, a reaction was set up using both of the M13 primers in combination with each other to assess the ligation reactions. Approximately 1µl aliquots of the ligation mixture were used for each reaction. PCR was carried out as described for the vectorette experiments, except that the annealing temperature was lowered to 55^{0}C and only 30 cycles were performed.

Approximately 1µl of the PCR reactions were aliquoted into a second PCR pre-mix containing the appropriate internal primers for both the plasmid and the pYAC4 vector. Specifically these are as follows: T7 for M13F, T3 for M13R; 1207 for 1089 and 1208 for 1091 (refer to figure 7.4, Chapter 7). Thirty cycles were carried out as before but at an annealing temperature of 57^{0}C (biased toward the annealing of 1207 and 1208, respectively). Approximately 10µl of each of the PCR reactions were examined by gel electrophoresis with or without prior EcoRI digestion (as described for *alu*-vector PCR). As before, the appropriately digested fragments were isolated for further hybridisation studies.

Appendix 1: Oligonucleotide Primers.

Plasmid primers	Primer sequence	Annealing[a]
M13 Universal sequencing primer (M13F)	GTAAAACGACGGCCAGT	55^0C
M13 reverse sequencing primer (M13R)	CAGGAAACAGCTATGAC	55^0C
T3	ATTAACCCTCACTAAAG	$55\text{-}57^0C$
T7	AATACGACTCACTATAG	$55\text{-}57^0C$
pYAC4 primers		
1089 (left-hand-side)	GCGACCCGTTCTCGGAGCACTGTCCGACC	$55\text{-}60^0C$
1091 (right-hand-side)	CGATATAGGCGCCAGCAACCGCACCTGTGG	$55\text{-}60^0C$
1207 (left-hand-side; internal)	GTTGGTTTAAGGCGCAAG	57^0C
1208 (right-hand-side; internal)	GTCGAACGCCCGATCTCAAGC	57^0C
224 (Universal vectorette primer)	CGAATCGTAACCGTTCGTACGAGAATCGCT	$55\text{-}60^0C$
Repetitive sequence-based primers		
Primer IV	CAGAATTCGCGACAGAGCGAGACTCCGTCTC	55^0C
517	CGACCTCGAGATCT(C/T)(G/A)GCTCACTGCAA	55^0C
559 (TC-65)	AAGTCGCGGCCGCTTGCAGTGAGCCGAGAT	55^0C

a Annealing temperatures used in this thesis

Appendix 1: Oligonucleotide primers (cont'd).

PDJ-33	GCCTCCCAAAGTGCTGGGATTACAGGY[b] R[c] TGAGCCC	60^0C
PDJ-34	TGAGCYRW[d]GATCKM[e]K[f]MCACTGCACTCCAGCCTGGG	60^0C
PDJ-66	GATGGTAGTAGGCCACTGCACTCCAGCC	60^0C
LIHs	CATGGCACATGTATACATATGTAAC(A/T)AACC	60^0C

MCC gene primers

MCC, exon fragment 'a'-1[g]	GAATTCATCAGCACTTCT	55^0C
MCC, exon fragment 'a'-2	CAGCTCCAAGATGGAGGG	55^0C
MCC, exon fragment 'b'-1	ATGTTGATTAATCCGTTGGC	55^0C
MCC, exon fragment 'b'-2	ACCCCAGAGCAGAAGGCT	55^0C

---

b Y = adenine or guanine

c R = cytosine or thymine

d W = adenine or thymine

e M = guanine of thymine

f K = adenine or cytosine

g Amplification using 'a'-1 and 'a'-2 gives rise to fragment 'a' which contains *MCC* exon nucleotides exon nucleotides 391-533; 'b'- 1 and 'b'-2 gives rise to MCC fragment 'b' which contains *MCC* exon nucleotides 1679 to 1862

Appendix 2: Sources of DNA probes.

Probe	Source	Reference
YN5.48	Dr. Ray White	Nakamura et al., (1988)
ECB220F.1	D. Lily Varesco	Varesco et al., (1989)
p3.1	Mr. Tristan Ward	Spurr et al., (1989)
SW15 (*MCC*)[a]	Dr. Kenneth Kinzler	Kinzler et al., (1991a)
MCC40cI (*MCC*)[b]	Dr. Kenneth Kinzler	Kinzler et al., (1991a)
λ12.75	Ms. Sally Cottrell	Thomas (1991)[c]
λ14.16	Ms. Sally Cottrell	Thomas (1991)
λ14.17	Ms. Sally Cottrell	Thomas (1991)
λ12.46	Ms. Sally Cottrell	Thomas (1991)
EF5.44	Dr. Yusuke Nakamura	Dunlop et al., (1990)
L5.62	Dr. Yusuke Nakamura	Dunlop et al., (1990)
CB83.6	Dr. P. Meera Khan	Breukel et al., (1991)

Other probes cited in this thesis (generated during the project) are described in the text.

[a] SW15 is a cDNA probe representing nucleotides 133 to 1918 of the *MCC* gene.

[b] MCC40cI is a cDNA probe representing nucleotides 1634 to 3969 of the *MCC* gene.

[c] The 'l' series clones were generated from genomic libraries constructed from radiation reduced hybrids containing fragments of chromosome 5.

Appendix 3. Sources of human-hamster somatic cell hybrids.

Hybrid	Source	Reference	Comments
PN/Ts-1	Dr. Huw Thomas	Varesco et al., (1989)	Segregating a normal human chromosome 5 from patient PW as the only human material.
PD/Ts-1	Dr. Huw Thomas	Varesco et al., (1989)	Segregating an abnormal chromosome 5 (5del(5q21-q22)) in the absence of the normal homolog from patient PW.
MD/Ts-1	Dr. Huw Thomas	Varesco et al., (1989)	Segregating an abnormal chromosome 5 (5del(q21-q22) as the only human material. This somatic cell hybrid was constructed from a patient (MD) different from the patient used for the derivation of PN/Ts-1 and PD/Ts-1 (i.e. PW).
SD/Ts-1	Dr. Huw Thomas	Hampton et al., (1992), submitted	Segregating an abnormal chromosome 5 (5del(q22.1-q23.3) in the absence of the normal homolog from patient PS.
HHW1155	Dr. J. Wasmuth	Joslyn et al., (1991)	Segregating a chromosome 5 with a small, (≈260kb) interstitial deletion which encompasses the *APC* gene. This hybrid was derived from patient 3214 (Joslyn et al., 1991).
HHW1159	Dr. J. Wasmuth	Joslyn et al., (1991)	Segregating the normal chromosome 5 from patient 3214 in the absence of the abnormal homolog.

Chapter 3.

Repetitive sequence polymerase chain reaction (PCR).

Chapter 3.

Repetitive sequence polymerase chain reaction (PCR).

Part 1. Rapid generation of human chromosome-specific DNA probes from a somatic cell hybrid.

3.1. Introduction.

The experiments described in the first part of this chapter were designed to exploit the presence of genomic repeated sequences as templates in a PCR-based strategy to generate human chromosome-specific DNA probes useful for mapping studies.

Mammalian genomes contain a number of interspersed repetitious DNA sequence families which are broadly categorised into two main classes: (1) short interspersed repetitive sequences (SINES), most commonly represented in the human genome by the *alu* sequences, which number some 9×10^5 copies (Britten et al., 1988), and (2) long interspersed repetitive sequences (LINES), the most common of which are the LIHs or KpnI family, present in about 10^4 - 10^5 copies (Jelinek and Schmid, 1982). These sequences are believed to have arisen in such numbers by retrotransposition (see van Ardshell et al., 1981; Bains, 1986), a process analogous in part to retroviral integration in mammalian genomes (reviewed in Varmus, 1989). Retrotransposition is an ongoing process, particularly noticeable in unstable genomes such as those in tumour cells (Bratthauer and Fanning, 1992). However, the finding of an integrated LIHs element in the Factor IX gene (Kazazian et al., 1988) suggests that the process occurs naturally in the germ-line.

Alu sequences are approximately 300bp in length (Deininger et al., 1981) and are about 85% homologous to each other (Bains, 1986). They are generally surrounded by sequences rich in adenine and thymine nucleotides which has suggested that these may be preferential integration sites in the genome (Bains, 1986; Kariya et al., 1987). *In situ* hybridisation of *alu* sequences to human metaphase chromosomal spreads shows a clear bias in distribution toward G-negative regions (Korenberg and Rykowski, 1988) which are highly AT rich. *Alu* sequences can be orientated either toward each other or in the

same direction. In general, where *alu* sequences are spaced more closely than 500bp, they tend to be orientated in the same direction (reviewed in Moyzis et al., 1989). On a gross genomic level, it has been estimated that *alu* elements are in opposing orientations about 35% of the time (Moyzis et al., 1989). Equivalents to the human *alu* sequences exist in other mammalian genomes and are particularly well characterised in mouse, hamster and ape (Jelinek and Schmid, 1982). Sufficient sequence divergence between man and rodent exists, however, to enable differentiation in primer template design (see below).

Based on a revised *alu* consensus sequence (Kariya et al., 1987), Drs. Finbarr Cotter and Bryan Young (ICRF Medical Oncology Laboratories) designed oligonucleotide primers to various positions of the *alu* element. Cotter et al., (1990) showed that these primers (which are schematically depicted in figure 3.1) when used alone or in combination, could allow amplification between adjacent *alu* elements in the same or inverted orientations in human genomic DNA. More formal evidence for this 'inter-*alu*' amplification was provided by direct di-deoxy sequencing of a heterogenous mixture of the genomic amplification products. These experiments clearly demonstrated that the PCR products did indeed represent amplification between adjacent elements, thus 'trapping' potentially single-copy DNA sequences (Cotter et al., 1990).

I reasoned that useful human-specific DNA probes could be derived from more complex sources, such as human chromosomes or chromosomal fragments retained in inter-species somatic cell hybrids, given that sufficient sequence divergence did indeed exist between rodent and human *alu* sequences (Jelinek and Schmid, 1982). Such a method was expected to allow direct and rapid access to sequences in sub-chromosomal regions of interest, obviating the need for time-consuming construction and screening of genomic libraries.

3.2. Results.

3.2.1. *Alu*-PCR from a somatic cell hybrid.

As a model system to test this hypothesis, a somatic cell hybrid - PN/Ts-1, was chosen. This somatic cell hybrid contains a normal chromosome 5 as its only human complement on a hamster background (Varesco et al., 1989). The PCR

experiments performed by Cotter et al., (1990), in addition to the results of exhaustive sequence analysis (Moyzis et al., 1989), predicted that *alu* elements could reside in opposing orientations sufficiently often to allow each primer to be used in isolation for assessment of human specificity (see figure 3.2(b)). Of each of the four primers tested (figure 3.1), primer IV appeared to be the most human-specific, yielding amplification products from the hybrid and only a very faint background smear in the hamster DNA under the PCR conditions used (see figure 3.2(a)). In addition, mouse DNA was also included in the example shown in figure 3.2(a). Mouse DNA gave rise to a reproducibly specific band at about 2kb with no background smear, suggesting that, like hamster-based hybrids, this primer would also be useful on inter mouse-human hybrids. It can be seen that amplification of the PN/Ts-1 hybrid DNA gave rise to about 15-20 discrete PCR products on a background smear. These products ranged in size from 300bp to 2kb with the majority between 600bp and 1.4kb. Amplification of different PN/Ts-1 hybrid cell DNA isolates demonstrated that these banding patterns were highly reproducible.

3.2.2. Cloning of inter-*alu* PCR products.

To investigate the origin of these PCR products, it was decided to clone the amplification products and to use individual cloned inserts for hybridisation analyses. The oligonucleotide primers depicted in figure 3.1 had originally been designed to contain sites for specific restriction enzymes at the 5' end of the sequences: for example, primer IV contains an EcoRI recognition site (underlined in figure 3.1). A number of attempts were made to clone the inter-*alu* PN/Ts-1 PCR products by digestion with EcoRI and ligation of the products to EcoRI digested pBluescript plasmid. However, even with careful DNA purification and the use of excess restriction enzyme, no recombinant colonies could be recovered. This was most likely due to the fact that the restriction site is too close to the 5' end of the primer. It has been suggested that specific restriction enzymes require a significant number of nucleotides protruding past the 5' end of the primer sequence for efficient digestion (Aslanidis and deJong, 1990; Dr. Pieter deJong, Lawrence Livermore National Laboratory, U.S.A., personal communication).

To circumvent this problem, I devised a simple strategy. It was assumed that the PCR products should contain 'blunt' or 'flush' ends due to the nature of the PCR amplification reaction and thus be amenable to cloning into pBluescript cut at an appropriate 'blunt'-site. EcoRV was chosen as the

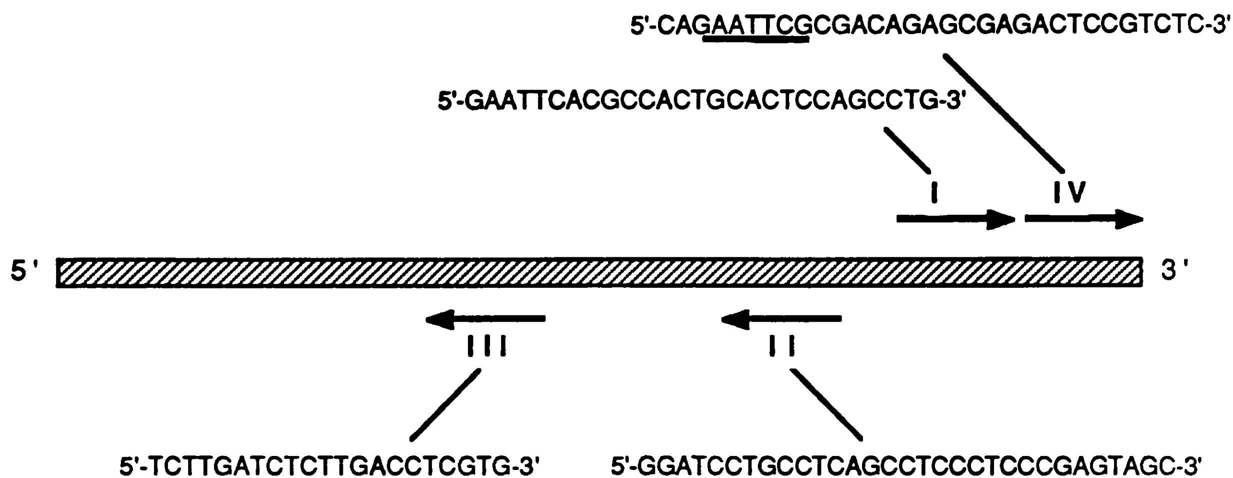


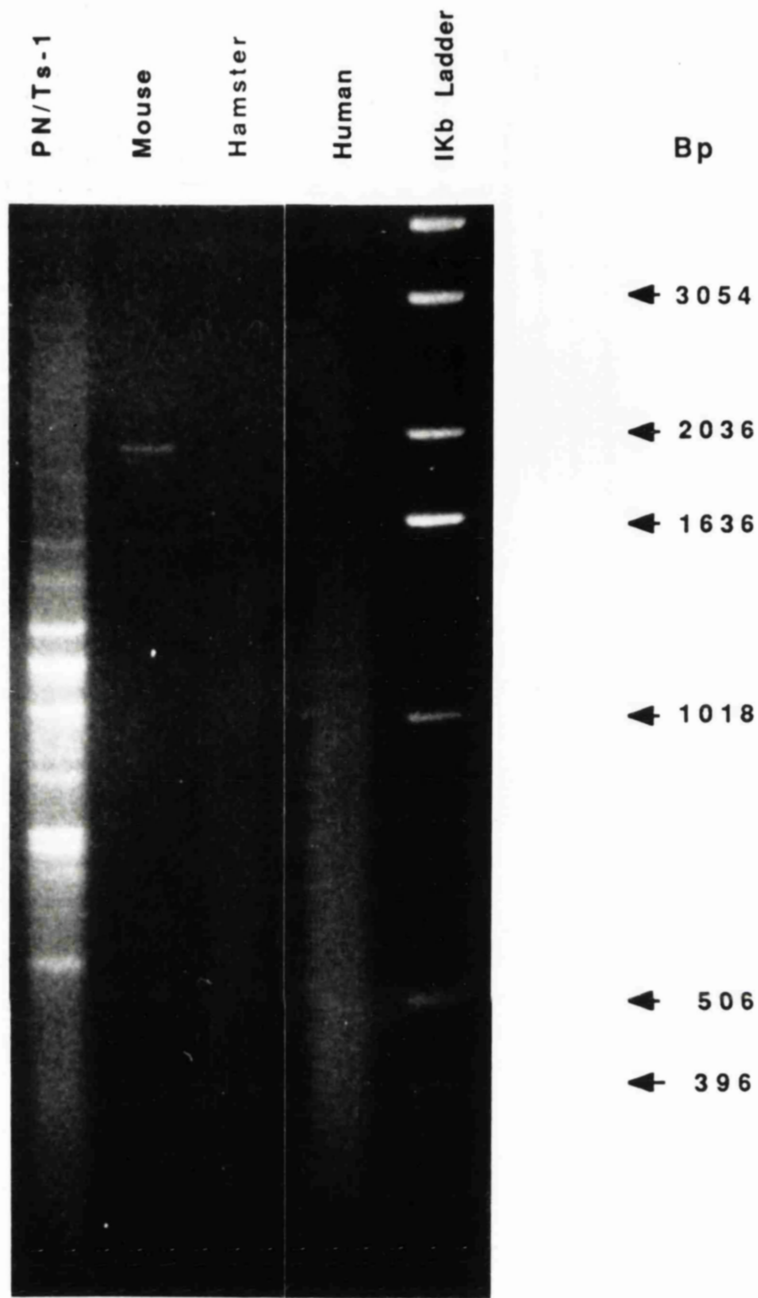
Figure 3.1. Position and sequence of *alu*-based oligonucleotides.

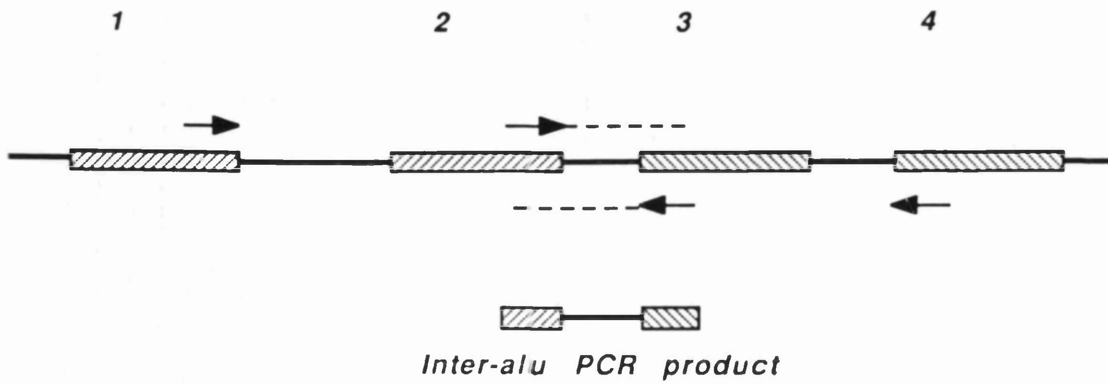
The hatched box represents the 280bp human *alu* consensus sequence. The relative positions and orientations of the four oligonucleotide primers (I, II, III, IV) are denoted by the arrows above and below the schematic *alu* element. The EcoRI site in primer IV is indicated by the bold line.

Figure 3.2. ***Alu*-directed amplification of hybrid PN/Ts-1.**

(A) Electrophoresis of *alu*-primed PCR products (1/10th of total) on a 1.5% agarose gel electrophoresed at 5V/cm for 10 hours and stained with ethidium bromide. The identity of the genomic DNA samples used in each of the PCR reactions are indicated above the figure. The size distribution of the amplified products is illustrated by the 1kb ladder DNA markers displayed on the right side of the figure. Note the appearance of a specific band in mouse DNA at approximately 2kb.

(B) Schematic illustration of inter-*alu* PCR. The figure represents a random stretch of human genomic DNA with a number of interspersed *alu* sequences. Elements **1** and **2** are depicted in the 5'-3' orientation, elements **3** and **4** are depicted in the 3'-5' orientation with respect to the consensus sequence (Kariya *et al.*, 1987). An oligonucleotide designed to the 3' end of the *alu* element (e.g. **primer IV**) is denoted by the arrows. An inter-*alu* PCR product can only be generated between elements **2** and **3** which are in inverse orientation with respect to each other.


A
PN/Ts-1
Mouse
Hamster
Human
IKb Ladder
Bp
3054
2036
1636
1018
506
396


B
1
2
3
4
Inter-alu PCR product

'blunt' end generating enzyme since it cuts extremely well and leaves few undigested molecules. Although 5'-dephosphorylation of the vector would have prevented re-circularisation of the plasmid without an insert (and be more efficient), the oligonucleotides did not contain phosphate groups at their 5' ends and would, therefore, have been unclonable. To avoid this, simple ligations, without prior enzymatic dephosphorylation of the vector were performed.

To attain a reasonable ratio of recombinants to non-recombinants, a fixed amount of PCR material (approximately 100 nanograms; ng) was ligated to varying amounts of vector DNA in the presence of Polyethylene Glycol (PEG) at a final concentration of 5% to enhance the ligation efficiency. This compound serves as a volume 'excluder' such that the local concentration of available DNA ends in solution is very high (Hayashi et al., 1986; Ludecke et al., 1989). A molar ratio of about 1:0.1 PCR material to vector (i.e. about 2.5ng) was usually ideal and gave libraries with ratios of about 70:30 recombinants to non-recombinants. The cloning efficiency was calculated at about 5×10^4 clones per microgram of PCR material using transformation competent cells with a theoretical efficiency of about 10^7 colonies per microgram of DNA.

3.2.3. Analysis of the 'inter-*alu*' PCR libraries.

For rapid analysis of potentially recombinant plasmids, single bacterial colonies were analysed by a PCR-based protocol as follows: Bacterial cells, presumed to contain recombinant plasmids, were added directly to a PCR pre-mix with sterile toothpicks and the cells lysed simply by heating in a PCR thermocycler at 94^0C for 10 minutes (see Chapter 2). *Taq* polymerase was added and the PCR reaction cycled in the normal way. Using this protocol, it was possible to analyse up to 50 clones, albeit amplified in individual eppendorf tubes, in a single experiment with no preparative procedures. Direct colony PCR was performed using the original *alu* primer IV, such that the fidelity of the cloning reaction could be assessed. In most cases recombinant clones gave rise to PCR products suggesting that inter-*alu* products must have been cloned. Examples of direct colony-PCR are shown in figure 3.3(a). It can be seen that these inserts represent a variety of sizes, essentially reflecting the size range demonstrated for the *alu*-PCR amplification of PN/Ts-1 in figure 3.2(a).

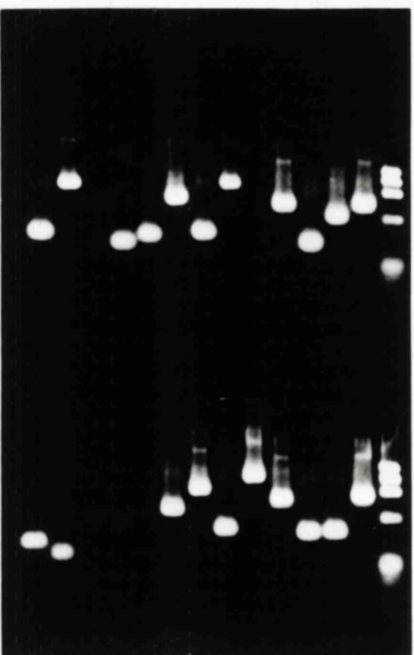
3.2.4. Hybridisation analysis of 'inter-*alu*' PCR products.

Figure 3.3. Analysis of an inter-*alu* PCR product library from hybrid PN/Ts-1.

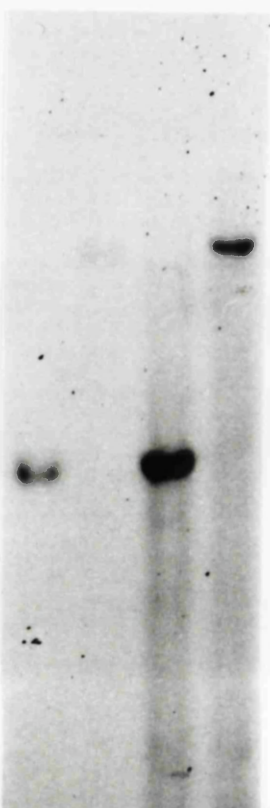
(A) Electrophoresis of PCR products from 26 individual bacterial colonies amplified using *alu* primer IV on a 1.5% agarose gel stained with ethidium bromide (Lanes **1-13** and **14-26**). Lanes **16-18** represent amplifications from 'blue' colonies (i.e. non-recombinant clones not expected to give products). Some white (i.e. potentially recombinant) colonies did not give rise to any detectable products (lanes **3** and **9**). Lanes marked **(M)** contain phage phiX174 DNA digested with HaeIII.

(B) Southern blot hybridisation analysis using probes generated from three random inter-*alu* PCR fragments. Each fragment was radiolabelled, pre-annealed to sheared and sonicated human placental DNA (Chapter 2) and hybridised to filters containing: 10μg of human **(Hu)** and 10-15μg of PN/Ts-1 **(PN)** DNAs digested with EcoRI **(E)** and HindIII **(H)**. The sizes of the hybridising fragments are indicated to the left of each 'panel'. The signal strengths of hybridisation fragments in the higher molecular weight size range are weak due to poor Southern transfer. The filters were each washed to a final stringency of 0.2 x SSC/0.1% SDS at 65^0C and exposed to autoradiographic film for 2 to 4 days.

A



B



Forty-eight different recombinant clones derived from an inter-*alu* PCR library of PN/Ts-1 were gridded out onto nylon membranes and screened with [α^{32}P]-dCTP labelled total human genomic DNA to select apparently 'single-copy' inserts for hybridisation analysis. In this experiment, 6 clones gave highly intense signals and 2 clones exhibited weak hybridisation signals. The other 40 clones did not give any signal with the genomic probe which suggested that only about 15% of the inserts contained either highly divergent *alu* sequences or elements of other common repeat sequence families. Six of the 40 inserts negative in the screening experiment were chosen for hybridisation analyses on Southern blots containing digested human and PN/Ts-1 DNAs. The labelled inserts were competed with unlabelled and sheared human placental DNA to quench any repetitive signal contributed by the 30 bases of *alu* sequence bounding each side of the inter-*alu* products. All of the 6 inserts were shown to be human in origin and all hybridised to PN/Ts-1 DNA with patterns identical to genomic DNA. These experiments confirmed the human origin of the PCR products, most of which are likely to be derived from chromosome 5. Examples of three of these hybridisations are shown in figure 3.3(b).

3.3. Conclusions.

Alu-PCR amplification of human genomic DNA in addition to direct sequencing of the PCR products (Cotter et al., 1991), together confirmed the expectation that the *alu*-based primers could allow specific amplification of genomic DNA between two *alu* sequences (i.e. inter-*alu* PCR).

With this evidence, the four *alu*-based primers depicted in figure 3.1 were then applied to the genomic DNA of a somatic cell hybrid. Of the four primers tested with PN/Ts-1, primer IV appeared to be the most human specific. Comparative analysis of the consensus sequences between species (Jelinek and Schmid, 1982) suggests that this specificity is probably due to the following factors: (1) the terminal cytosine nucleotide at the 3' end of the human *alu* consensus sequence (included in primer IV) is not present in the hamster *alu* equivalent and (2) the cytosine and adenine residues at positions -5 and -6 from the 3' end of the human consensus sequence are usually represented by thymine and guanine nucleotides, respectively, in hamster (Jelinek and Schmid, 1982). A lack of inter-species homology of these nucleotides at the 3' end of the primer is probably sufficient to preferentially

bias the reaction, in a competitive sense, toward the amplification of human sequences. In the absence of any human DNA sequences, amplification gives rise to only a very faint background smear.

Amplification of the PN/Ts-1 hybrid with the single *alu* primer IV, produced discrete and reproducible sets of PCR products suggesting that only a limited number of potential 'inter-*alu*' sequences had been selected for amplification in the PCR reaction. These results are similar to those published by Nelson et al., (1989) at the end of these initial studies. The restricted set of PCR products from these somatic cell hybrids is probably due to an inter-play of three key factors: (1) that the *alu* repeat sequences need to be sufficiently close to each other, (2) that they have sufficient homology to the primer and, (3) that they must be in opposite orientations to allow amplification to occur when only a single primer is used in the reaction.

It has been estimated that *alu* sequences are about 85% homologous to each other (Bains, 1986). Although absolute homology will not be required, it is probable that a template competition effect exists in the PCR reaction favouring those pairs of sequences most homologous to the primer. The background smears seen in all amplifications of single chromosome hybrids in this and other similar studies (Nelson et al., 1990) probably represent a very large number of minor amplification events, less favourable in a competitive reaction. This competitive effect is discussed further in Chapter 9.

The distribution of *alu* sequences is also quite variable. Although the average spacing of repeats was originally estimated at approximately 4kb (Rienhard et al., 1981), the true distribution of these sequences can be best explained by a bimodal model assuming '*alu*-rich' and '*alu*-poor' domains with average spacings of 1 kb and 10 kb respectively (Moyzis et al., 1989). In the case of the PN/Ts-1 amplifications, it is likely that '*alu*-rich' domains are selectively targetted for PCR since the reaction will probably tend to favour smaller amplification products. As suggested in Cotter et al., (1990), modification of PCR conditions to allow amplification of longer PCR products (e.g. see Ponce and Micol, 1992), may increase the complexity of the amplification mixture.

Part 2. Applications of the *alu*-PCR method: comparative analysis of somatic cell hybrids

3.4. Introduction.

On the basis of the discrete and highly reproducible patterns generated from the PN/Ts-1 hybrid using primer IV (figure 3.2(a)), I reasoned that if a disease-related deletion were present in a chromosome retained as the sole human constituent of a somatic cell hybrid, a number of the *alu* sequences in the same region would be deleted. If these particular sequences were present and amplified as discrete products in the normal homolog of that chromosome, a change in the banding pattern of the other should be observed.

At the time of these studies, Nelson et al., (1989) also demonstrated the amplification of human specific sequences from somatic cell hybrids retaining whole human chromosomes using oligonucleotide primers different from primer IV described above. These primers, 517 and 559 were both designed to the same position in the *alu* sequence but are reverse complements of each other and therefore prime in opposite directions from the *alu* elements.

In addition, Ledbetter et al., (1990) described another oligonucleotide that specifically primes from the conserved 5' region of the 6kb LIHs/KpnI LINES human repeat family. When used to amplify somatic cell hybrid DNA, this primer also gives rise to discrete and reproducible patterns. The numbers of products generated are fewer, however, than the numbers generated by *alu*-PCR and probably reflect the lower numbers of this repeat sequence family in the human genome. The authors also demonstrated that both the *alu* and LIHs primers could be used in the same reaction to produce discrete, but entirely different patterns than those generated by either primer used alone. To encompass the use of both primers, Ledbetter et al., (1990) have used the term 'Interspersed Repetitive Sequence (IRS)-PCR' to describe these experiments.

For the experiments described in the following sections, two somatic cell hybrids were used: PN/Ts-1 and another hybrid; MD/Ts-1. The latter hybrid contains a single copy of chromosome 5 with a cytogenetically visible deletion thought to encompass the *APC* gene, around region q21-q22 (Varesco et al., 1989). *In situ* hybridisation using biotin dUTP labelled total human DNA suggested that like PN/Ts-1, this hybrid contains chromosome 5 as the only

human DNA constituent. Thus these hybrids could be used for direct comparison by IRS-PCR.

3.5. Results

3.5.1. IRS-PCR of somatic cell hybrids: Primer IV

To test the hypothesis of differences in the *alu*-PCR patterns between the normal and deleted chromosomes 5, primer IV was used to amplify the PN/Ts-1 and MD/Ts-1 hybrid DNAs (see figure 3.4(a)). As shown before in figure 3.2(a), amplification of PN/Ts-1 gave rise to about 15 to 20 discrete products ranging from about 300bp to 2kb in size. The pattern generated from MD/Ts-1 was extremely similar, if not identical. None of the intense PCR products thus appeared to originate from within the region defined by the MD deletion. It is, however, likely that pairs of *alu* sequences in the appropriate orientation and distance apart do exist in this region. However, *alu* sequences slightly diverged from the primer sequence will appear very weakly or not at all, presumably due to the competitive effect.

3.5.2. IRS-PCR of somatic cell hybrids: Primers 517, 559 and LIHs

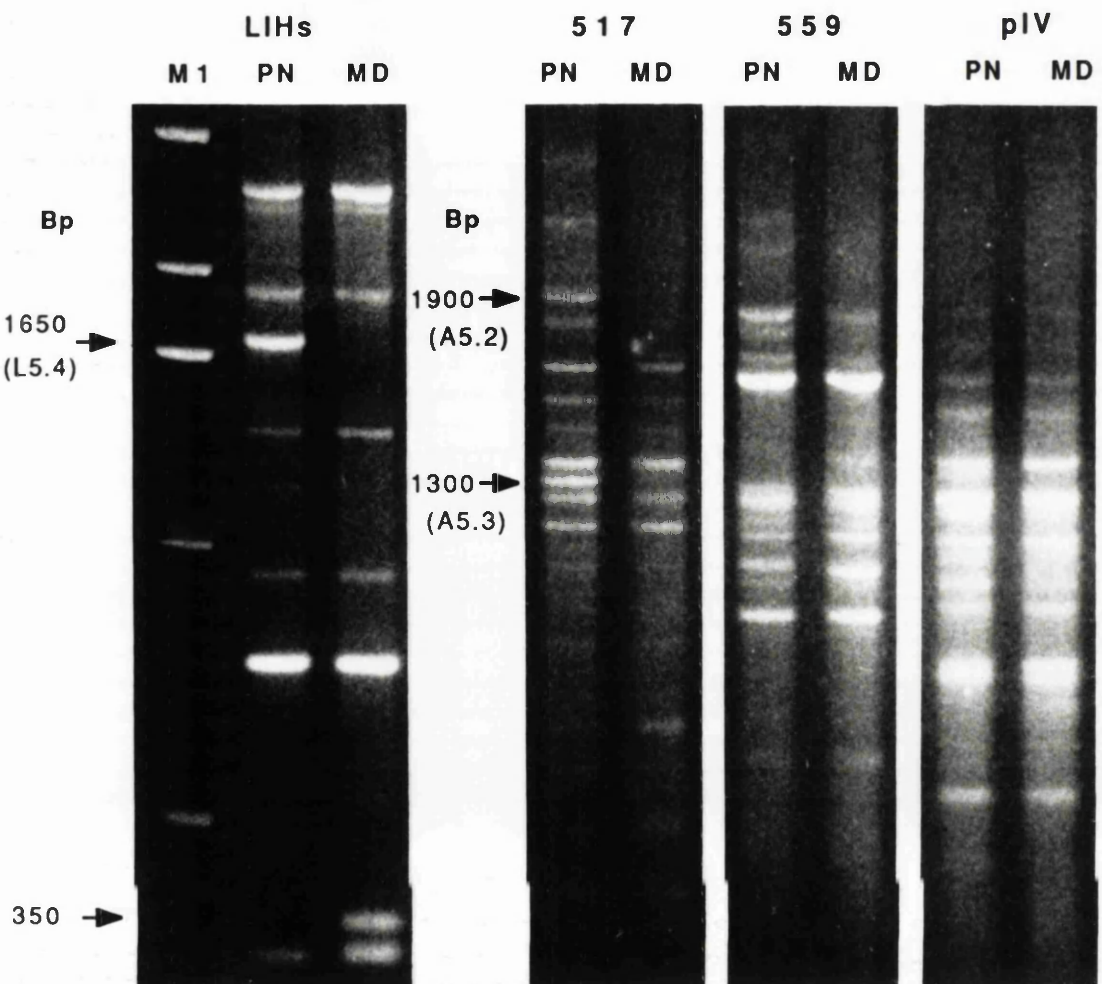
The primers 559, 517 and LIHs, described by Nelson et al., (1989) and Ledbetter et al., (1990), respectively, were then used individually to amplify both the normal and deleted chromosome 5 cell hybrids (figure 3.4(a)). As for primer IV, it can be seen that the patterns generated from both of the hybrids are very similar for each individual primer used. However, the products are entirely dissimilar when different primers are used on the hybrids. Amplification with the LIHs primer generated some 13 major and minor products. One band at 1.65kb (designated L5.4) was clearly present in PN/Ts-1 but not in MD/Ts-1. Interestingly, there was an intense band at 350bp which was present in the deleted but not the normal chromosome. Possible reasons for the appearance of this additional band in MD/Ts-1 are discussed in the conclusions to this chapter.

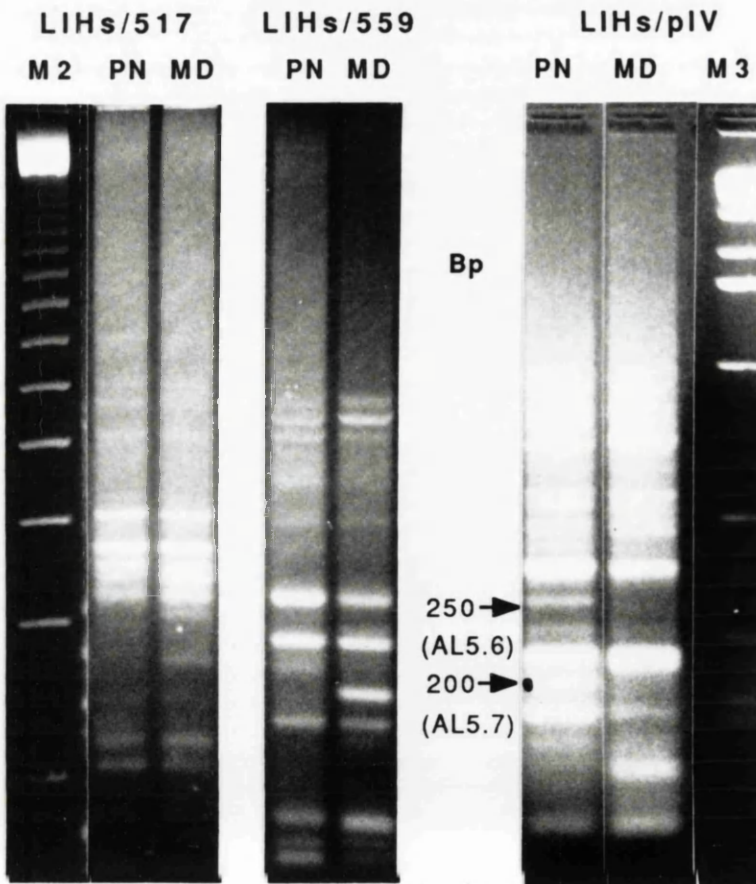
Alu primer 517 generated about 20-25 major products with a size range of 400bp to 3kb while primer 559 generated approximately 15 major products from 600bp to 2.5kb. The 517 amplification reactions revealed two banding pattern differences between the normal and deleted hybrids at approximately

Figure 3.4. Comparative repetitive sequence PCR analysis of normal and deleted chromosomes 5.

(A) Electrophoresis of PCR reaction products from PN/Ts-1 **(PN)** and MD/Ts-1 **(MD)** hybrids on a 1.5% agarose gel electrophoresed at 5V/cm for 10 hours and stained with ethidium bromide. The primers used for each amplification reaction are indicated at the top of the gel lanes. Lane **(M1)** contains the 1kb ladder DNA size markers. Sizes of fragments **L5.4, A5.2** and **A5.3** are indicated by arrowheads to the left of the figures.

(B) Electrophoresis of PCR reaction products from PN/Ts-1 **(PN)** and MD/Ts-1 **(MD)** hybrids on 3% *Nusieve*™ agarose gels electrophoresed at 7V/cm for 10 hours and stained with ethidium bromide. The primer combinations used for each of the amplification reactions are indicated above the gel lanes. Lanes **(M2)** and **(M3)** contain the 100bp ladder and phage phiX174 DNA digested with HaeIII, respectively. Sizes of fragments **AL5.6** and **AL5.7** are indicated by arrowheads to the left of the figure. Note the extra bands in **(MD)** that are not present in the **(PN)** reaction products in the LIHs/559 combination.


A (alu)
LIHs
M1 PN MD
517
PN MD
559
PN MD
pIV
PN MD
Bp
1650
(L5.4)
350
Bp
1900
(A5.2)
1300
(A5.3)


B (alu, lines)
LIHs/517
M2 PN MD
LIHs/559
PN MD
LIHs/pIV
PN MD M3
Bp
250
(AL5.6)
200
(AL5.7)

1.9kb (A5.2) and 1.3kb (A5.3). However, no such differences were observed between the amplification products of the two hybrids using primer 559.

The LIHs primer was also used in combination with each of the three *alu* primers (figure 3.4(b)). For each combination, the patterns generated were entirely different from those generated by any of the primers when used alone and, because of the generally smaller size, had to be resolved on 3% NuSieve GTG agarose gels for long periods of time (see Chapter 2). The products ranged in size from approximately 100bp up to a maximum of approximately 1.0kb with a heavy background smear in each case. Two fragments of 250bp (AL5.6) and 200bp (AL5.7) were observed in the normal but not the deleted chromosome 5 using the LIHs-pIV combination. No additional products were seen in the normal chromosome when compared to the deleted chromosome for the LIHs-559 or LIHs-517 combinations, but, it can be seen that a number of products appeared only in the MD/Ts-1 hybrid, especially in the LIHs-559 reactions (see conclusions).

3.5.3. Isolation and mapping of IRS-PCR 'difference' products

To generate enough DNA product for labelling and mapping of these PCR fragments, the amplification products were eletrophoresed in low melting point agarose gels and the difference products excised. These gel slices were diluted with an equal volume of sterile water and 5µl of this dilution used in a further PCR reaction. Most of the PCR fragments, except for the two difference products generated by primer 517, could be re-amplified. These two differences were resistant to re-amplification under a large variety of PCR conditions, even after re-isolation from a number of low melting point agarose gels. In general, a complex smear was obtained in each case and the reason for this is unknown.

The successfully re-amplified products; L5.4, AL5.6 and AL5.7 were labelled with [α^{32}P]-dCTP and mapped with respect to a somatic cell hybrid panel containing PN/Ts-1, MD/Ts-1 and another hybrid, PD/Ts-1. This independently available hybrid contains a homolog of chromosome 5 with a deletion that is cytogenetically indistinguishable to that of MD. However, at a molecular level, hybridisation of a number of chromosome 5-specific DNA probes have shown that the MD deletion extends further distally than this PD deletion (Varesco et al., 1989). Each probe was competed with unlabelled and sheared total human placental DNA according to Sealy et al., (1985) to quench

residual *alu* and/or LIHs repetitive sequences. Under the conditions used, it was only possible to map the inter-LIHs product, L5.4. The other two products generated from the LIHs-*alu* combinations; AL5.6 and AL5.7, gave no appreciable hybridisation signals. This was also found to be the case for these two probes using alternative competitive hybridisation conditions described by Nelson et al., (1989).

L5.4 hybridised to the normal chromosome 5 hybrid DNA but not to the other two deletion chromosome hybrids (figure 3.5), and thus mapped to the minimal region near the *APC* gene defined by the two interstitial deletions.

3.5.4. Cloning of L5.4.

The L5.4 fragment was purified and cloned as described for the construction of the PN/TS-1 *alu*-PCR library. Potential recombinants were tested by direct colony PCR using the LIHs primers. Each of 10 colonies presumed to contain recombinant plasmids gave rise to products that were of the correct size as determined by agarose gel electrophoresis. In addition, purified PCR product from one of the colonies (termed pL5.4) was labelled and mapped with respect to the somatic cell hybrids, confirming that the correct sequence had been cloned.

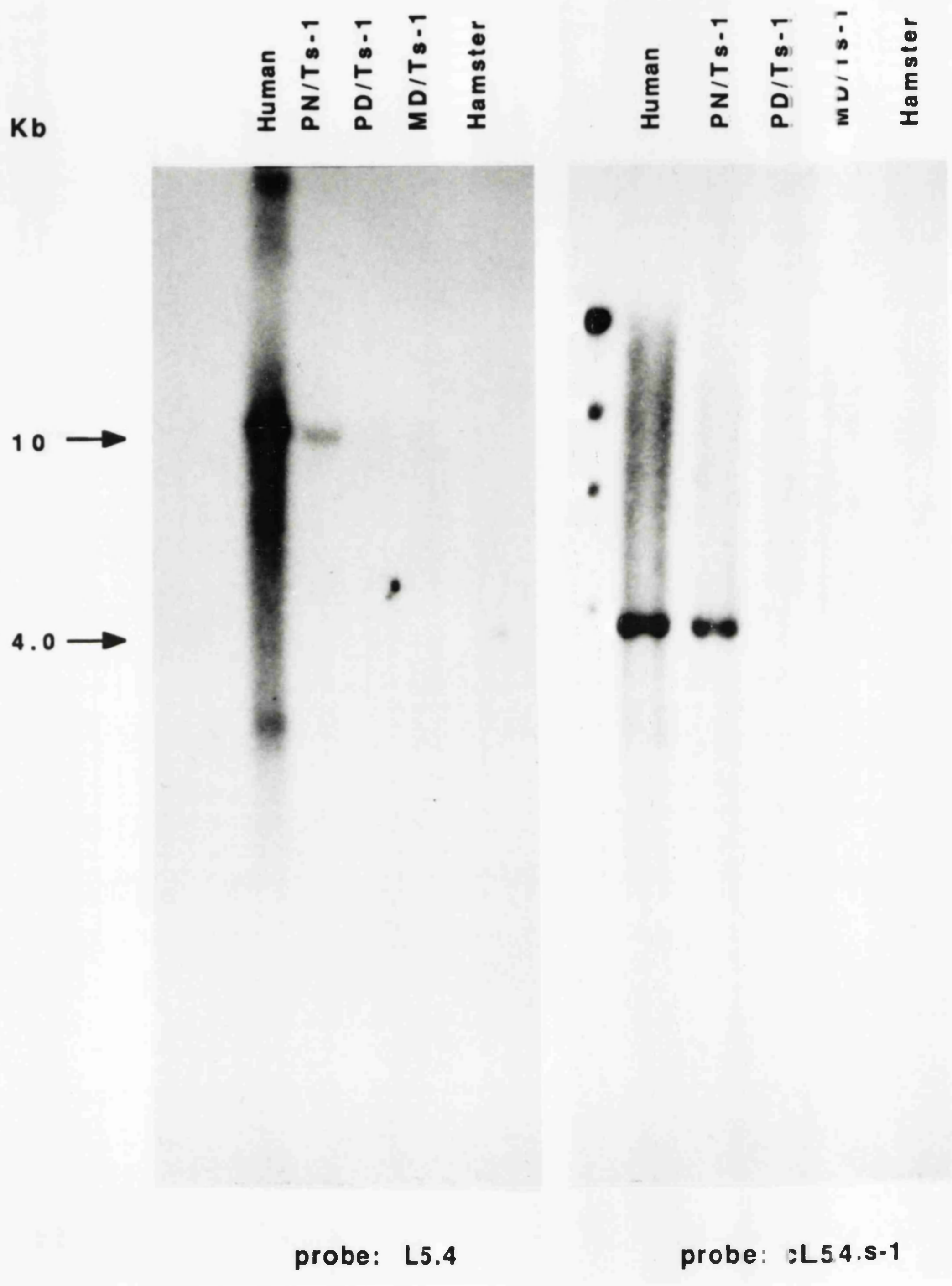
3.5.5. Isolation of a cosmid containing the L5.4 sequence.

Hybridisation of pL5.4 to human DNA shows that this fragment is still quite repetitive even after competition with placental DNA (see figure 3.5). As explained in Chapter 5, it was intended to use each clone isolated in the region around the *APC* gene in a hybridisation assay for the selection of yeast artificial chromosomes (YACs) to expand the amount of cloned DNA around 5q21-q22. L5.4 could not be used in such an assay as the ICRF YAC hybridisation filters required very high signal-to-noise ratios for the selection of positive clones (see Chapter 5). To generate an appropriate single-copy fragment, it was decided to isolate a cosmid for L5.4.

The cosmid library used for this experiment was constructed from flow-sorted chromosome 5 in the sCOS-1 cosmid vector (described in Evans and Wahl, 1987) at the Los Almos National Laboratory, U.S.A.. Library clones, representing approximately two genome equivalents of chromosome 5, were spotted in high density gridded arrays of approximately 9000 clones 22 x 22cm

Figure 3.5. Southern blot analysis of probes L5.4 and cL5.4.s-1.

Southern blot hybridisations of the inter-LIHs product **L5.4** (left hand side) and the cosmid sub-clone **cL5.4.s-1** (right hand side) on filters containing EcoRI digested genomic DNAs, the identities of which are indicated above the panels. The sizes of the hybridising fragments are indicated to the left of the figure. **L5.4** was pre-annealed with sheared and sonicated human placental DNA as desribed (Chapter 2). Sub-clone **cL5.4.s-1** was not pre-annealed. Both filters were washed to a final stringency of 0.2 x SSC/0.1% SDS at 65^0C and exposed to autoradiographic film for 3 days.


Kb
Human
PN/Ts-1
PD/Ts-1
MD/Ts-1
Hamster
Human
PN/Ts-1
PD/Ts-1
Hamster
10
4.0
probe: L5.4
probe: cL54.s-1

nylon filters (see Nizetic et al., 1991) by Dr. Anna-Maria Frischauf (Analysis of Mammalian Mutation laboratory, ICRF). Hybridisation of L5.4 resulted in the detection of one positive cosmid, designated cL5.4. The sCOS-1 cosmid vector contains sites for the rare-cutting restriction endonuclease; NotI, which flank the BamHI cloning site. The insert size could therefore be determined simply by digestion of cosmid DNA with NotI and electrophoresis on 0.3% agarose gels. The cL5.4 cosmid insert was found to be approximately 38kb.

The cosmid was shown to be a true positive as follows. Firstly, single colonies of the cosmid containing bacteria were directly amplified using the LIHs primer with appropriate negative controls. Each of the colonies tested gave rise to only one inter-LIHs product of 1.65kb, identical to in size to L5.4. Secondly, the cosmid DNA was labelled, competed extensively with sheared placental DNA and hybridised to the deletion hybrid panel described above. The cosmid hybridised to PN/Ts-1 DNA, exhibiting 8 bands including a 10kb band identical to that identified by pL5.4. All of these cosmid bands were absent in both of the hybrids containing the chromosome 5 interstitial deletions.

To generate single-copy fragments, cosmid DNA was digested with the following restriction enzymes: EcoRI, HindIII, BamHI, PvuII and BglII. Southern blots of the restriction digests were hybridised with labelled total human genomic DNA. Three fragments that showed no signal with genomic DNA were isolated and used as probes on the mapping panels. A 1.1kb BglII fragment hybridised with the least repetitive sequence background at about 4kb on EcoRI digested DNAs. This sub-fragment is clearly absent in both deletions hybrid cell DNAs (figure 3.5). The use of this fragment for the isolation of YACs is presented in Chapter 5.

3.6. Conclusions.

The IRS-PCR amplifications of chromosomes derived from different individuals were shown to be extremely constant and can be thought of as 'PCR-karyotypes' of the chromosomes. However, differences were observed in banding patterns between normal and deleted chromosomes 5, one of which, probe (L5.4), was found to map near the *APC* gene. The simplicity of the method together with the lack of any specialised techniques compares it favourably to other subtractive cloning methods such as the phenol enriched reassociation technique (PERT; Kunkel et al., 1985).

The lack of appreciable hybridisation signal from products AL5.6 and AL5.7 was probably due to the fact that these fragments contain residual *alu* and LIHs sequences representing approximately 25 and 30% of each probe, respectively (perhaps more if one or both contain additional repeat elements). Taken together with the small size of the fragments (250bp and 200bp respectively) and the unavoidable reduction of 'single-copy' signal by competition, these results are not unexpected.

It is clear that in addition to 'missing' bands in MD/Ts-1, there were a number of additional products present in the MD/Ts-1, but not the PN/Ts-1 hybrid (see figure 3.4(a) and (b)) which was unexpected. There are a number of possible reasons for the 'appearance' of these additional bands:

It can be argued that the bands reflect sequence deviation between two chromosomes from two unrelated individuals. Such sequence differences could affect amplification in any one of three ways. Firstly, DNA sequence polymorphism could exist in one of the priming sites of an '*alu*-pair' such that amplification (under competitive conditions) might not occur. Secondly, continuing retrotransposition of repeat elements in the germline may give rise to presence-absence polymorphisms, in which case amplification would not occur under any condition. Thirdly, it is possible that length polymorphism may occur due to a variability in the numbers of AT-rich di- and tri-nucleotide repeats, particularly at the 3' end of an *alu* repeat. In this case, amplification may well occur in both homologs, but the products of the PD/Ts-1 hybrid may be 'hidden' behind other amplification products because of a differing fragment length. Such length polymorphisms, which give rise to unexpected 'additional' products, have recently been reported (Guzzetta et al., 1991).

It is also possible that additional human material may be present in the MD/Ts-1 hybrid cell that had not been detected by *in situ* hybridisation with total human DNA (Varesco et al., 1989). These possibilities were, however, not addressed in this thesis. The nature of the differences could be investigated by mapping, either on the deletion panels, or perhaps by chromosomal *in situ* hybridisation.

The background smears obtained in the *alu*-LIHs combinations probably reflects a very high degree of template competition (discussed further in

Chapter 9) since the two repeat elements are being used in combination with each other. Interestingly, the large amount of potential and discrete PCR products generated in these combinations clearly demonstrates that there must be significant overlaps in the distributions of both of these repeat elements in the human genome.

Chapter 4.

Mapping of microdissected genomic DNA probes in the *APC* gene-region.

Chapter 4

Mapping of microdissected genomic DNA probes in the *APC* gene-region.

4.1. Introduction.

The experiments described in this chapter were aimed specifically at the isolation of a large number of DNA probes mapping to the *APC* gene region. To this end, initial collaborative investigations were made of two genomic libraries constructed by the physical microdissection and microcloning of DNA from region 5q21-q22 to assess their potential use in the large scale isolation and mapping of clones with respect to the somatic cell hybrid panel outlined in Chapter 3.

The microdissection libraries (hereafter referred to as library I and library II) were constructed as outlined in Chapter 1. The physical dissections and microcloning manipulations were carried out by Drs. Uve Claussen and Gabriele Senger (Institut fur Humangenetik, Erlangen, F.R.G). The PCR amplification and subsequent cloning reactions were carried out by Drs. Hermann-Josef Ludecke and Bernhard Horsthemke (Institut fur Humangenetik, Essen, F.R.G.). The libraries were initially investigated in collaboration with Dr. Wolfgang Balhaussen and Ms. Gabriele Leuteritz (Institut fur Humangenetik, Erlangen, F.R.G). The primary characteristics of each of the two libraries are summarised in table 4.1(a). Four examples of the physical dissections around region 5q21-q22 are shown in figure 4.1(a).

4.2. Results.

4.2.1. Characterisation of Library I.

The strategy followed was to identify single-copy microclone inserts by on Southern blots containing human genomic DNA digested with EcoRI and HindIII (or BglII). Suitable inserts would then be used for sub-regional mapping with respect to the somatic cell hybrid chromosome 5 panel (see Chapter 3).

Analysis of the first 10 clones in this way led to the identification of only one single-copy sequence; mc1, the others being either repetitive or giving no

Table 4.1(a). Characteristics of microdissection libraries.

Library	Library I	Library II
Number of dissections	20	20
Estimated amount of DNA	300fg[a]	250fg
Number of PCR cycles	26	30
Number of recombinant clones	5×10^4	5×10^4
Average size of insert (base pairs)	170[b]	150[c]

Table 4.1(b). Preliminary analysis of microclone inserts from microdissection libraries I and II.

Library	Library I	Library II
Number of inserts analysed	64	78
Inserts exhibiting no signal	7	5
Inserts exhibiting repetitive signals	39	54
Inserts exhibiting single-copy signals	18	19
Number of inserts mapped	10[d]	8
Number on chromosome 5	9	8
Number deleted in MD & PD	2	1
Number deleted in MD only	1	3

a fg; fentograms. Estimated by Drs. U. Claussen and B. Horsthemke.
b estimated from analysis of 64 clones (see table 4.1(b)).
c estimeted from analysis of 78 clones (see table 4.1(b)).

d including microclone insert - mc1.

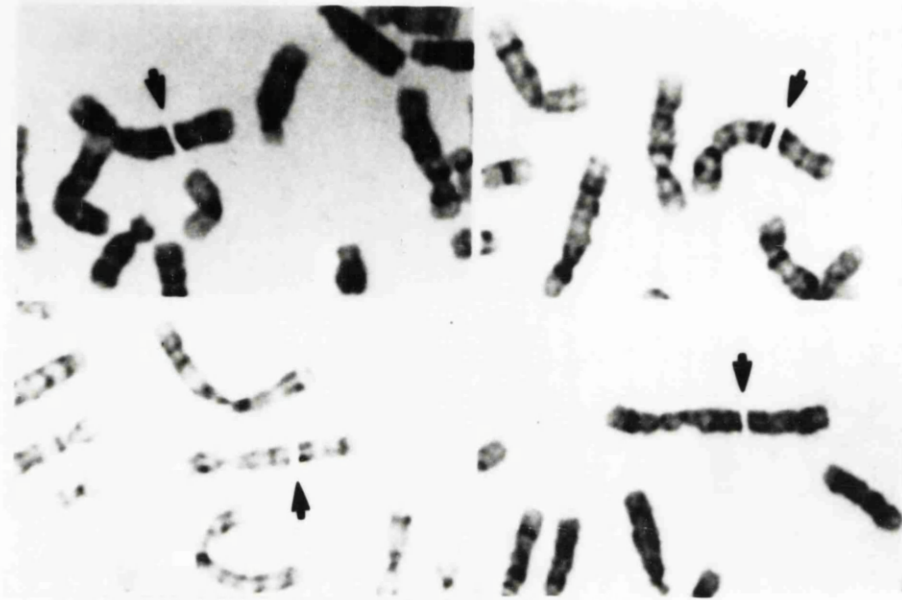
Figure 4.1. Microdissection of chromosome region 5q21.

(A) Four examples of the physical microdissection of chromosomal region 5q21-q22 from giemsa-trypsin-giemsa (GTG)-banded human metaphase spreads. Arrows indicate the position of the dissected chromosomal DNA.

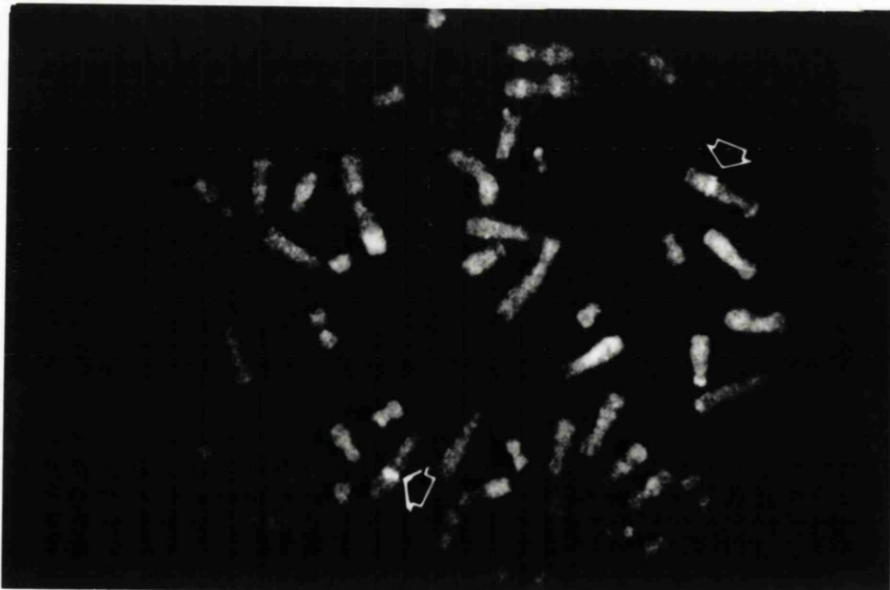
(B) Non-isotopic CISS-hybridisation using microdissection library II as a probe (see text). The chromosome 5q21-specific signals are indicated by arrows.

Photographs were kindly provided by Drs. Udo Trautmann and Wolfgang Balhaussen, Institut fur Humangenetik, Erlangen, F.R.G.

A



B



appreciable hybridisation signals. Hybridisation of the mc1 insert on the mapping panels showed, unequivocally, that it did not originate from chromosome 5. *In situ* filter hybridisations of this insert to a large number of the microclones demonstrated that this clone was represented a number of times in the library (Dr. W. Balhaussen, personal communication). It was decided that a further microdissection library should be constructed on the basis that this microclone may have represented one of a larger number of contaminant sequences. In addition, *in situ* filter hybridisations with a number of other randomly selected microclones suggested a high degree of clone redundancy.

However, during construction of the second library, 18 apparently independent single-copy inserts were identified in library I, ten of which were chosen for further analysis. Initial attempts at sub-regional mapping of these inserts proved difficult and inconclusive under the hybridisation conditions used, primarily because of very poor signal-to-noise ratios (i.e., high levels of non-specific hybridisation signal). The non-specific signal was thought to be due to the inclusion of dextran sulphate in the hybridisation buffer (Buffer I; see Chapter 2), but also perhaps because of insufficient blocking by the salmon-sperm DNA. Thus, the original hybridisation buffer was changed to a simple buffer containing sodium dodecyl sulphate (SDS) in 0.5 molar (M) buffered sodium-phosphate (Church and Gilbert, 1984; see Chapter 2). This buffer significantly reduced the high degree of non-specific background hybridisation signals previously obtained thus allowing extended autoradiographic exposure times for the detection of insert-specific signals.

Further, post-hybridisation washing conditions were also modified by reducing the stringency from 1 x SSC, 0.1% SDS at 65^0C in the original experiments to 2 x SSC, 0.1% SDS at 55^0C, proving ideal for very small sized inserts. In this way, signals for 9 out of 10 of these microclones could be clearly identified on the mapping panels. All of these inserts were found to map to chromosome 5. In addition, two of the inserts, mc14 (figure 4.2(a), panel 2) and mc81 were absent from the DNAs of the two interstitial deletions while one other insert, mc43, was absent from the DNA of MD/Ts-1 only. The overall data for the hybridisation analyses of library I are presented in table 4.1(b). Further characteristics of the three deleted clones are presented in table 4.2.

4.2.2. Characterisation of Library II.

Library II was constructed as described for library I, except that new reagents were used in all steps of the procedure and a new cloning vector - pT3T7-18U was used instead of pUC 8 (Drs. U. Claussen and B. Horsthemke, personal communications). The emphasis of physical dissections in this library was placed on sub-region 5q22 (Dr. U. Claussen, personal communication). Seventy-eight random library II inserts were analysed by hybridisation to human genomic DNA. Of 19 apparently single-copy inserts, 14 were used for sub-regional mapping. Eight of these sequences could be unequivocally mapped on the somatic cell hybrid panels. Of these, one clone - mc5, was absent in both hybrids and another three inserts - mc2 (figure 4.2(a), panel 1); mc320 and mc901 were absent in the MD/Ts-1 hybrid only (see tables 4.1(b) and 4.2).

Examples of regional mapping analyses using microclone inserts from Libraries I and II are shown in figure 4.2(a), panels 1-3.

4.2.3. Chromosomal *in situ* suppression (CISS)-hybridisation.

In separate experiments conducted by Dr. U Trautmann (Institut fur Humangenetik, Erlangen) and Dr. W. Balhaussen, a small amount of the original ligation mixture was amplified and concomitantly labelled with biotin-dUTP by M13-PCR. This labelled mixture was used as a probe on normal human metaphase spreads after suppression of repetitive sequences using Cot-1 human placental DNA (Trautmann et al., 1991). The result of one of these hybridisation experiments is shown in figures 4.1(b). The results are impressive in showing clear and unequivocal hybridisation around subregions 5q21-q22; the approximate regions of the physical dissections.

4.2.4. Conclusions from the initial characterisations of the microdissection libraries.

The primary analysis of library I suggested that contamination had occurred during construction. One of the identified contaminant microclones, mc1, was subsequently analysed by di-deoxy sequencing and found to represent a microclone derived from a library constructed at the Prader-Willi syndrome gene region at 15q11.2-q13 (Dr. B. Horsthemke, personal communication).

Table 4.2. Characteristics of deleted clones from preliminary library analyses.

Probe	Library	Deleted in	Insert size (in bp)	EcoRI signal[a] (in kb)
mc14	I	PD+MD	250	3.8
mc43	I	MD only	210	6.7
mc81	I	PD+MD	180	2.1
mc5	II	PD+MD	220	7.5+1.9[b]
mc2	II	MD only	290	4.2
mc320	II	MD only	230	1.8
mc901	II	MD only	110	6.5

a on genomic DNA

b internal EcoRI site present in this insert; observed with hybridisation of PCR prepared insert DNA

After maximisation of the microclone hybridisation procedures, however, subsequent analysis did show that 9 single-copy microclone inserts from this library mapped to chromosome 5 and that three of these sequences were absent in one or both of the deletion hybrids. Thus, the library constructed did contain sequences mapping to the regions of physical dissection, but, with a major clone contaminant and a relatively high degree of clone redundancy.

The mapping of some of the clones from both of the libraries into just the MD deletion alone confirmed that observation of Varesco et al., (1989) that, on a molecular level, the deletion in MD must extend further distally than that of PD, even though these two hybrids are cytogenetically indistinguishable.

4.2.5. Large scale mapping of inserts from Library II.

Although library I did, in fact, contain sequences useful for regional mapping, it was decided to use library II for a large-scale mapping effort in localising sequences in the *APC* gene-region. To expedite the analysis of a large number of clones from this library, two steps were taken. Firstly, 300 individual colonies were inoculated into bacterial medium (LB-ampicillin; see Chapter 2) in 96-well microtitre dishes and grown to saturation overnight. It was expected that the number of cells per well would be relatively uniform for the rapid preparation of insert DNA as described for the inter-*alu* PCR product libraries (see Chapter 3). The dishes also provided a secure means in which to store the clones for future reference and analysis (see Chapter 2).

Secondly, to increase the rapidity of analysis and the numbers of potentially 'readable' clones, it was decided to screen the PCR amplified inserts *directly* on the mapping panel filters, as opposed to a preliminary screen on human genomic DNA to select for 'single-copy' sequences (see conclusions below).

The results of these large scale mapping efforts are summarised in table 4.3.(a). Of one hundred and twelve individual hybridisations that I carried out in this separate analysis, forty-one inserts could be clearly mapped using the somatic cell hybrids. Of these inserts, 17 were found to be deleted in both PD and MD; 7 clones in MD only and 15 mapping to chromosome 5 but not deleted. Similar analyses of 90 clones were independently performed by Ms. Kathy Howe (Somatic Cell Genetics Laboratory, ICRF) which gave rise to 9 clones in PD and MD; 10 clones in MD only and 15 mapping to chromosome 5 but not deleted.

To check if the inserts that had been mapped within the minimal region defined by the overlap of the MD and PD deletions (containing the *APC* gene), were different from one another, a detailed comparison of the hybridisation patterns and PCR product sizes for each of the inserts was carried out. In total, it was found that 6 clones were redundant (one clone represented 4 times; one clone represented 3 times and 4 clones represented twice), leaving a total of 19 independent clones in this minimal chromosomal interval (see table 4.3(a)). Originally this figure was reported as 22 clones (Hampton et al., 1991) but, more rigorous analysis has since modified these results.

Possible redundancy among microclone inserts mapping into the region defined by the distal end of the MD deletion, or mapping to chromosome 5 outside the region defined by either of the deletions, was not addressed as these clones were not considered useful to the polyposis project. It is likely that an approximately 30-40% redundancy exists in these sets of clones based on the detailed characterisations described above.

It should be noted that about 25% of the microclones described here (including those deleted and non-deleted) showed cross-hybridisation at varying intensities to rodent DNA. An example of a highly conserved microclone - mc408 is shown in figure 4.2(b), panel 2 (see conclusions).

4.2.6. Definition of a new minimal region encompassing the *APC* gene.

Further into these studies, a 25 year old male patient exhibiting polyposis and a chromosome 5 rearrangement was identified (Cross et al., 1992). The rearrangement on chromosome 5 was thought to be complex, involving intra-chromosomal recombination resulting in a deletion around region 5q22-5q23.2. The abnormal allele was segregated in a somatic cell hybrid designated SD/Ts-1 (Thomas, 1991; Hampton et al., submitted). If deletion was the correct interpretation, the proximal breakpoint of this deletion would potentially divide the previous minimal region into two halves further narrowing the position of the polyposis gene. Therefore, all of the sub-regionally localised microclones (that is, mapping to one or both deletions), including those from the original library characterisation studies, were re-mapped with respect to this new hybrid.

Table 4.3. **Summaries of large scale mapping analysis.**

A. Primary analysis.

Number of inserts hybridised	Number that could be mapped	Deleted in: PD+MD	MD	On ch. 5 but not deleted
112[a]	41	19	7	15
90[b]	34	9	10	15
202[c]	75	28(19[d])	17	30

B. Secondary analysis (SD deletion mapping).

Number of inserts hybridised	Deleted in: PD+MD	PD+MD+SD	MD+SD	MD only
36	7[d]	12[d]	0	17

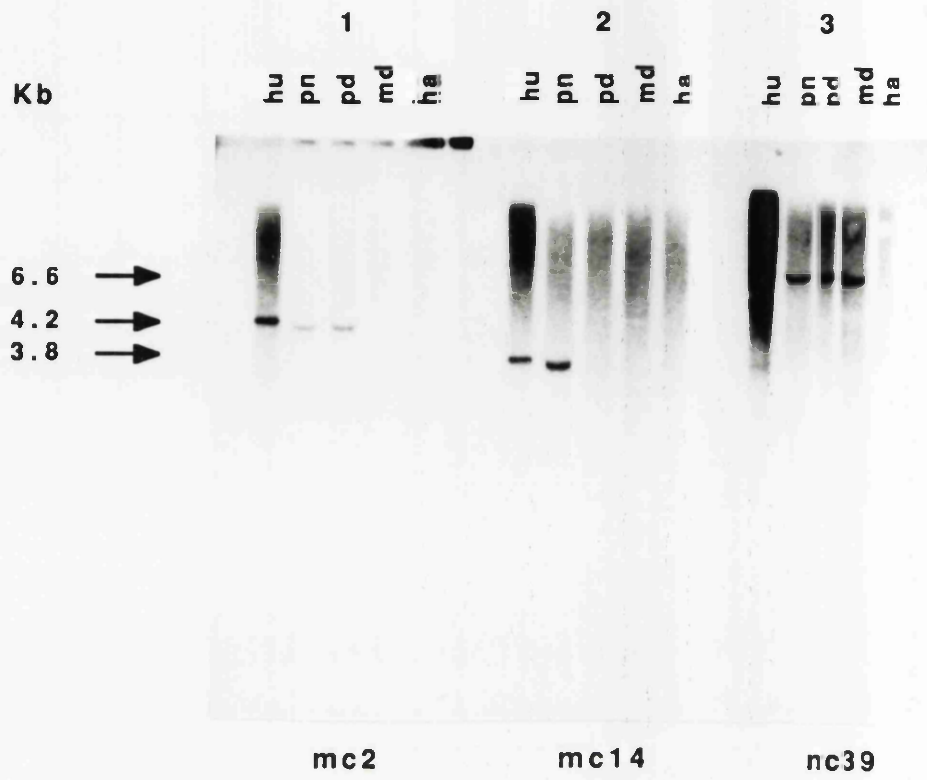
a Mapped by G. Hampton
b Mapped by K. Howe
c Totals
d known to be independent

Figure 4.2 Southern blot analysis of microclone inserts.

(A) Preliminary hybridisation analyses of microclones from libraries I and II on hybrid mapping panels (I, 2, 3) representing the different chromosomal intervals defined by deletions MD and PD (i.e. deleted in MD only (1), in both (2) and in neither (3)). The identity of the EcoRI digested DNAs are indicated above the figure: human genomic DNA (**hu**), PN/Ts-1; normal chromosome 5 (**pn**), PD/Ts-1; 5del(q21-q22) (**pd**), MD/Ts-1; 5del(q21-q22) (**md**) and hamster genomic DNA (**ha**). The acronym of each hybridised microclone is indicated below the panels (refer to **tables 4.1(b)** and **4.2** and to the idiogram in **figure 4.3**). The sizes of the hybridising fragments are shown to the left of the figure. Each of the filters were washed at a final stringency of 2 x SSC/0.1% SDS at 55^{0}C and exposed to autoradiographic film for 2-5 days.

(B) Examples of microclone hybridisations from the large scale analysis of library II mapping into each of intervals I, II and III defined by deletions MD, PD and SD. The identity of the hybrids are indicated above the figure as described in (**A**), except for the addition of the SD/Ts-1 hybrid (5del(q22-q23.2)) (**sd**). The acronym of each hybridised microclone is indicated below the panels (refer to **tables 4.1(b)** and **4.2**, and to the idiogram in **figure 4.3**). Note the strong rodent cross-hybridisation with microclone mc408. The positive signal of mc519 on SD/Ts-1 DNA in panel (3) is less intense upon photographic reproduction. Each of the filters was washed at a final stringency of 2 x SSC/0.1% SDS at 55^{0}C and exposed to autoradiographic film for 2-5 days.

A



B

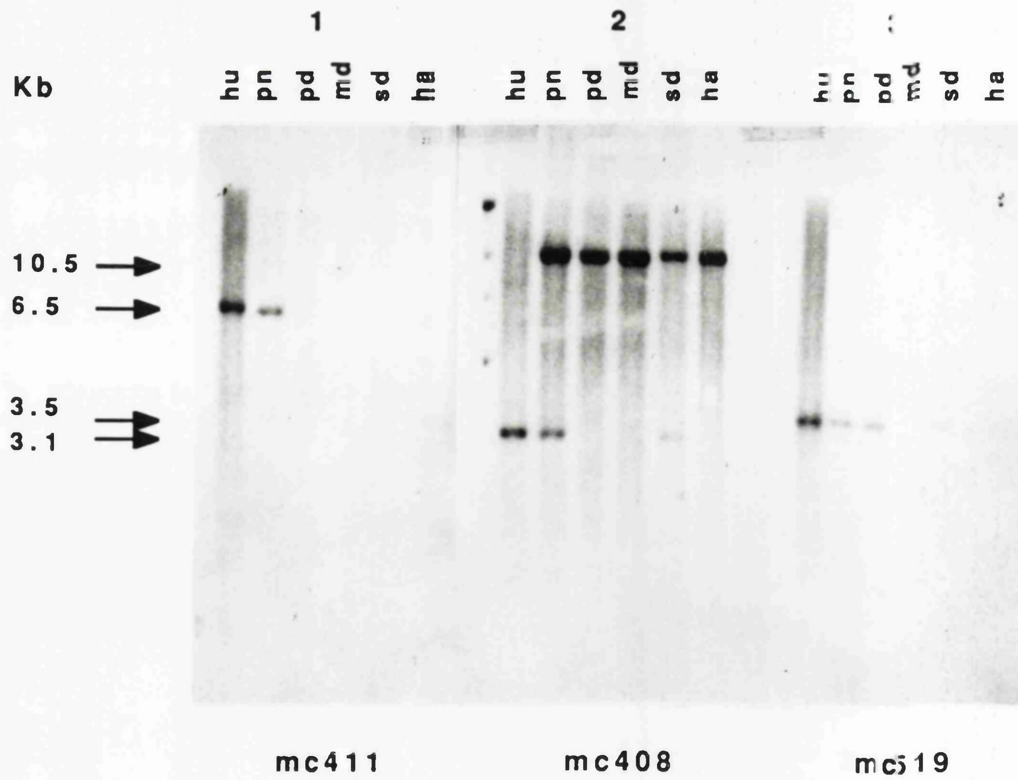
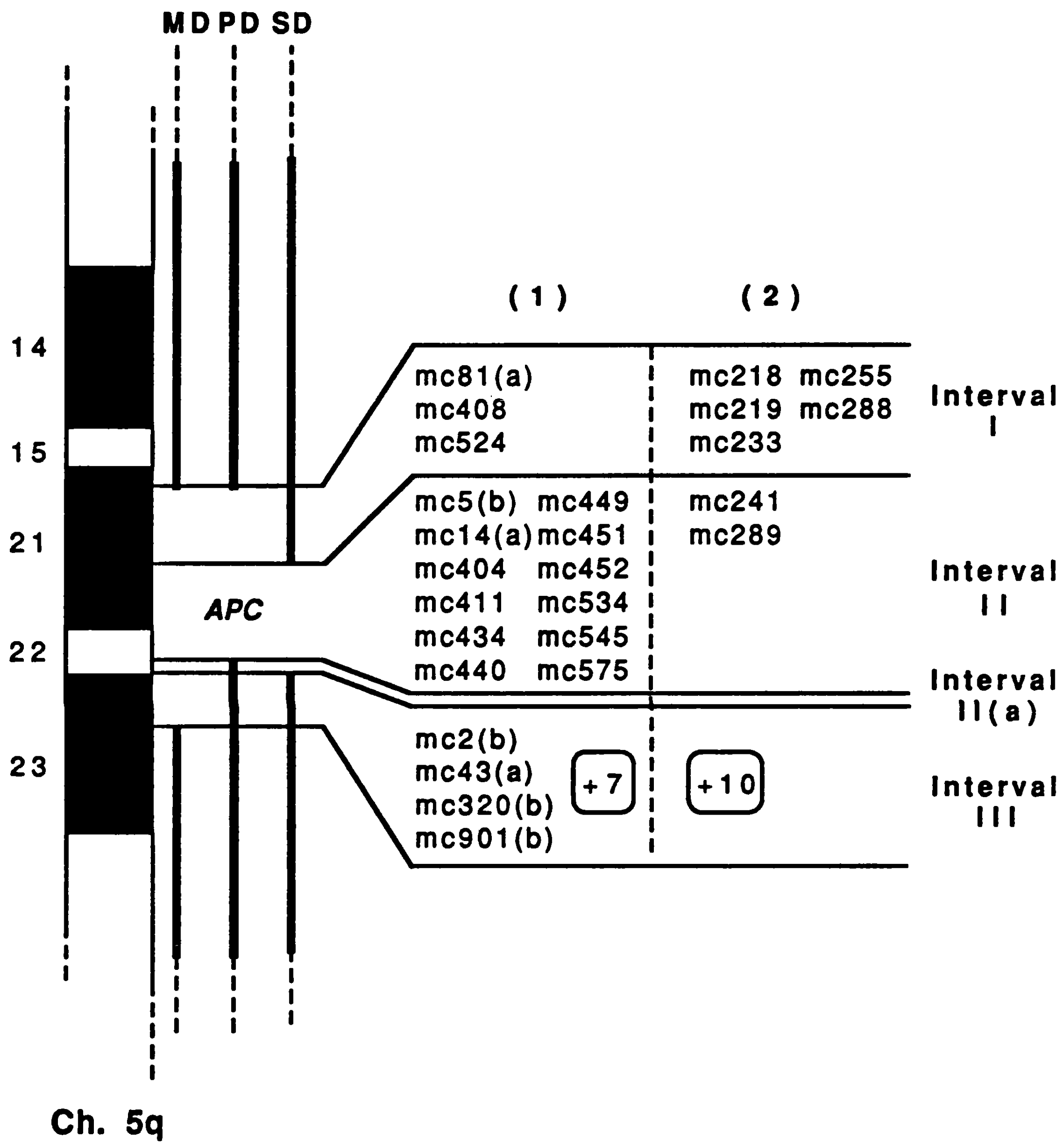


Figure 4.3. Regional localisation of microclone inserts.

The left side of the figure depicts a partial idiogram of chromosome 5q (ISCN, 1985). The deleted chromosomes 5 are depicted by black bars to the right of the idiogram; **MD** = MD/Ts-1, **PD** = PD/Ts-1, **SD** = SD/Ts-1. The relative extents of the deletion breakpoints, which are represented by breaks in the continuous bars, are approximate and based on available cytogenetic data (Williams *et al.*, 1991; Ms. Tania Jones, Human Cytogenetics Laboratory ICRF, personal communication). Notes **(1)** and **(2)** refer to the microclone inserts regionally localised to each of the three major chromosomal intervals by G. Hampton and Ms. K. Howe (Somatic Cell Genetics Laboratory ICRF), respectively. **(a)** and **(b)** denote clones regionally localised during preliminary analyses of microdissection libraries I and II, respectively (Hampton *et al.*, 1991a). **Interval II(a)** has been defined by an anonymous marker, λ12.75 (Ms. S. Cottrell, Cancer Genetics Laboratory, ICRF, unpublished data). None of the microclones were found to map into this region. The numbers enclosed by boxes in **interval III** represent clones absent in MD/Ts-1 DNA only which were not assessed for independence with respect to one-another.


MD PD SD
(1)
(2)
14
15
21
22
23
APC
mc81(a)
mc408
mc524
mc218 mc255
mc219 mc288
mc233
Interval
I
mc5(b) mc449
mc14(a) mc451
mc404 mc452
mc411 mc534
mc434 mc545
mc440 mc575
mc241
mc289
Interval
II
Interval
II(a)
mc2(b)
mc43(a)
mc320(b)
mc901(b)
+7
+10
Interval
III
Ch. 5q

The microclones were all re-isolated from the permanent stocks, amplified and then mapped to allow, in addition to this study, a confirmation of the initial mapping work. The hypothesis of a deletion was confirmed and the mapping of the microclones enabled three new intervals (I, II, III) to be defined, including a new minimal region (interval II) around the *APC* gene (refer to the idiogram in figure 4.3). Twelve of the clones from the large scale mapping studies and 2 clones from the initial analyses were found to map into this new minimal region (see table 4.3(b)).

Examples of microclones mapping into each of these three defined intervals are shown in figure 4.2(b), panels 1-3. The idiogram in figure 4.3 schematically illustrates the regional mapping of all of the microclones (from libraries I and II) to the first and second of these newly defined intervals.

4.3. Conclusions.

The use of microdissection libraries, in combination with a somatic cell hybrid mapping panel, led to successful identification of 14 independent clones mapping to the minimal region around the *APC* gene (that is, interval II; figure 4.3). The use of such libraries was found to be difficult because of the generally small size of the inserts (170bp to 150bp for library I and II, respectively). Successful large scale mapping of microclones was probably dependent on a number of factors.

Firstly, effective mapping of the microclones was dependent on the optimisation of the hybridisation conditions. The simple SDS-sodium phosphate buffer (Church and Gilbert, 1984) proved to be ideal in that it could effectively block non-specific hybridisation backgrounds allowing for the prelonged exposure of the filters and, therefore, detection of the insert-specific signals. In general, the mapping panel filters required exposures of between 2 and 7 days. In addition, low stringency post-hybridisation filter washes ensured that these single-copy signals could be detected in the first instance. Previous experience had shown that even a small decrease of salt concentration to 1 x SSC (at 65^0C) would, in some cases, result in a lack of signal detection.

The second important factor was that direct hybridisation of the inserts across a mapping panel (as opposed to prior hybridisation to genomic DNA filters) allowed the identification of a greater number of 'readable' clones from

library II than was suggested from the initial analysis (37% versus 26%). This is simply due to the fact that the single-copy signals can still be read on the somatic cell hybrids even with a repetitive background on human DNA because of the reduced human DNA content of the hybrids (for example see figure 4.2(a), panel 3). Few if any of the inserts exhibited very highly repetitive signals, in line with other published reports (see Ludecke et al., 1990)

It has been argued that more efficient methods of labelling, such as incorporation of [α^{32}P]-dCTP during PCR (Hirst et al., 1991) or the co-labelling of clones with both [α^{32}P]-dCTP and [α^{32}P]-dATP (Buiting et al., 1990; Davis et al., 1990), would provide better hybridisations due to a potentially higher specific activity. In this study such methods were not investigated as the specific activity generated was usually in the order of 10^9 counts per minute (cpm) per μg of DNA. The PCR 'tails' generated from the plasmid multiple cloning site (MCS) region at the ends of the inserts may have contributed to the positions available for [α^{32}P]-dCTP incorporation during random primed labelling. However, it is likely that for much smaller inserts (for example less than 100bp), inclusion of the radionucleotide(s) in the PCR reaction would certainly lead to better overall labelling. In general, such small inserts were not amenable to mapping in the studies described in this thesis.

The observation of cross-species homology in a relatively large proportion of microclones is intriguing (for example see figure 4.2(b), panel 2). Approximately 25% of the microclones described in this study showed some degree of sequence conservation in rodent DNA. This has been reported a number of times in similar microclone libraries (MacKinnon et al., 1991; Hirst et al., 1990; Fiedler et al., 1991; Buiting et al., 1990) and, in the case of a library constructed from chromosome 15q11.2-q13, some 80% of the clones were found to be conserved across a variety of species (Buiting et al., 1990). This feature of such microclone libraries is discussed further in Chapter 9.

A crude estimate of about 5 megabases (Mb) was made at the time for the minimal region (that is, Interval II) around *APC*, based largely on available cytogenetic data, including *in situ* hybridisation analyses (see Williams et al., 1991), and minimal distances in the region spanned by large pulsed field gel fragments (ICRF Polyposis group meetings). Assuming a random distribution of these clones, it was calculated that the region had been saturated at a density of about one landmark clone per 230kb. This was

thought to be a reasonable density for the isolation of yeast artificial chromosomes (YACs) and the construction of a detailed physical map around the *APC* gene region. Isolation of yeast artificial chromosomes (YACs) with some of these microclones, in addition to other regionally localised DNA probes, is the subject of the Chapter 5.

Chapter 5.

Isolation of yeast artificial chromosomes (YACs).

Chapter 5

Isolation of yeast artificial chromosomes (YACs).

5.1. Introduction.

The single DNA probe isolated by IRS-PCR (see Chapter 3, part II) and the 22 independent probes derived by sub-regional microdissection (see Chapter 4) were thought to represent a reasonable saturation of the minimal region around the *APC* gene, originally defined by the proximal and distal PD deletion breakpoints (that is, intervals I and II, figure 4.3; Chapter 4). It was at first thought that this minimal region was of the order of about 8Mb of DNA and, assuming a random distribution of the 23 DNA probes, implied a coverage of approximately one probe per 350kb. As described in Chapter 4, section 4.3.1, this minimal region was re-defined by the characterisation of the SD deletion giving rise to estimates of approximately one landmark DNA probe every 330kb for 15 probes in a 5Mb region.

By analogy with other positional cloning strategies pursued for the cloning of the Retinoblastoma, Duchenne Muscular Dystrophy and Cystic Fibrosis genes (reviewed in Goodrich and Lee, 1990; Monaco et al., 1986; Koenig et al., 1987; Rommens et al., 1990), it was expected that, in an initial strategy, genomic 'walking' from key landmark probes expanded as large pieces of cloned DNA in cosmids or bacteriophage vectors would serve as the most effective approach toward cloning the *APC* gene. At the time of the studies described in this chapter, large pieces of genomic DNA cloned as yeast artificial chromosomes (YACs) (Burke et al., 1987) had been shown to be effective for the coverage of *megabase* regions of complex genomes (Coulson et al., 1988; Silverman et al., 1989) and for the isolation of genes using novel approaches (Elvin et al., 1990). Such YAC libraries had been constructed at ICRF (Dr. Anthony Monaco, Genome Analysis Laboratory; presently at Human Molecular Genetics Laboratory, ICRF, Oxford) and were available to independent ICRF investigators.

On the basis of the dramatically increased capacity of YACs to faithfully propagate cloned genomic DNA (in the order of 5 to 10 times that in cosmid vectors (Larin et al., 1991)), I decided to select as large a number of YACs as possible in the expectation that a YAC contig could be constructed over most,

if not all of the minimal region around the *APC* gene. It was anticipated that other strategies, such as allele-loss in sporadic colorectal tumours or the identification of relatively small deletions or constitutional rearrangements on chromosome 5 in FAP patients with the same set of probes, would then indicate the YAC or YACs on which to concentrate most effort (see Chapter 8).

The use of microclones to select YACs by hybridisation was expected to present difficulties because of the generally small insert size and poor hybridisation qualities (see Chapter 4). Although selection of YACs by PCR-based approaches had been demonstrated to be highly efficient and largely error-free (Green and Olson, 1990a; Green and Olson, 1990b), only hybridisation filters of the ICRF human YAC library were available and pools of YAC clone DNAs had not yet been constructed. However, this hybridisation method, if successful, would be much more efficient in screening large numbers of YACs with relatively less effort. The ICRF human YAC library used in these studies is described in Chapter 2 and in Larin et al., (1991). DNA from 9216 YAC clones were spotted in high density gridded arrays on 22 x 22cm nylon membranes (Mark Ross, Genome Analysis Laboratory, ICRF), essentially as described for the chromosome 5-specific cosmid library in Chapter 3, part II.

Before initiation of these studies, a strategy for efficient YAC analysis was devised. Due to the unambiguous grid co-ordinates inherent in the ICRF YAC filters, it was thought that probes common to single YACs could be identified simply by primary hybridisation obviating the need to isolate the same YAC more than once. Moreover, a graph of the numbers of probes used, versus the number of YACs identified, would be an indication of YAC 'contig' formation. After a large number of hybridisations, such a curve was expected to become flat implying that further hybridisations would add little or no new information. The success of this approach would be entirely dependent on the successful hybridisation of the microclones and on the assumption of a random, or at least sufficiently dispersed probe distribution.

During these studies a candidate gene for *APC* mapping in interval II; *MCC* (for Mutated in Colorectal Carcinoma) was identified by Vogelstein and colleagues on the basis of allele-loss studies in sporadic colorectal tumours (Kinzler et al., 1991a). Although a small number of mutations were identified in some of these sporadic tumour DNAs, none had been reported in FAP kindreds at this time (discussed in Chapters 8 and 9). At the time of this

publication, it was decided to select YACs for the *MCC* gene, as well as for microclone inserts, for two principal reasons. Firstly, if the *MCC* gene did, in fact, represent the *APC* suppressor gene, then YACs would provide an ideal resource to study this gene in a biological context by introduction of the YAC or YACs into colorectal tumour cell lines (Huxley and Gnirke, 1991) in preference to over-promoted cDNA expression constructs. If, on the other hand, *MCC* did not represent the *APC* gene, such YACs would themselves, or as part of an expanded contig, be useful in identifying genes in the immediate genomic region (see Chapter 8). The isolation of YACs for *MCC* is described in this chapter as part of the contig construction strategies pursued in general.

It should be noted that at the outset of the YAC isolation experiments, microclone inserts identified in the large scale mapping effort (see section 4.3, Chapter 4) had not been localised with respect to the SD deletion. Thus, microclones chosen for YAC studies described here included one clone, mc524, that maps to Interval I (figure 4.3, Chapter 4), that is, outside the newly defined minimal region.

5.2. <u>Results 1.</u> Isolation of yeast artificial chromosomes (YACs).

5.2.1. A feasibility study for YAC isolation by microclone hybridisation.

To study the feasibility of selecting YACs from the gridded DNA filters by microclone hybridisation, insert mc14 (see figure 4.3., Chapter 4) was chosen, principally because this sequence exhibited perhaps the strongest 'single-copy' hybridisation characteristics of the entire microclone set. Initially, this probe was hybridised to duplicate YAC filters and control mapping panels as for the original hybridisation experiments described in Chapter 4 (that is, labelled with 30μCi [α-^{32}P]dCTP, with post-hybridisation washes using 2 x SCC, 0.1% SDS at 55^{0}C). Under these conditions, no positive signals were detected on the duplicate YAC filters. Also, signals on the control mapping panels were weak and ambiguous even after 3- to 4-day autoradiographic exposures. In an attempt to increase the signal-to-noise ratios, mc14 DNA was then labelled with 50μCi of [α-^{32}P]dCTP and the filters washed extensively at slightly lower stringencies (that is, 2 x SSC/0.1% SDS at 30 to 40^{0}C; see Chapter 2). This gave rise to two convincing positive signals on the YAC filters and a clear result on the mapping panels after extended 5-day autoradiographic exposures (see table 5.1).

5.2.2. Isolation of additional YACs by microclone hybridisation.

After the initial feasibility experiment (see above), further microclones were chosen on the basis of hybridisation qualities relative to mc14 (typically with the largest insert sizes). Of a pool of 15 sequences that I had localised into both the PD and MD deletions (that is, Intervals I and II), only 8 of these were deemed suitable for YAC hybridisation assays (see table 5.1). Hybridisation of microclone inserts by Ms. Kathy Howe resulted in a pool of 7 additional sequences mapped into PD and MD. After re-localisation with respect to the SD deletion, two of these independent probes - mc241 and mc289, were found to lie in Interval II and both appeared to be suitable for YAC hybridisation assays. This provided a pool of 9 microclones to continue the YAC isolation studies.

Using the modified conditions worked out for mc14 (see above), these 9 microclones best suited to YAC hybridisation assays (detailed in table 5.1) were successively screened, usually two to three times each on the duplicate YAC filters. Only three of these clones - mc524, mc534 and mc575, in addition to mc14, gave rise to positive YAC signals (see table 5.1). Figure 5.1 shows an example of the signals identified by probe mc534.

5.2.3. Conclusions from the microclone hybridisation experiments.

In total, only 4 out of the 10 microclones used gave rise to positive YAC signals (that is, 40%; see table 5.1), even with careful modification of hybridisation conditions. Some of the microclones, particularly mc411 and mc449, were judged to be unequivocally negative on these filters with reference to the positive control mapping filters. It is not known whether sequencing of the inserts and subsequent PCR selection of YACs for mc411, mc449 or any of the other weakly hybridising clones would have provided a better assay system. This is discussed further in Chapter 9.

In line with the observations made during the YAC hybridisation experiments, the use of these clones in the construction of genomic physical maps by PFGE analysis was similarly limited to those microclones exhibiting strong single-copy hybridisation characteristics (my own studies, Ms. S. Cottrell, Mr. T. Ward, personal communications). Thus, for the construction of detailed physical maps (and for the further isolation of additional YACs),

Table 5.1. **Identification of YACs by probe hybridisation.**

(a) Microclones	Numbers of YAC signals
mc5	0
mc14	2
mc411	0
mc449	0
mc452	0
mc524	2
mc534	3
mc575	1
mc241	0
mc289	0
sub-total	8

(b) Other probes	Numbers of YAC signals
cL5.4.s-1	1
MCC ('a'+'b')	3[a]
sub-totals	4
Totals	**12**

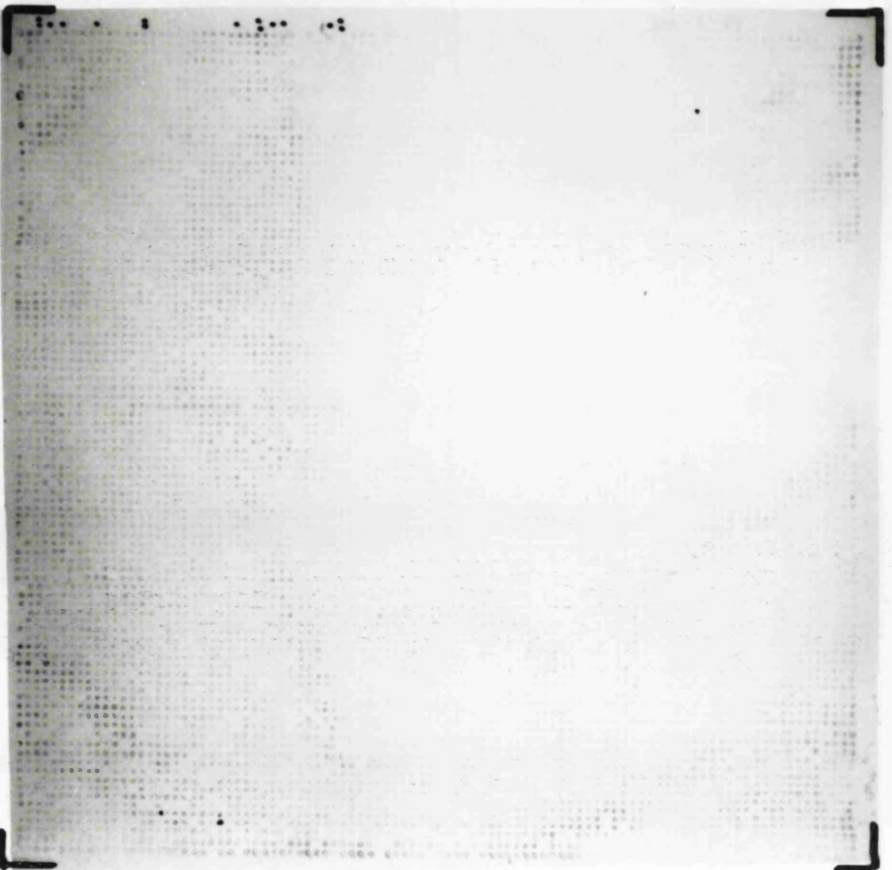
[a] One of these YACs was identified both by mc575 and probes from the *MCC* gene (see also table 5.2).

Figure 5.1. **Identification of positive YAC signals with probe mc534**

(A) A 22 x 22cm nylon filter containing DNA samples from 9216 YAC clones screened with [α^{32}P]-dCTP labelled pBR322 plasmid. The variability in signal is due to (1) non-equalised loading of YAC clone DNA and, (2) non-equal distribution of labelled probe.

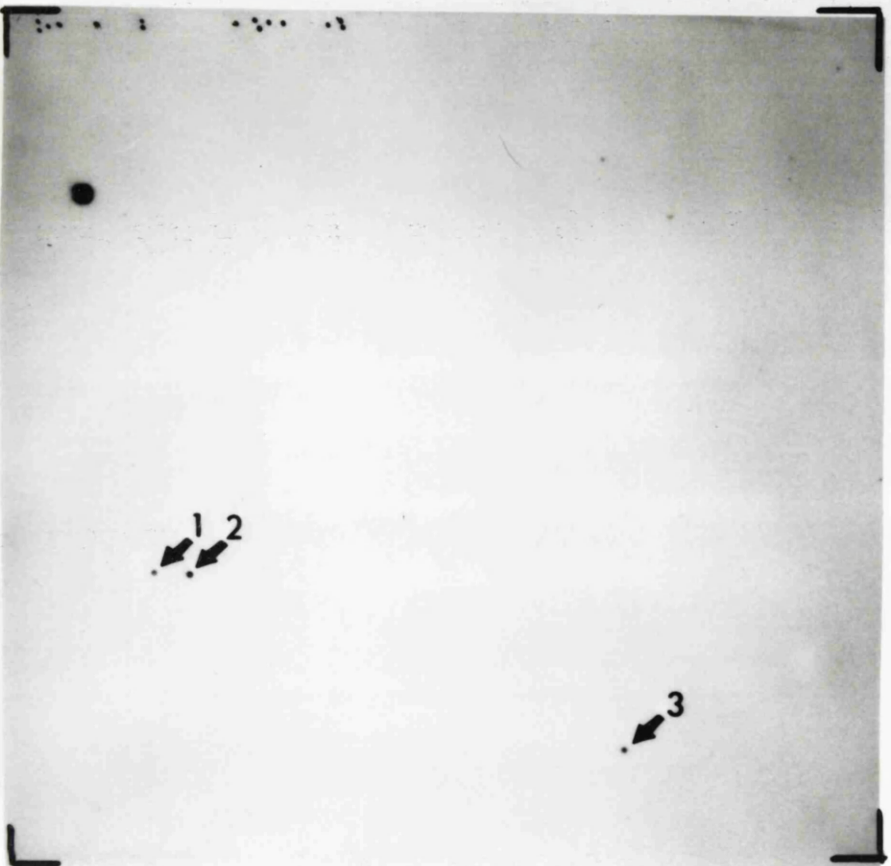
(B) The same filter stripped and hybridised with 50µCi [α^{32}P]-dCTP labelled probe mc534. The filter was washed to a final stringency of 2 x SSC/0.1% SDS at approximately 40^{0}C and exposed to autoradiographic film for 6 days. The three positive signals (see **tables 5.1** and **5.2**) are denoted by arrows. Positives **1, 2** and **3** are YACs ICRFy900A0819, ICRFy900A0919 and ICRFy900D073, respectively. The grid co-ordinates of the positive signals were calculated by aligning the filters at the robot marks indicated in the top left hand side of the autoradiographs.

A



pBR322

B



mc534

attempts were made to isolate cosmids for these microclones by other members of the polyposis group (Mr. T. Ward and Ms. K. Howe).

5.2.4. Isolation of a YAC for probe cL5.4.s-1.

Hybridisation with the 1.1kb BglII L5.4 cosmid subclone; cL5.4.s-1 (see Chapter 3, part II) under the modified hybridisation conditions resulted in the identification of one potential positive (see table 5.1).

5.2.5. Isolation of YACs for the *MCC* gene.

To effect isolation of YACs for the *MCC* gene, oligonucleotide primers were designed to the boundary sequences of two intron-exon junctions essentially representing both the 3' and the 5' ends of the *MCC* gene (Kinzler et al., 1991; see Chapter 2, appendix 1). Using these primer sets, PCR amplification of human genomic DNA generated two products of approximately 250bp in size (referred to here as products 'a' and 'b' containing exon nucleotides 391-533 and 1679-1862, respectively). Amplification of somatic cell hybrid DNAs with these two primer sets demonstrated that both exons mapped to interval II, as expected. Both of these PCR products were found to be quite repetitive when used as probes on the somatic cell hybrid mapping panels but, under the appropriate competitive conditions, could be successfully used on Southern blots. Thus, these two exon-containing probes were labelled and hybridised as a mixture to the YAC filters. This resulted in the identification of 3 positive YAC signals (see table 5.1).

5.2.6. Probes coincident to the same YAC(s).

During the YAC hybridisation experiments, grid co-ordinates were constantly cross-referenced on the ICRF genome data bases (Dr. Gunther Zehetner, Ms. Chrystal Douglas, Genome Analysis Laboratories, ICRF), as well as by inspection of the primary hybridisation filters. Two cases of probes coincident to the same YAC(s) were identified. In the first case, inspection of the primary filters revealed that probe mc575 was coincident to one of the YACs identified by sequences from the *MCC* gene (see tables 5.1 and 5.2). This sequence, in fact, lies some 100kb distant to the 5' end of *MCC* gene (see Chapters 8). In the second case, it was found that an anonymous probe, termed p3.1 (Spurr et al., 1991; screened on separate sets of YAC filters by Mr. T. Ward), recognised 2 of the 3 positive signals identified by microclone mc534 (see tables 5.2 and 5.3).

This clearly demonstrates the power of reference data bases when using unambiguous grids on the YAC filters. To some extent these results supported the multi-hybridisation approach outlined in the introduction to this chapter. In addition, this data provided valuable information for the ascertainment of small contigs around the *MCC* gene and around probes mc534 and p3.1. The use of coincident YAC mapping is expanded upon further in Chapter 6.

5.2.7. Summary of candidate YAC clone detection.

Summaries of the hybridisation assay experiments are provided in table 5.1. In total 13 probes were used in such hybridisation assays. Of these, 6 were successful in identifying a total of 12 positive signals. Because of the coincident detection of one of the *MCC* YACs by probe mc575, this left a total of 11 different YACs to investigate further.

5.3. Results 2. Assessment of candidate YAC clones.

In each of these 11 cases, the YAC clones were investigated by a series of four tests to assess whether they represented true positives. The results of these experiments are summarised in table 5.2. at the end of this chapter.

5.3.1. Secondary screening.

Yeast cells from the microtitre wells representing the positive grid co-ordinates were streaked out on selective yeast minimal media lacking uracil to obtain single colonies (Chapter 2). After selective growth of the yeast, 5 or more single colonies were re-streaked. These yeast colonies, in addition to AB1380 control yeast cells (see Chapter 2), were transferred to nylon membranes and hybridised with the original probes at high stringency. The secondary screening procedure proved to be essential since some of the original microtitre-well stocks appeared to be 'mixed', that is with additional YAC-containing yeasts. This was observed in 40% of cases. In only 1 out of 11 cases (identified by probe mc524) did a potentially positive YAC turn out to be negative (table 5.2).

5.3.2. Analysis by pulsed field gel electrophoresis (PFGE).

Individual colonies from each of the YAC-containing yeasts, apparently positive by secondary screening, were used to make large-scale yeast DNA preparations in agarose 'plugs' (see Chapter 2). These 'plugs' were subjected to PFGE to separate the yeast chromosomes in the 100kb to 1Mb size range. One or more yeast DNA plugs containing different YACs were always included as negative controls in these experiments. Filters of the pulsed field gels were hybridised with the test probe to determine the size, as well as positivity of the YACs. These filters were then stripped and probed with radiolabelled pYAC4 vector to check for the presence of other unrelated YACs resulting from the co-transformation of the yeast hosts during library construction. Co-transformation was found in 2 of 10 cases (table 5.2). Figure 5.2 (a) shows an example of the positive identification and sizing of YACs detected by microclones mc14 and mc534.

5.3.3. Comparative human and yeast Southern blot hybridisation.

To assess whether the YACs were positive by virtue of the sequences in the test probe in question, and not by virtue of some cross-sequence homologies, whole yeast and human genomic DNAs were digested in parallel with HindIII and EcoRI. Total human genomic DNA was used in 200-fold excess to compensate for the difference in yeast and human genome sizes. Hybridisation of Southern blots with the test probe showed, in each case, the same sized hybridisation fragments in yeast and human genomic DNAs. This test, in addition, allowed an assessment of the fidelity of YAC cloning in each of the candidate YACs. Specifically, the identification of the similarly sized restriction fragments in both the YAC-containing yeast DNA and human genomic DNA implies that the cloned sequences are in the same relative 'form'. Altered sized fragments in the yeast would imply potential rearrangement (although it may also imply that the test sequences reside in the terminal restriction fragment). An example of this type of experiment on the YAC identified by probe cL5.4.s-1. is shown in figure 5.2(b).

5.3.4. Fluorescence *In situ* hybridisation (FISH) of total YAC-containing yeast DNAs.

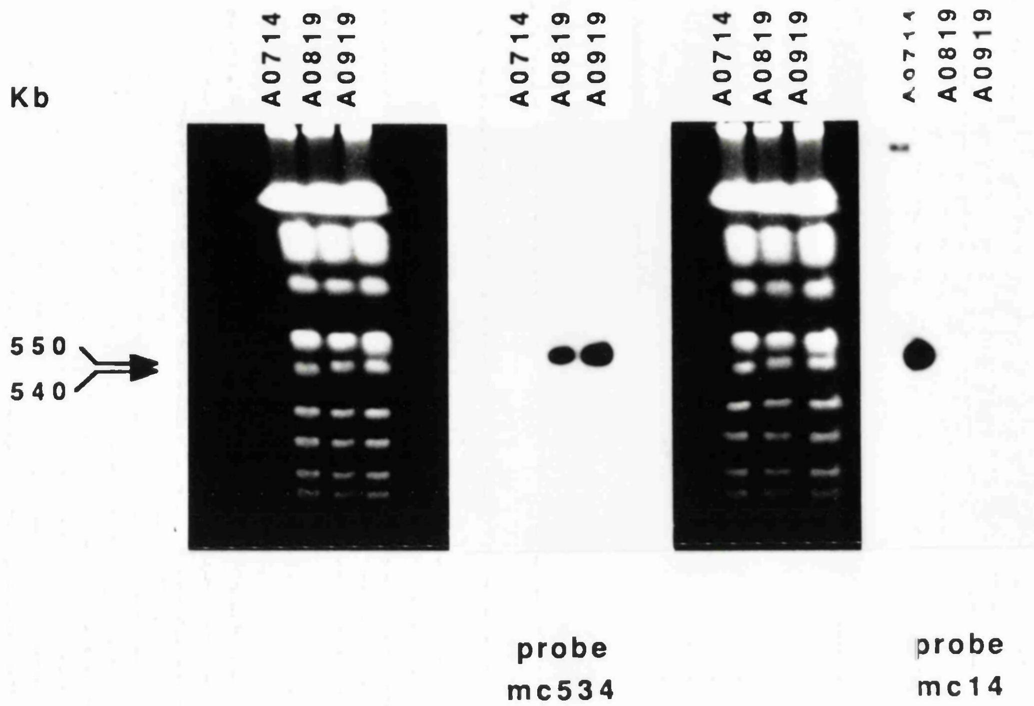
FISH experiments were carried out by Ms. Tania Jones (Cytogenetics Laboratory, ICRF). For these experiments, whole yeast DNAs from the agarose plugs were purified, labelled with biotin-dUTP and hybridised to normal human lymphoblastoid metaphase spreads with or without signal

Figure 5.2. Pulsed field Gel electrophoresis (PFGE) and comparative Southern blotting of potentially positive YAC clones.

(A) Left hand side: PFGE of YAC-containing yeasts. The identity of each of the YAC-containing yeasts are indicated above the figure. Each lane represents one half-block (40µl) electrophoresed in a 1% gel at 5V/cm for 20 hours with a linear switching time of 20 to 84 seconds. The YACs (**540kb**, **550kb** and **550kb**, respectively) can be seen by eye on the gel and are indicated to the left of the figure. Right hand side: Southern blot of the PFG, hybridised with probe **mc534**. The other half of the Southern blot filter was hybridised with probe **mc14**. Southern blot filters were washed to a final stringency of 0.2 x SSC/0.1% SDS at 65^{0}C and exposed to autoradiographic film overnight.

(B) Left hand side: Electrophoresis of EcoRI and HindIII digested human and yeast (ICRFy900A10109) genomic DNAs on a 0.8% agarose gel. The gel was electrophoresed at 2.5V/cm for 48 hours and stained with ethidium bromide. Right hand side: Southern blot of the agarose gel probed with **cL5.4.s-1**. The sizes of the hybridising fragments are indicated to the right of the figure.

A



B

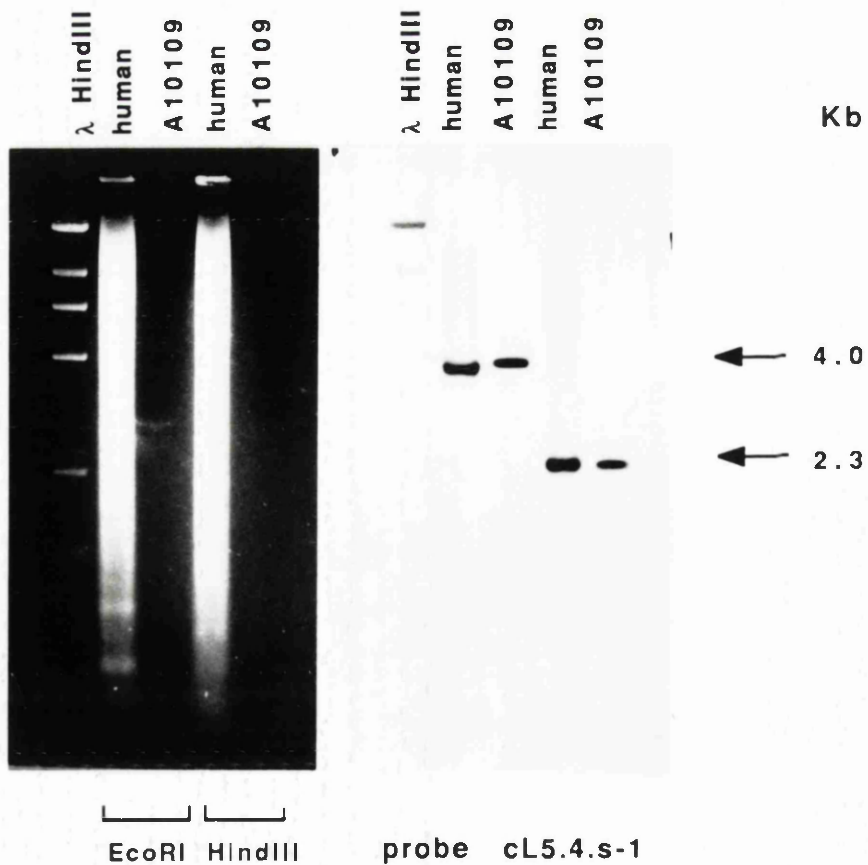
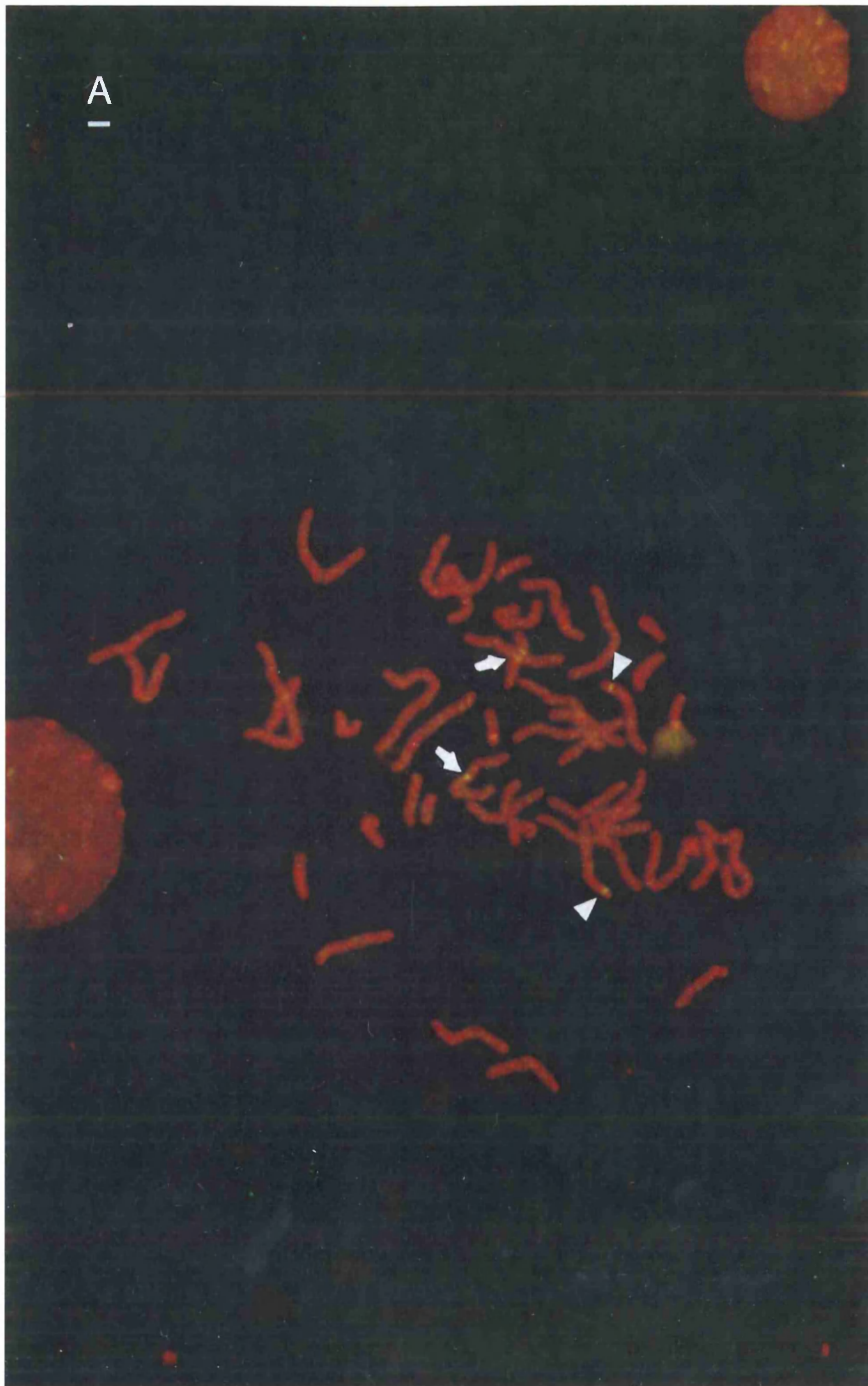
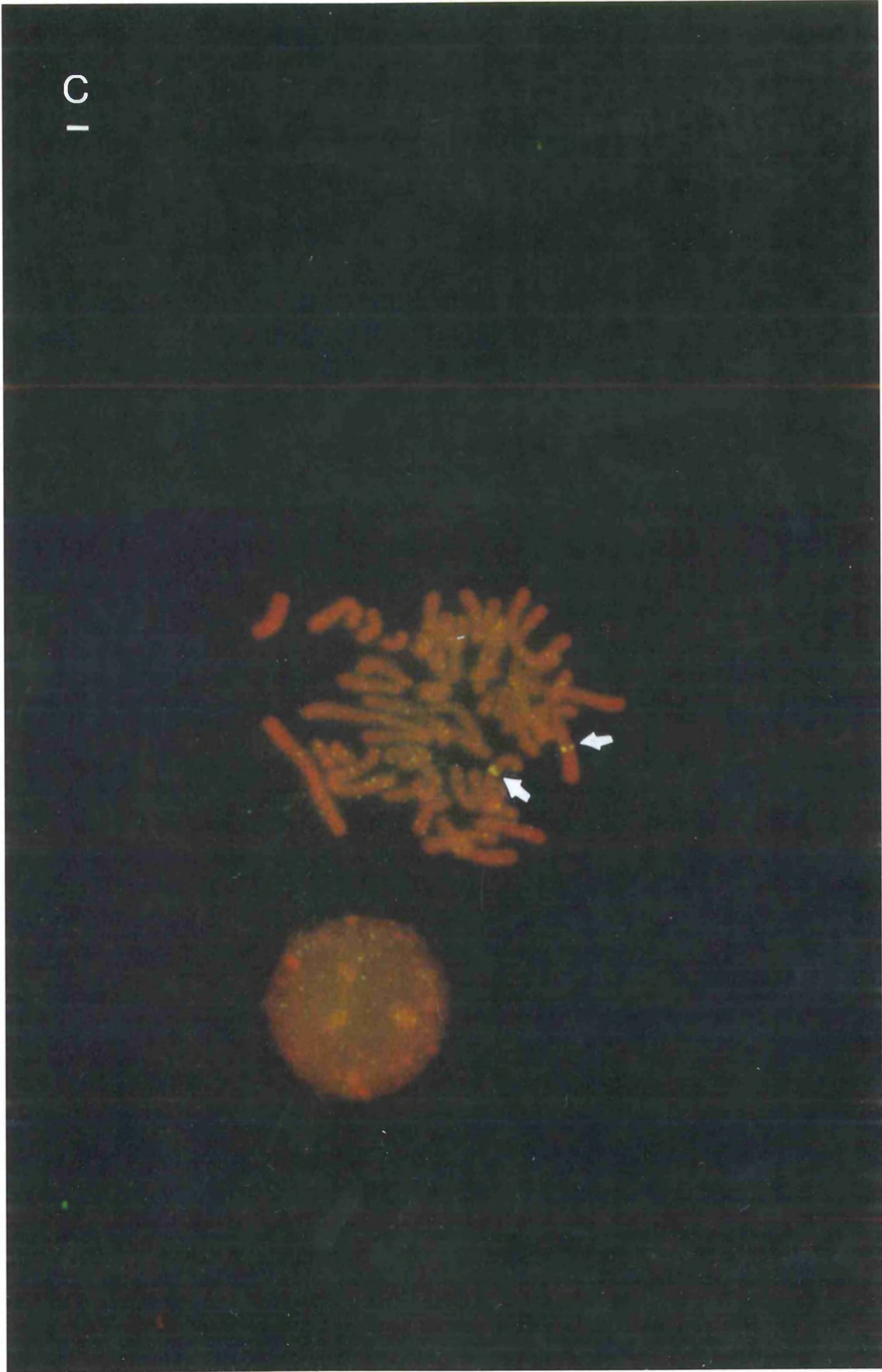


Figure 5.3. Fluorescence *in situ* hybridisation (FISH) of total yeast DNAs on human chromosome metaphase spreads.

Three examples of non-isotopic YAC FISH analysis demonstrating **(A)** chimeric (B0879), **(B)** co-transformed (B0624) and **(C)**, non-chimeric (A1010) YACs. Arrows indicate chromosome 5q21-specific signals; arrowheads indicate non-chromosome 5 signals on chromosome 8p and a C-group chromosome in **(A)** and **(B)**, respectively. FISH analysis was carried out using total yeast DNA suppressed with human genomic DNA essentially as described in Williams *et al.*, (1991). Photographs were kindly provided by Ms. Tania Jones, Human Cytogenetics Laboratory, ICRF.


A

B


C

amplification. All of the YACs, except one (that identified by probe cL5.4.s-1) were shown to contain sequences originating from the 5q21-q22 chromosomal region. The cL5.4.s-1 YAC (see table 5.2), gave rise to highly dispersed and inconsistent signals; perhaps due to a high content of repetitive sequences. In 6 of the other 8 cases, the most consistent signals were only observed at 5q21-q22. In two cases however, consistent signals were observed in regions of the genome, in addition to 5q21-q22. From one of the YACs identified by the mc534 probe; ICRFy900D073 (table 5.2), signals were consistently observed at the telomeric regions of chromosomes 5 and 1. These signals were quite intense, in fact more so than the signals in the 5q21-q22 region. In one other YAC isolated by the mixed *MCC* 'exon-containing' probe; ICRFy900B0879, signals of equal intensity were observed on 8p, in addition to 5q21-q22. Examples of the FISH analyses demonstrating 5q21-q22-only and chimeric YACs, in addition to a co-transformed yeast are shown in figure 5.3.

5.4. Conclusions.

The experiments described above confirmed that 10 different cognate YACs had been identified with 6 probes (see table 5.2). In the polyposis project as a whole, a further 9 YACs were identified with an additional 4 loci sub-localised to Interval II. Some of the characteristics of these YACs are detailed in table 5.3. Taking into account the coincident detection of YACs by probes mc534 and p3.1, a total number of 17 YACs had been isolated in Interval II.

YAC-FISH experiments proved to be a powerful method for visualising the genomic origin of most of the YACs studied. Potentially chimeric YACs were identified in two cases. This had been suspected at least for the larger of the two YACs (ICRFy900B0879) simply on the basis of its large size (table 5.2). YACs of this size are extremely rare in the ICRF library (Larin et al., 1991) and the YAC was thought to result from the co-cloning of two 600kb average sized EcoRI partial DNA fragments. In combination with comparative Southern blot analysis of restriction fragments in YAC and human DNA, the FISH procedure adds a very high degree of confidence to the appropriate assignment and fidelity assessment of YACs. However, it should be noted that the FISH technique would be ineffectual for the identification of YACs containing small fragments from other regions of the genome. The only unequivocal method for the rigourous investigation of the YACs in this sense, would be the isolation of YAC insert termini which is the subject of Chapter 7.

Table 5.2. Characteristics of YACs isolated in this study.

Probe hybridised	No. of positives primary screen	No. of positives second screen	YAC designation	Size (kb)	FISH results[a]	Comments
mc14	2[b]	2[c]	ICRFy900G016	420	5q21	
			ICRFy900A0714	550	5q21	
mc524	2	1	ICRFy900A1242	580	5q21	+ YAC @ 440Kb
mc534	3	3	ICRFy900A0819	550	5q21	
			ICRFy900A0919	550	5q21	
			ICRFy900D073	680	5q21 + 1/5qter	
mc575	1	1	ICRFy900B0624	620	5q21	+ YAC @ 440kb[d]
cL5.4.s-1	1	1	ICRFy900A10109	690	Dispersed	
MCC ('a'+'b')	3	3	ICRFy900A1010	600	5q21	
			ICRFy900B0624	620	5q21	+ YAC @ 440kb[e]
			ICRFy900B0879	1200	5q21 +8p	Co-ligated

a Experiments performed by Ms. T. Jones
b Signals found by hybridisation on gridded YAC filters
c Determined by secondary filter and PFG hybridisations
d This additional YAC is referred to as 624-D, whereas the YAC identified by mc575/*MCC* is referred to as 624-575
e Same as that in yeast B0624 identifed by mc575

Table 5.3. Characteristics of additional independent YACs localised to Interval II.

Probe hybridised[a]	YAC designation	Size[b] (kb)	FISH results[c]	Comments
ECB220F.1	yECB220.1	240	nd	
	yECB220.2	270	nd	
	yECB220.3	590	nd	
YN5.48	yYN5.48.3	640	nd	
	yYN5.48.5	660	nd	
	yYN5.48.6	1100	nd	Unstable
λ12.75	y12.75	350	nd	
p3.1	ICRFy900A0819			(probe mc534; see table 5.2)
	ICRFy900A0919			(probe mc534; see table 5.2)
	ICRFy900D027	920	5q21 + C-group	

a Experiments performed by Dr. A Monaco and Mr. T. Ward
b Experiments performed by Mr. T Ward
c Experiments performed by Ms. T. Jones

Chapter 6

Ascertainment of YAC contigs 1. Coincident hybridisation analysis and physical mapping.

Chapter 6

Ascertainment of YAC contigs 1. Coincident hybridisation analysis and physical mapping.

6.1. Introduction.

The experiments described in this, and the next chapter, were designed to determine both the physical distribution and characteristics of the YACs described in Chapter 5 (table 5.2), and, to some extent, those isolated by other members of the ICRF Polyposis Group (Table 5.3). In total, 17 different YACs had been isolated by 10 independent probes mapping to interval II on chromosome 5q21-q22. These represented 5 contigs (that is, part of a set of at least 2 overlapping YACs) and two YACs not part of any contig. By taking the minimal sizes of each of these YACs (and the smallest size of one of the YACs within each of the YAC contigs), it was suspected that these YACs may have represented as much as 50-70% of interval II. However, few of the probes (and corresponding cognate YACs) mapping to this interval had been physically localised at this time. Thus, two principal questions emerged.

Firstly, it was important to establish the existence of any potential overlaps between YACs isolated by independent probes to assess the true extent of YAC contigs present across interval II. To achieve this, it was reasoned that a number of the microclones, in addition to other anonymous DNA markers in interval II (which were either negative or not tested on the primary YAC filters) could be used in hybridisation assays on the set of isolated YACs. Based on the assumption that most of the remaining probes were randomly distributed about interval II, it was hoped that some of the clones might fortuitously detect the existence of inter-YAC overlaps. In the least successful outcome, it was proposed that these assays would at least localise some of the probes. Markers that remained unlocalised after these tests would then represent primary candidates for the isolation of additional YACs, most likely by cosmid sub-fragment hybridisation assays.

The second issue was the fact that most of the microclones failed to identify any further YACs by hybridisation which suggested that genomic 'walking' would also have to be integrated into the YAC project to effect complete regional coverage of interval II. However, the isolation of all of the YAC insert termini at the time of these studies (34 in total) would have been

difficult and largely inefficient with respect to the methods then available (see Chapter 7). Thus, it was intended to construct crude physical maps of the YACs. Physical mapping would indicate the relative orientations of the YACs with respect to each other in terms of the 'left' and 'right' YAC-vector arms. This would then indicate the termini of interest for genomic 'walking' (and further coincident hybridisation assays). Integration of the results from both of these experiments was expected to provide valuable information about interval II (that is, genomic size and regional coverage) and about the sub-regions on which most effort should be concentrated.

6.2. Results.

6.2.1. Coincident mapping.

Yeast DNAs (approximately 300 to 500ng) containing the interval II YACs were digested with EcoRI, electrophoresed in 0.8% agarose gels and Southern blots were prepared. In total 11 microclones and 6 other anonymous DNA markers generated at ICRF and elsewhere (the source and description of which is provided Chapter 2, appendix 2) were hybridised on the panels. Microclones mc14 and mc575 were used as control probes to assess the sensitivity of the hybridisation assay filters. Under low stringency hybridisation conditions, these 2 microclones identified their cognate YACs after 3 to 4 hours autoradiographic exposure.

Subsequently, 17 test probes were hybridised across these panels. The results of these experiments are summarised in table 6.1. In total, 6 of the test probes (5 microclones and one anonymous DNA marker, CB83.6, (Breukel et al., 1991)) were found to be coincident with some of the interval II YACs. None of the probes, however, detected overlap amongst the independent YACs. Interestingly, 3 of the microclones, known to be different from each other, were found to be coincident to all three of the YACs isolated with probe YN5.48 (Nakamura et al., 1989). This was entirely unexpected and suggested that the microclones might not be as evenly distributed as originally anticipated (see conclusions). The YN5.48 marker, and the corresponding YACs were known to map to the telomeric end of interval II and distal to the *APC* gene by linkage analysis (Nakamura et al., 1988) and genomic physical mapping (see figure A7.1, appendix to Chapter 7).

Table 6.1. Detection of coincident hybridisation on Interval II YACs.

YACs are labelled at the top of the table using shortened acronyms, e.g. 220.1 is YAC yECB220.1, D027 is YAC ICRFy900D027 etc. The probes in **(A)** are those localised to Interval II that were successful in identifying YACs (marked with a plus sign below the respective cognate YACs; see **tables 5.2** and **5.3**, Chapter 5). The probes in **(B)** are those localised to Interval II that were used to detect coincident mapping on the set of Interval II YACs. A minus sign indicates that no cross-hybridisation was observed, a plus sign indicates that a signal was observed on the panels (see text).

YACs	220.1	220.2	220.3	10109	12.75	D027	0819	0919	D073	0624	1010	B0879	G016	0714	48.3	48.5	48.6
Probes																	
(A)																	
ECB200F1	+	+	+														
cL5.4.s-1				+													
λ12.75					+												
p3.1						+	+	+									
mc534							+	+	+								
mc575										+							
MCC										+	+	+					
mc14													+	+			
YN5.48															+	+	+
(B)																	
mc5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mc241	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mc289	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
mc404	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
mc411	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mc434	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
mc440	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
mc449	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mc451	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
mc462	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mc545	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CB83	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L5.62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EF5.44	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
λ14.16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
λ14.17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
λ12.46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

6.2.2. YAC physical mapping.

The large size of the YAC inserts and the presence of yeast chromosomes in the host cells precludes simple mapping of these inserts by multiple restriction enzyme digestions and inspection by gel electrophoresis. Moreover, the use of standard restriction enzymes with six-base recognition sites generally cut too frequently in the genome to allow resolution of the YAC insert fragments (see Nelson, 1990 for a discussion). Thus, in general, enzymes that cut infrequently in the human genome have to be used in combination with PFGE and Southern blotting for efficient mapping of the YAC inserts (Burke, 1990; Nelson, 1990).

For small YAC inserts, published reports suggested that multiple digestions of the yeast DNA with a number of infrequently cutting enzymes individually or in combination would be sufficient to generate crude physical maps (e.g. see Huxley et al., 1990). The restriction fragments could be visualised by hybridisation of PFGE blots with total human DNA, the repetitive content of which is expected to identify most of the human insert restriction fragments. Hybridisation with probes specific for the 'right' and 'left' arms of the pYAC4 vector can then be used to indicate the relative orientations of the insert restriction fragments.

Initially, experiments of this type were performed on YACs G016, A0714 A0819 and A0919 (representing the smaller size range: see table 5.2, Chapter 5) using the restriction enzymes SacII and MluI. These two restriction enzymes, the recognition sites for which are rich in cytosine and guanine nucleotides (GpG), cut infrequently in human genomic DNA. This is due, in part, to the paucity and non-random distribution of such CpG clusters in the genome, but also due to the fact that many of these clusters are methylated *in vivo* and resistant to cleavage. However, human genomic DNA isolated in YACs does not appear to be methylated at these sites by the yeast host, and the frequency of digestion is therefore somewhat higher and ideal for YAC insert mapping (Huxley et al., 1990).

These experiments were only partially successful in generating physical maps for these YACs. Nonetheless, using this approach, YACs A0819 and A0919 were found to be identical; at least within the resolution afforded by the mapping techniques used. These experiments were generally hindered by the

fact that total human DNA, or in fact probes specific for '*alu*' and 'LIHs' sequences, would not always detect each restriction fragment. For example, in the 550kb A0819 YAC, repetitive sequences entirely failed to detect a number of restriction fragments; particularly those generated by cleavage of sites nearest to the 'left' and 'right' termini of the insert, respectively. In this case, some of the mapping information could only be inferred by comparison of fragments in single digestion reactions that were found to be 'missing' after double-digestion. It was felt, therefore, that this method would be unsuitable for rigorous mapping of the YACs.

An alternative method is the use of partial enzyme digestions of the yeast DNA and subsequent hybridisation using left and right 'arm'-specific pYAC4 vector probes (Burke et al., 1987). This method, although more technically demanding, allows the visualisation of a series of extended fragments, the difference between each representing the distance between successive restriction enzyme sites.

Partial digestions can be achieved either by varying the times of digestion with a constant amount of restriction enzyme, or by dilution of the enzyme, serially or logarithmically. Preliminary experiments indicated that the latter was the most effective. It was found that cessation of the reactions in which a time course was being carried out was not efficient even after immediate incubation in 0.5M EDTA on ice (see Chapter 2). This is probably because of the time taken for chemical diffusion through the agarose plugs. The effect is unimportant, however, for extensively diluted enzymes.

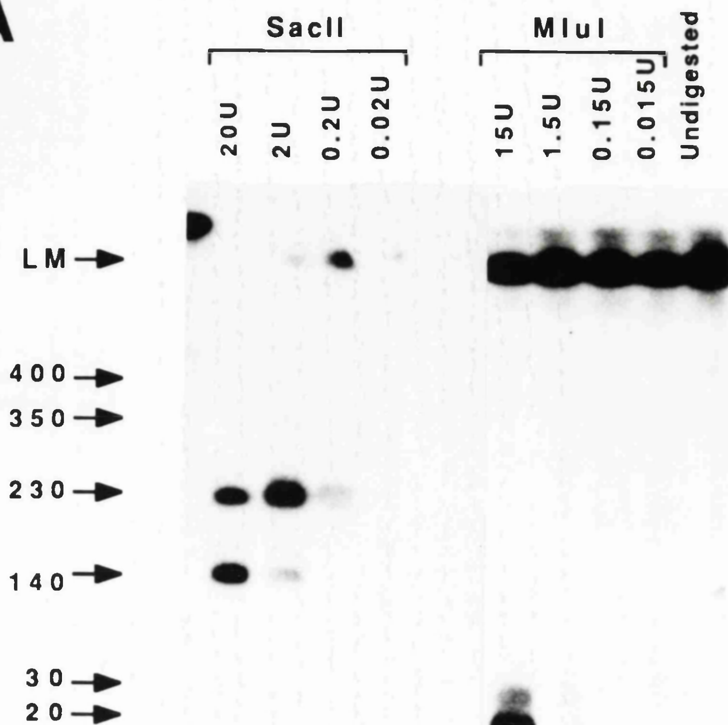
To assess the approximate activities of each enzyme, dilutions were carried out empirically on a number of YAC plugs. Even though all of the YAC DNA preparations varied, the correct range for each enzyme could usually be chosen to maximise the preferential appearance of partial digestion products (see Chapter 2). This method was successful in generating unambiguous maps for all of the YACs (including those described above). An example of the partial mapping technique on YAC D073 and the inferred physical map are shown in figures 6.1. From this example it can be seen that two of the SacII restriction sites in the centre of the YAC insert could be identified by hybridisation with the 'left-' or 'right-arm' specific probes, with nearly exact correlation in the estimated sizes detected by both probes.

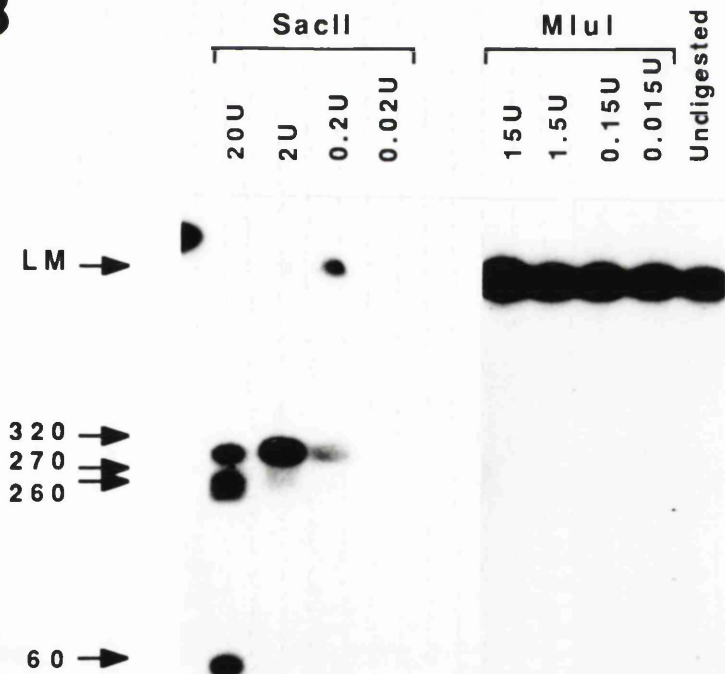
Figure 6.1. Partial restriction enzyme digest mapping of the YAC in yeast host ICRFy900D073.

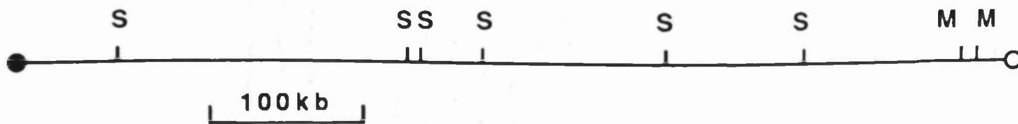
PFGE filters of partial SacII and MluI restriction digests of ICRFy900D073 sequentially hybridised with left-arm **(A)** and a right-arm **(B)** pYAC4-specific probes. One-third agarose blocks were incubated with either SacII (left hand side) or MluI (right hand side) for one hour with varying amounts of enzyme (indicated in units (U) above the figure). The last lane of the right hand side of the figure panels contains a one-third agarose block incubated with no enzyme (undigested). The filters were prepared from a 1% agarose gel run at 5V/cm for 14 hours with a linear switching time of 1 to 50 seconds. The sizes of the hybridising fragments are indicated to the left of each figure (calculated from co-electrophoresis of λ multimers (not shown)). **LM** indicates the point of limiting mobility of restriction fragments in the agarose gel (≈650kb). Filters were washed to a final stringency of 1 x SSC/0.1% SDS at $65^{0}C$ and exposed to autoradiographic film overnight (shown here) or up to 2 days to ensure visualisation of all partial restriction fragments (not shown).

The MluI digested DNA fragment detected by the right-arm probe migrates at the limiting mobility (LM) of the PFG (lower panel, right hand side). The size of this fragment (≈650kb), when added to the MluI fragments detected by the left arm probe (≈30kb; top panel, right hand side) equals the size of the undigested YAC (680kb).

(C) The inferred restriction map is depicted by the solid black line below the autoradiographs. The unfilled and filled circles represent the left and right arms, respectively. The restriction sites (S = SacII; M = MluI) are depicted above the line.


A
SacII
MluI
20U
2U
0.2U
0.02U
15U
1.5U
0.15U
0.015U
Undigested
LM
400
350
230
140
30
20


B
SacII
MluI
20U
2U
0.2U
0.02U
15U
1.5U
0.15U
0.015U
Undigested
LM
320
270
260
60


C
S
SS
S
S
S
M M
100kb

6.2.3. YAC contigs.

The physical maps derived for all of the other YACs studied are depicted in figure 6.2.

Contig 1:

Contig 1 (figure 6.2(a)) represents the two YACs identified by probe mc14; that is, G016 and A0714 (refer to table 5.2, Chapter 5). The overlap of these two YACs could be inferred unequivocally as they share a large proportion (400kb) of genomic DNA and span a total distance of only 650kb. Probe mc14, and the probe found to be coincident on these YACs - mc451, lie on opposite sides of coincident SacII and MluI restriction sites which are also detected in genomic DNA (indicated above the contig).

Contig 2:

Contig 2 represents the YACs identified by the *MCC* gene (A1010, B0624 and B0879) and by microclone mc575 (B0624). FISH analysis on the largest of these three YACs; B0879, suggested that this YAC was chimeric (see figure 5.3(a), Chapter 5). This was confirmed by attempts to align the physical maps. It was concluded that this YAC contained only a small proportion of DNA from chromosome 5 and was therefore not investigated further. The other two YACs; B0624 and A1010 were known to overlap at the *MCC* gene by probe hybridisation on the primary YAC filters (see Chapter 5). The actual extent of overlap was suspected to be quite small and could not be assessed by digestion with SacII and MluI alone. The true extent of overlap (100kb), was established by the use of termini-specific fragments on multiple digests of the A1010 YAC (see figure 8.3, Chapter 8).

Due to the co-transformation of the B0624 yeast host cell with an unrelated YAC, the YAC specific for mc575 and the *MCC* gene (referred to as 624-575; see table 5.2, Chapter 5) could not be readily mapped. To circumvent this problem, sequences representing both the left and right hand insert termini of this YAC were isolated and used for partial restriction fragment detection. The isolation and characterisation of these insert termini is presented in Chapter 7. Figure 6.2(b) depicts the SacII and MluI map derived for these two YACs, which together span a total distance of 1,100 kb. It has also been possible to align this contig on the genomic physical map by virtue of

coincident and identical MluI and SacII restriction sites which lie centromeric to the *MCC* gene in the YAC and genomic DNAs (indicated above the contig).

Contig 3:

Contig 3 represents the YACs isolated by probes mc534 and p3.1 (that is, D073, A0819 and D073). Alignment of these three YACs was found to be extremely difficult since two of them appeared to be chimeric. The most probable alignment of YACs D073 and A0819 is shown in figure 6.2 (c). It is assumed that YAC A0819 does not contain sequences from regions of the genome other than chromosome 5 (and certainly not at the end hybridising to probe p3.1) whereas D073 almost certainly does on the basis of FISH analysis. The dotted lines in the latter are intended to indicate the region of divergence from chromosome 5 sequences. The physical map implies that about half of this YAC originates from 5q21-q22 which is in line with the relative intensities of disparate genomic signals observed by FISH analysis (see section 5.8.4. and table 5.2, Chapter 5).

It has been impossible to align YAC D073 with respect to the other two YACs. The restriction map of this YAC is therefore depicted in isolation such that an inappropriate contig is not inferred by this study. It was subsequently found that this YAC in fact contains sequences from regions of the genome, other than chromosome 5q21-q22, at both ends of the YAC insert (see section 7.4.6. and table 7.1, Chapter 7). Attempts to orientate these three YACs by the use of further restriction enzymes or by cross-hybridisation analysis using randomly derived inter-*alu* probes was not pursued. This contig was not thought to be useful for the initiation of genomic 'walking' on the basis of the high degree of insert chimerism.

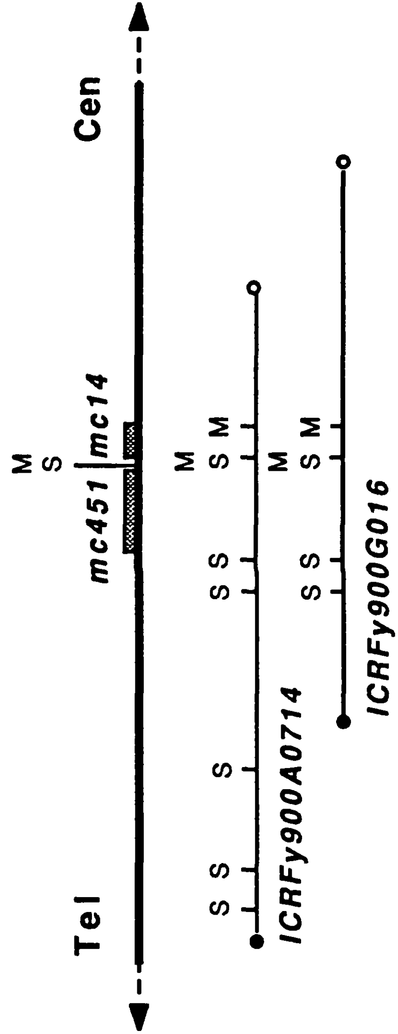
Contigs 4 and 5:

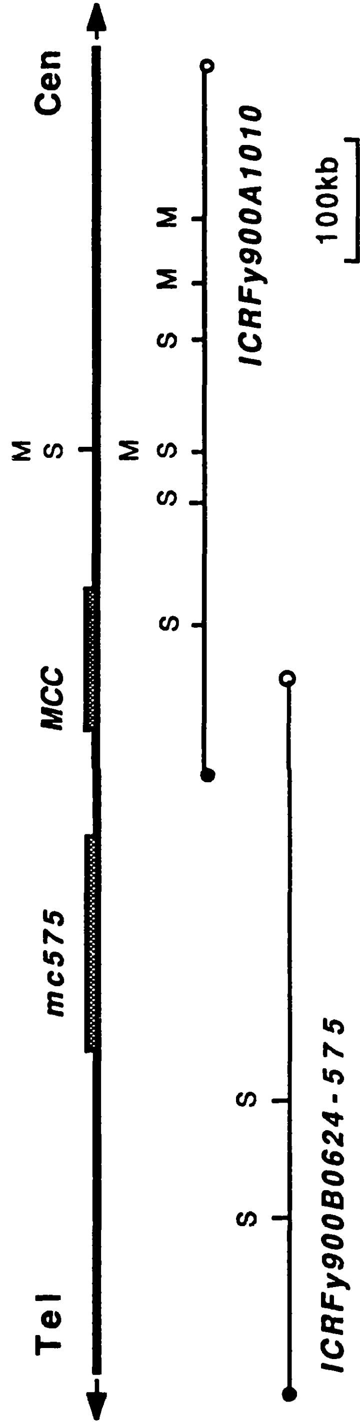
These two contigs, which are represented by three YACs isolated with each of the ECB220F1 and YN5.48 probes (see table 5.3, Chapter 5) have not been investigated by either FISH analysis or physical mapping at the time of writing (T. Ward, personal communication). Thus, the genomic distances that each of these YAC contigs covers can not be calculated at present.

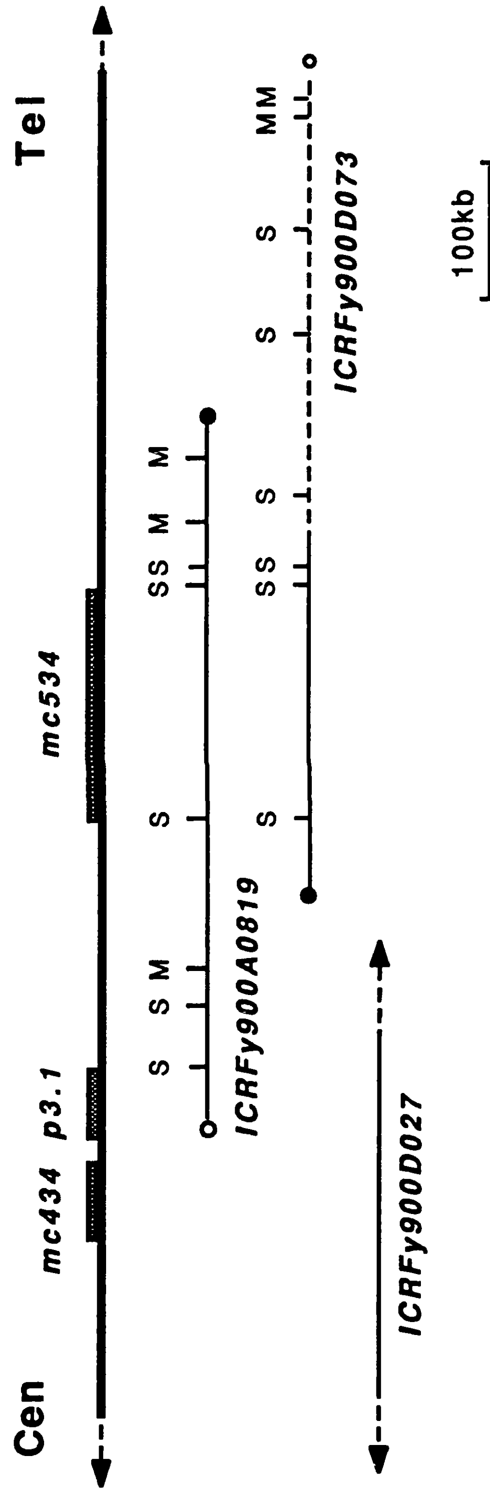
<u>Figure 6.2.</u> Inferred restriction maps and contig structures.

The YACs and YAC contigs depicted in **(a)** to **(e)** are represented by thin black lines and the name of each is denoted below. Unfilled and filled circles represent the left and right vector arms, respectively. The letters **S** and **M** denote SacII and MluI restriction sites, respectively. The thick black lines above the YACs in **(a)**, **(b)** and **(c)**, represent genomic DNA, the hatched lines with arrows at the end of these lines are intended to indicate a continuous stretch of DNA. The relative orientations of the YACs with respect to chromosome 5 in **(a)**, **(b)** and **(c)** are indicated in terms of the centromere (**CEN**) and telomere (**TEL**). In the two cases where some of the restriction sites could be matched in both the YAC and human maps (i.e., in **(a)** and **(b)**), the sites are depicted above the genomic DNA. Hatched boxes on the genomic DNA lines (and under the YACs in **(d)** and **(e)**) represent the smallest region that each probe hybridised to on the YAC digest PFGE filters.

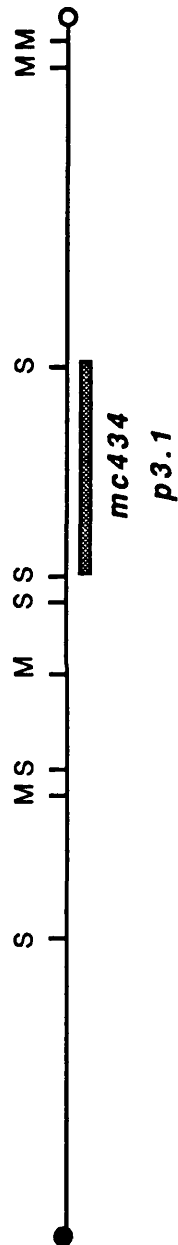
The contig that could be inferred from restriction mapping of two of the YACs; **ICRFy900D073** and **ICRFy900A0819** (i.e., contig 3), is depicted in figure **(c)**. The hatched line in **ICRFy900D073** represents the region of this YAC that is thought not to originate from chromosome 5. **ICRFy900D027** is illustrated below probes **mc434** and **p3.1**, but is not drawn to scale. Restriction sites in this YAC (which does not contain chromosome 5 DNA at either end of the insert; see Chapter 7) could not be aligned with respect to the other two YACs (represented by the hatched lines and arrows). The map of this YAC is illustrated in isolation in **(d)**. The restriction map of YAC **ICRFy900A10109** is depicted in **(e)**.

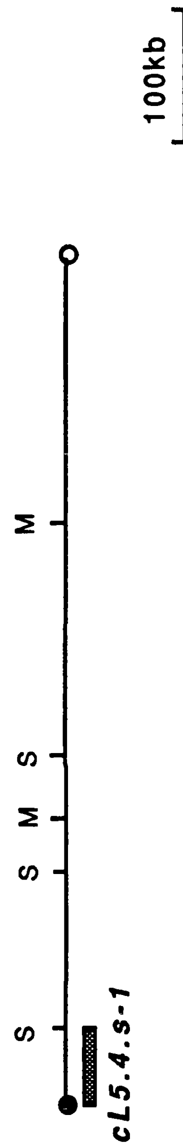

a
CONTIG 1
Tel
Cen
mc451
mc14
ICRFy900A0714
ICRFy900G016


b
CONTIG 2
Tel
Cen
mc575
MCC
ICRFy900A1010
ICRFy900B0624-575
100kb



CONTIG 3
c
Cen
Tel
mc434
p3.1
mc534
S
S
M
S
SS
M
M
ICRFy900A0819
S
SS
S
S
S
MM
ICRFy900D073
ICRFy900D027
100kb


d
ICRFy900D027
M M
S
S S
M
M S
S
mc434
p3.1


e
ICRFy900A10109
M
S
S M
S
cL5.4.s-1
100kb

YAC A10109:

The restriction map derived for this YAC is presented in figure 6.2(d).

6.3. Conclusions.

The use of coincident mapping on the YAC panels provided a direct means by which the distribution of a number of previously unlocalised probes could be established. It was not, however, successful in the detection of YAC overlaps. This was thought to be due to a number of factors, primarily the physical distribution of the microclones (described in the appendix to Chapter 7 and discussed in Chapter 9).

Partial restriction mapping of the YACs with infrequently cutting enzymes allowed the construction of accurate maps. In general, it was possible to establish, exactly, the overlaps between YACs that were known to contain common sequences. However, this was found to be difficult in cases where one or more YACs appeared to be chimeric (for example in contig 3). The apparent frequency of chimerism suggested that construction of a YAC contig across interval II would be more difficult than originally anticipated, particularly in the absence of further sub-localised markers. Chapter 7 deals with the isolation of YAC insert termini as an alternative method for the ascertainment of YAC contigs. These experiments, which were largely conducted at the same time as physical mapping support and considerably extend the observation of a high degree of chimerism in the ICRF YAC libraries.

Chapter 7

Ascertainment of YAC contigs 2. Detection of inter-YAC overlap and isolation of terminal sequences

Chapter 7

Ascertainment of YAC contigs 2. Detection of inter-YAC overlap and isolation of terminal sequences.

7.1. Introduction.

The results of the YAC coincidence mapping presented in Chapter 6 suggested that other methods should be sought to identify potential overlaps within the set of Interval II YACs.

Most of the methods described for random overlap detection are based on repetitive sequence content (e.g. Coulson et al., 1986). For example, the YACs could be digested with one or more commonly cutting restriction enzymes and Southern blots probed with fragments containing *alu* or LIHs sequences to create a fingerprint. Simple comparison would then give an indication of potential overlaps. However, this method is limited in that any observed similarities are inferred and cannot be directly followed up. The level of confidence in the detection of overlapping YACs by this method can only be based on experience. Recent work published at the time of writing has suggested that this approach can be effective, but, only with the use of a number of probes recognising high, moderate and low copy-number repeats in combination with a suitably derived algorithm for the calculation of likelihood scores (Schlessinger et al., 1991; Zucchi and Schlessinger, 1992).

An extension of this method; *alu*-PCR (see Chapter 3), has been applied directly to YACs (Nelson et al., 1989). In this case any potential overlaps can be further investigated by isolation of the apparently common products and hybridisation to the YACs in question. This system, however, has been found to be limited by the numbers of PCR products generated even from primers highly conserved with respect to *alu* consensus sequence. In general, overlaps can be confirmed in cases where the YACs have been identified by the same locus or where YACs have been identified by walking (Nelson et al., 1991; Butler et al., 1992) but rarely in cases where the degree of overlap is small (Butler et al., 1992).

However, the use of a number of highly conserved *alu*-PCR primers on each YAC would give rise to a very high density of products, especially if the *alu*-

based oligonucleotides were designed to prime from both directions of the repeat elements. The overlaps could then be detected by hybridisation directly onto a panel of *alu*-PCR products of the YACs to be investigated. This method, which is really an extension of that described by Ledbetter et al., (1991) and Bicknell et al., (1991), was assessed on Contigs 1 and 2 (figure 6.2, (a) and (b)), which represent the extremes of overlap so far detected in the interval II YACs. Using three highly conserved primers, termed PDJ-33, -34 and -66 (recently published in Butler et al., (1992)), overlap was unequivocally detected in Contig 1 but not in Contig 2. Since Contig 2 represents the degree of overlap likely for independent YACs, this implied that some potential overlaps would certainly be missed. This pilot experiment suggested that other more robust methods must be assessed.

In view of the considerations outlined above, it was decided to isolate systematically the termini of the YAC inserts, particularly those that represented the distal extents of contigs inferred by coincident hybridisation and, in some cases, by physical mapping. These probes could then be used on the YAC coincident mapping panels. It was proposed that the 'end-probes' would provide the maximum amount of information possible and not be dependent on the distribution of either repeat elements in the YACs or on the distribution of probes on the genomic physical map. In cases where a YAC was chimeric, it was expected that other approaches, such as *alu*-PCR-based methods, would be equally disadvantaged. In any case, chimeric YACs were only thought to occur in about 1/5 cases (20%; Dr. Anthony Monaco, personal communication; Monaco et al., 1992). Moreover, these terminal sequences could then be used directly for genomic walking from the appropriate YACs.

A number of methods have been described for the isolation of YAC insert termini which depend on (1) library construction, (2) re-circularisation of restriction fragments followed by sequence-specific amplification by PCR or genetic selection and (3) amplification of vector-insert junction sequences by *alu*-vector PCR or 'vectorette-PCR'. These methods are discussed in Chapter 1.

After consideration of these different available methods, I decided to investigate the use of '*alu*-vector' PCR on a subset of the YACs because of the experience gained using the *alu*-PCR method (Chapter 3) and the availability of suitable oligonucleotide primers.

7.2. <u>Results 1.</u> Isolation of YAC insert termini.

7.2.1. *Alu*-vector polymerase chain reaction (PCR).

For these experiments YACs G016 and A0714 (see table 5.2, Chapter 5 and figure 6.2(a), Chapter 6) were chosen since the mc14 and mc451 probes had not yet been physically localised within Interval II. Isolation of terminal fragments would be useful for localising the contig in this respect. In addition, YACs B0624 and A1010 (see table 5.2, Chapter 5 and figure 6.2(b)) were also assessed for attempts to expand the contig around the *MCC* gene by genomic walking.

The strategy followed was to amplify each of the YACs in parallel with the *alu* primer alone and then in combination with primers specific for the 'left' and 'right' vector arms. It was anticipated that the appearance of a novel product in the *alu*-vector primer reactions would indicate a potential end-fragment that could be further investigated. The vector-specific primers; 1089 (left arm-specific) and 1091 (right arm-specific) were based on those described by Riley et al., 1990 (see appendix 1) which are positioned 280 and 160bp from the left and right pYAC4 vector arm EcoRI sites, respectively. A set of four *alu* primers; 517, 559 (Nelson et al., 1989; see Chapter 3, part II), primer IV (see Chapter 3, part I) and another primer; PDJ66 (Dr. P. deJong, personal communication; Butler et al., 1992), known to be highly conserved with respect to the *alu* consensus sequence, were used for these experiments.

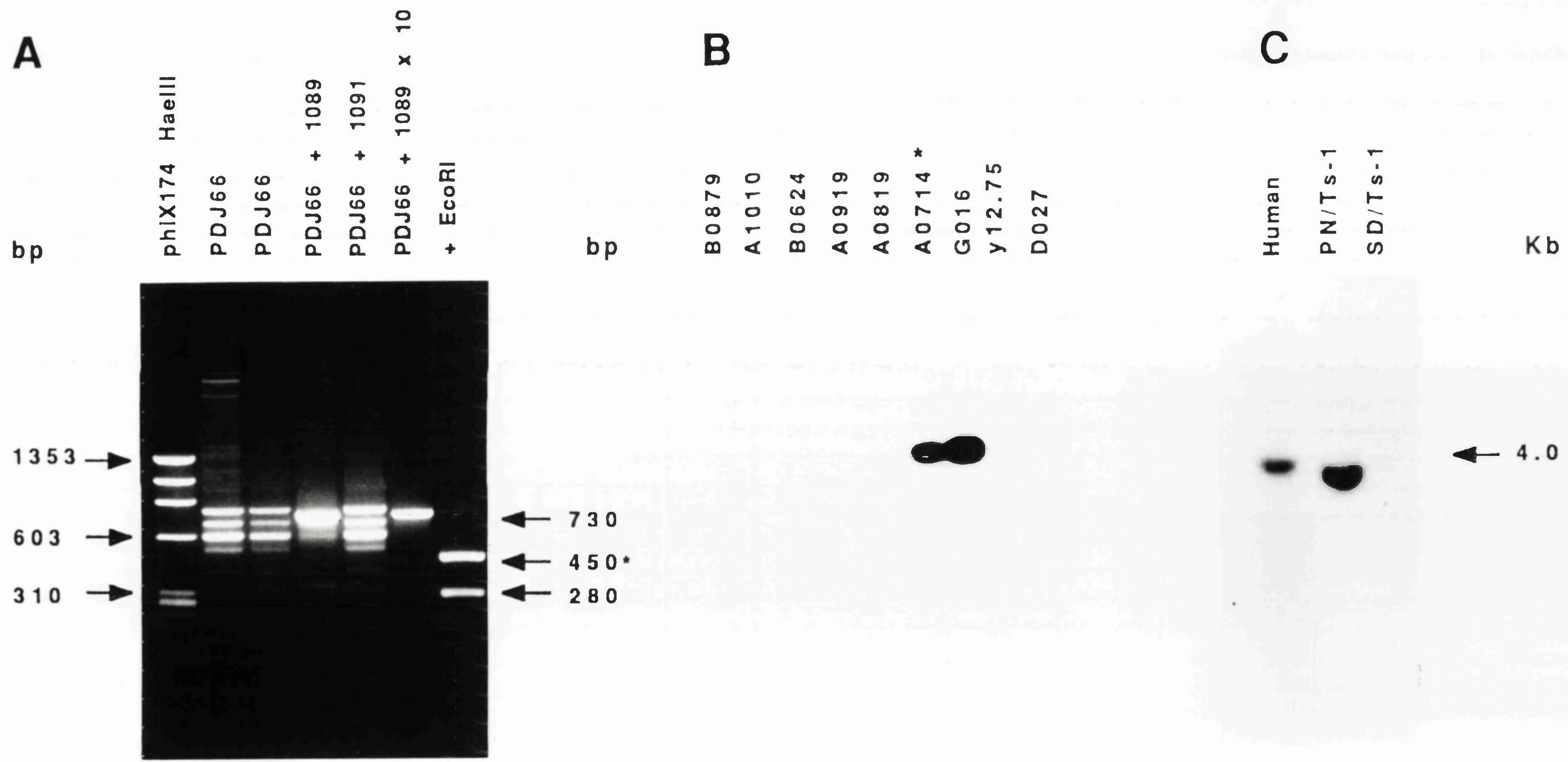
In total, two of the *alu*-vector PCR reactions gave rise to new products (see table 7.1 at the end of this chapter) in YACs A1010 (1010LHE; primers 517 & 1089) and A0714 (714LHE; primers PDJ66 & 1089). Due to the positions of the vector-specific primers, PCR products representing true terminal sequences should result from amplification across the EcoRI cloning site of the pYAC4 vector. Therefore, digestion with EcoRI should produce two products; one of constant size representing the distance from the vector-specific priming binding site to the EcoRI cloning site and another of variable size representing the distance of the EcoRI cloning site to the *alu* element. Digestion of the 1010LHE and 714LHE products each gave rise to two fragments, one of which was the expected vector fragment size (280bp) and the other a variable size (850bp and 450bp respectively). Figure 7.1(a) demonstrates the *alu*-vector amplification of YAC A0714 with primers PDJ66 and 1089 and the fragments obtained after digestion with EcoRI.

Figure 7.1. Isolation and characterisation of YAC-insert terminus '714LHE'.

(A) Electrophoresis of *alu* and *alu*-vector PCR products from YAC ICRFy900A0714 on a 1.5% agarose gel stained with ethidium bromide: **PDJ66** is the *alu*-specific primer and **1089** and **1091** are the left (TRP1) and right (URA3) arm-specific primers, respectively (refer also to **figure 7.4**). *Alu*-PCR of the YAC is shown in duplicate. Note the highly specific amplification of the novel product when the pYAC4 **1089** primer is used in 10-fold molar excess (**PDJ66 + 1089 x 10**). The **450bp** insert terminus-specific product (714LHE), used as a probe in **7.2 (A)** and **(B)**, is indicated by the asterix.

(B) Southern blot hybridisation of a filter containing EcoRI digested whole yeast DNAs containing some of the interval II-specific YACs. The identity of each of the YACs is indicated above the figure. The filter was washed at 0.2 x SSC/0.1% SDS at 65^{0}C exposed to autoradiographic film overnight.

(C) Southern blot hybridisation of a filter containing total genomic and hybrid DNAs digested with EcoRI. The identity of the two hybrids DNAs are indicated above the figure (see text). This filter was washed to a final stringency of 2 x SCC/0.1% SDS at 65^{0}C and exposed to autoradiographic film for 3 days.


A
bp
phIX174 HaeIII
PDJ66
PDJ66
PDJ66 + 1089
PDJ66 + 1091
PDJ66 + 1089 x 10
+ EcoRI
bp
1353
603
310
730
450*
280
B
B0879
A1010
B0624
A0919
A0819
A0714 *
G016
y12.75
D027
probe: A0714LHE
C
Human
PN/Ts-1
SD/Ts-1
Kb
4.0
probe: A0714LHE

It can be seen from this experiment that the *alu*-PCR patterns specific for the YAC were greatly reduced in the *alu*-vector reactions in which novel products were generated. This effect could be greatly enhanced by performing the PCR reaction with 10-fold excess of the vector-specific primers (figure 7.1(a)) and was useful for the analysis of essentially 'pure' products.

Screening of these fragments on the coincident mapping panels described in Chapter 6, showed that both hybridised to the cognate YAC from which these sequences were derived. 714LHE also hybridised to YAC G016, which was expected on the basis of overlap detected by physical mapping of these YACs (see figure 6.2(a), Chapter 6). Both of these *alu*-vector junction fragments mapped into interval II by hybridisation on the chromosome 5 mapping panels. Examples of the coincident and genomic hybridisation assays for 714LHE are presented in figure 7.1 (b) and (c).

Further confirmation for the *bona fide* nature of the candidate termini was achieved by hybridisation of the fragments to SacII and MluI digests of the respective YACs. Each sequence mapped to the appropriate terminal restriction fragments from the YAC of origin and, in the case of the 714LHE, to the appropriate internal restriction fragments of YAC G016 inferred from the physical maps.

7.2.2. Conclusions on the *alu*-vector method.

The *alu*-PCR method led to isolation of only 20% of the insert termini investigated and was felt to be inadequate for further studies. Due to the availability of specific 'vectorette' cassettes amongst colleagues at ICRF, it was decided to assess the 'vectorette' method for its potential in generating the remaining insert termini of interest.

7.2.3. 'Vectorette'-polymerase chain reaction (PCR).

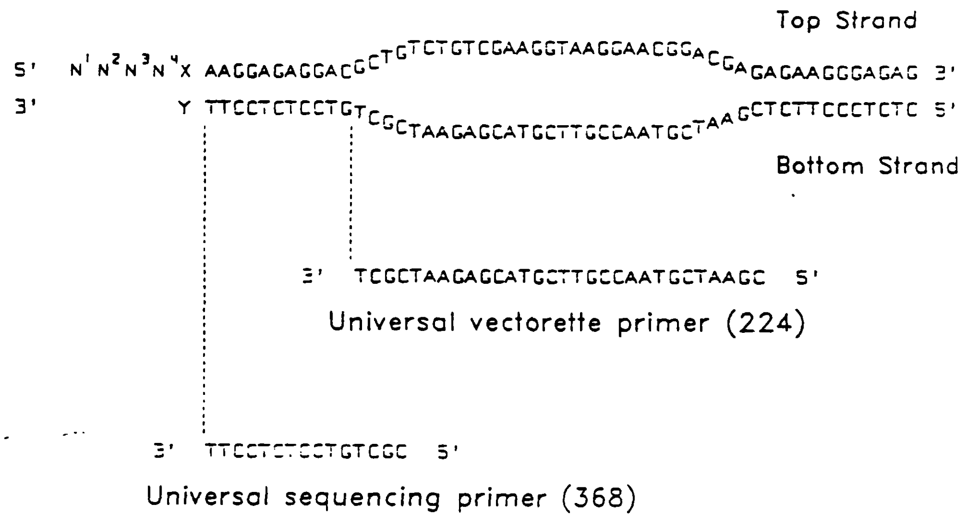
The principle of the 'vectorette' method is outlined in figure 7.2. To assess this system, a further subset of YACs was used. These included YACs D027 and D073, which were thought to represent the terminal YACs in the mc434-p3.1-mc534 contig prior to physical mapping (i.e. contig 3; figure 6.2(c), Chapter 6) and YAC A10109 since the cL5.4.s-1 probe had been mapped close to ECB220F1 (Mr. T. Ward, personal communication), within a distance for potential

Figure 7.2. Principal of the 'Vectorette-PCR' method.

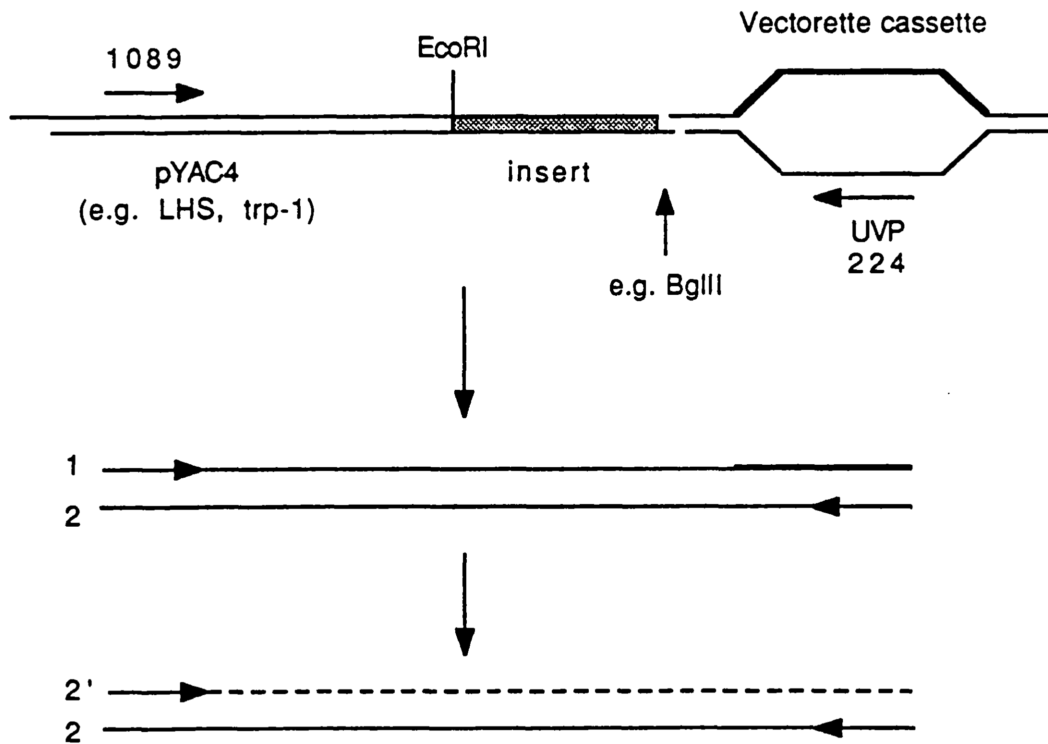
(a) Illustration of the vectorette cassette sequence (from Riley et al., (1990). The nucleotides (N^1, N^2, N^3, N^4) represent the positions of bases creating an overhang for ligation to digested yeast DNA (e.g. BglII).

(b) A schematic illustration of the vectorette method. In this illustration a fragment from the left arm (i.e. TRP1 gene) of pYAC4 with an associated portion of YAC insert DNA (cut by BglII) is ligated to a vectorette cassette. In this situation amplification takes place between primer **1089** (specific for the left arm; refer also to **figure 7.4**) and the universal vectorette primer (**UVP**) **224**. In the first round of PCR, a product is formed that still has non-complementary sequence at one of the ends (i.e. in **strand 1**; indicated by the bold portion of the line). Therefore, in the second round, only **strand 2** participates further in the PCR reaction allowing the synthesis of **strand 2'** (directed by primer **1089**) and so forth. Thus, in a situation where a vectorette cassette is ligated on both sides of a random piece of DNA, amplification could not occur after the first round of PCR since there would be no complementarity at either primer binding sites.

(a)



(b)



overlaps between these sets of YACs to be detected (refer to figure A7.1, appendix to this Chapter).

For these experiments, yeast DNAs from each of the YAC clones were digested with each of three restriction enzymes; BamHI, HindIII and PstI compatible with the vectorette 'cassettes' available at the time. The resultant digestions were checked by electrophoresis on 0.8% agarose gels. BamHI was rarely successful for complete yeast DNA digestions and was replaced with BglII, which is 'end-compatible' with this restriction enzyme. BglII consistently gave rise to total digest products. For these experiments an additional YAC, designated Y.2, was included as a positive control to assess the various stages of this protocol. This was found to be necessary in each case as the success of the 'vectorette' experiments tended to be variable. This YAC was known to generate a terminal-specific PCR product from the 'left' vector-insert junction using BglII digest libraries (kindly provided by Ms. Louise Sefton, Human Genetics Laboratory, ICRF).

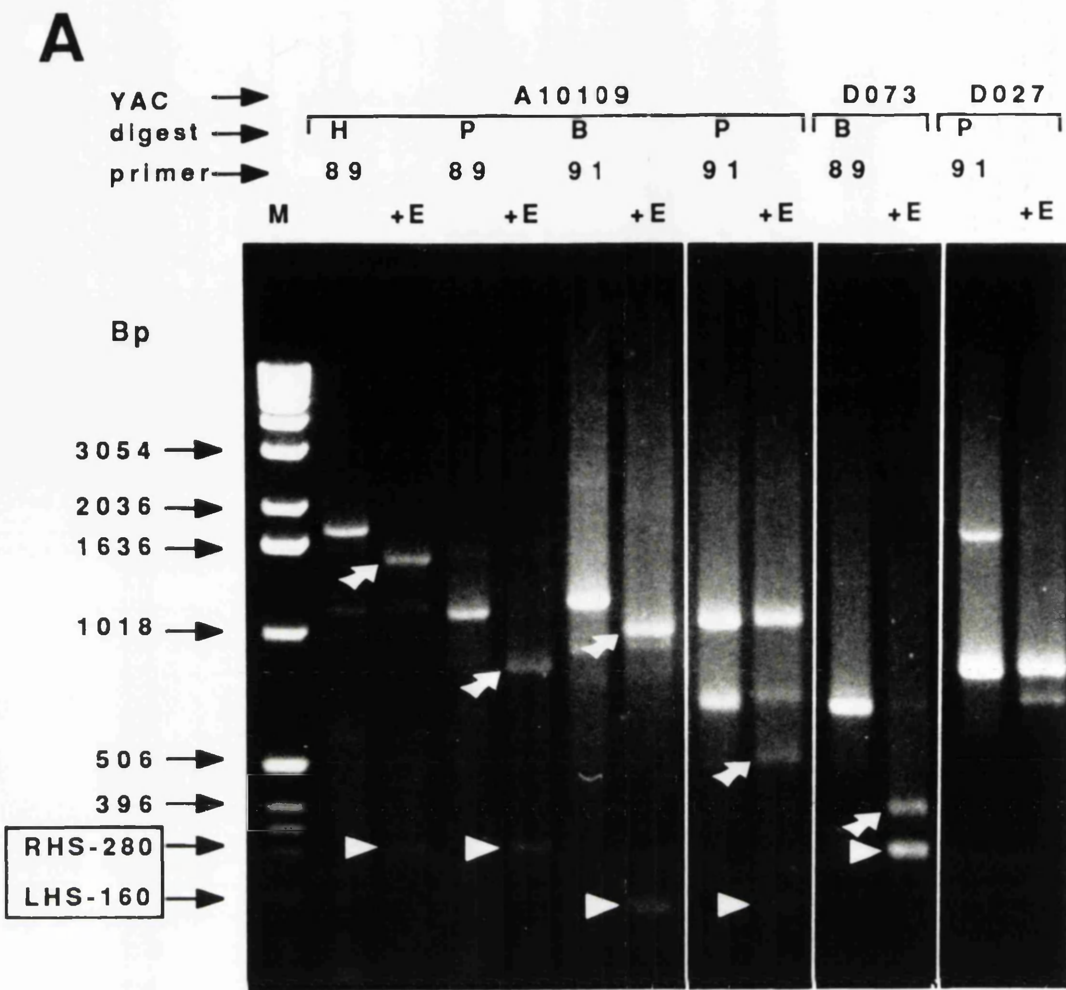
From the four YACs initially studied, 6 potential termini were generated; 2 each from both the left and right arms of YAC A10109, one from the left arm of D073 and one from the right arm of YAC D027 (summarised in table 7.1 at the end of this chapter). The appropriately sized PCR fragment was generated from the control Y.2 BglII library. The results of these experiments are shown in figure 7.3(a). It can be seen that a number of other quite intense PCR products were often generated in the amplification reactions in addition to the correct products. The reason for these is not understood but they could be eliminated by decreasing the amount of vectorette library DNAs used in each of the PCR reactions by about 10-fold. As for '*alu*-vector' PCR, these PCR products were assessed by EcoRI digestion ('+E') and, in each case, except for the right hand end of the D027 YAC insert, the appropriate sized vector-specific fragments were generated (indicated by the arrowheads).

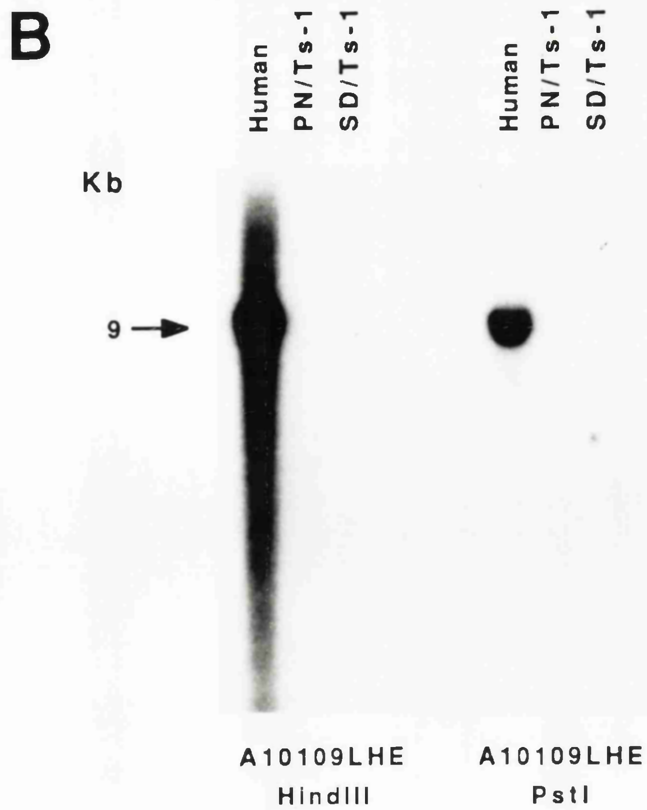
Of these 5 EcoRI-digested termini (indicated by the arrows in figure 7.3), 3 could be successfully mapped with respect to the somatic cell hybrid panel. The two fragments from the left end of A10109 mapped to the same EcoRI restriction fragment in human DNA but did not map to chromosome 5 (see figure 7.3(b)). Comparison of the signal intensities suggested that the larger fragment generated from the HindIII library (≈1.6kb) appeared to contain a repetitive element, presumably located between the PstI site in the smaller fragment from the same terminus (≈800bp) and the HindIII site of the larger

Figure 7.3. **Generation and characterisation of 'vectorette' PCR products.**

(A) 1.5% gel electrophoresis of 'vectorette' PCR products from YACs ICRFy900A10109, ICRFy900D073 and ICRFy900D027 before and after EcoRI digestion. The yeast digest libraries **(B)** BglII, **(H)** HindIII and **(P)** PstI and pYAC4 vector arm-specific primers **(89)** 1089 or **(91)** 1091 used in each case are indicated above the figure. EcoRI digests of the PCR products **(+E)** were run in alternate lanes as indicated. Arrows point to the EcoRI digested YAC-insert termini fragments used as probes in further characterisation experiments (e.g. as in **(B)**). Arrowheads point to the pYAC4-specific fragments released after EcoRI digestion (**LHS-280bp** or **RHS-160bp** for the left and right vector arms, respectively). The absence of a pYAC4-specific 160bp fragment in YAC **D027-P-91, +E** implies that this amplification product is probably an artifact. Lane **M** contains the 1kb ladder size markers. The sizes of some of the fragments are indicated to the left of the figure.

(B) Two examples of genomic mapping using ^{32}P-labelled fragments derived from the left-hand-end of YAC ICRFy900A10109 (fragments **10109-89-H** and **10109-89-P** on the left and right hand side filters, respectively). The genomic and hybrid DNAs were digested with EcoRI and the identities of each of the two hybrids are indicated above the figure. Roughly equal loading of all DNAs on the gels from which these filters were taken was demonstrated by hybridising probes within and outside the SD deletion (not shown). The filters were washed to a final stringency of 2 x SSC/0.1% SDS at 65^{0}C and exposed to autoradiographic film overnight.


A
YAC
A10109
D073
D027
digest
H
P
B
P
B
P
primer
89
89
91
91
89
91
M
+E
+E
+E
+E
+E
+E
Bp
3054
2036
1636
1018
506
396
RHS-280
LHS-160


B
Human
PN/Ts-1
SD/Ts-1
Human
PN/Ts-1
SD/Ts-1
Kb
9
A10109LHE
HindIII
A10109LHE
PstI

fragment. Neither of the BglII nor PstI fragments from the right end of this YAC could be mapped even under stringent competitive hybridisation conditions. The BglII fragment from the left end of the D073 YAC did not map to chromosome 5. Each of the 5 fragments were then investigated further by hybridisation to SacII and MluI digests of the YACs and all behaved as expected for *bona fide* insert termini.

7.2.4. Conclusions on the 'vectorette' method.

Without regard to the high frequency of apparent chimerism in the subset of YACs tested (see section 7.6 (conclusions) and Chapter 9), the frequency of termini generated by PCR in these experiments, combined with some associated technical difficulties, implied that a more robust method should be sought to isolate the remaining insert termini.

7.3. 'Junction-trapping': a new approach to the isolation of YAC-insert termini.

7.3.1. Introduction and strategy.

Experience with the techniques for the generation of insert termini led to the development of a novel and simplified protocol. This new protocol was devised on the basis of the following observations.

Firstly, the protocols for both the *alu*-vector and 'vectorette' PCR methods were seen to be highly dependent on the distribution of either repetitive elements or specific restriction enzyme sites in the YAC insert DNA, respectively. In both cases, the sequences of interest must be spaced closely enough to the EcoRI cloning site of pYAC4 to be within a distance amenable to PCR. For *alu*-vector PCR, this presents obvious problems in genomic regions with a poor representation of repetitive elements. It could be argued that the use of enough 'vectorette' linkers (specific to different restriction enzymes) should be capable of 'trapping' a sufficient number of closely spaced sites. However, this would greatly increase the expense and the relative labour effort involved for the generation of each terminus-specific fragment.

The use of a frequently cutting restriction enzyme such as Sau3A1, which cuts on average every ≈300bp in genomic DNA (Drmanac et al., 1986), should, in nearly every case, be spaced close enough to the pYAC4 cloning site for a PCR-

based strategy. However, because of the presence of a Sau3A1 recognition site between the 1091 primer binding site and the cloning site on the right pYAC4 vector arm, total digests could not be used. Partial restriction digests, on the other hand, should produce a number of molecules that avoid this internal restriction site, but which will still result from the cleavage of the sites in the insert DNA close to the pYAC4 vector arm.

Further, the use of partial digests was expected to provide a variety of progressively longer insert fragments which would still be within a distance amenable to PCR. This would then provide a choice of terminal fragments from which to work with. As illustrated by the 'vectorette' experiments above, a choice of fragments from a particular terminus may be useful in avoiding the positions of repetitive elements (compare the hybridisation signals generated by products 10109LHE (HindIII) and 10109LHE (PstI) in figure 7.3(b)).

Secondly, the 'vectorette' system was specifically designed to avoid amplification across random restriction fragments by the use of double-stranded linkers with a region of non-complementarity. This safe-guard is essential in a system where the linker can ligate to either side of random insert fragments or to themselves; a double-stranded linker that is complementary along its length would allow non-specific amplification even in the presence of pYAC4 vector-specific primers.

However, it was hypothesised that the yeast DNA fragments could be 'trapped' by ligation to BamH1-digested plasmid using conditions favouring circularisation. This was thought to be best achieved by 5'-dephosphorylation of the insert DNA and not the plasmid. Thus, a vast excess of insert DNA could be used in combination with very small amounts of the plasmid DNA. In this approach, co-ligation of insert DNA or co-ligation of plasmid molecules on either side of random fragments would be unlikely. In any case, the latter would only represent a very small proportion of the ligation events. The vector-insert junctions could then be amplified using a plasmid specific primer and the primers specific to the pYAC vector arms.

The result of such a PCR reaction was expected to be quite complex and a further two selection mechanisms were devised in an attempt to increase the overall specificity of the reactions:

(1) Bias of amplification: This could be achieved by using the pYAC4 vector-specific primer in 10-molar excess. This was based on the observation of the large bias effected in the *alu*-vector PCR reactions where the novel junction products were selected to the detriment of the inter-*alu* products (see figure 7.2(a), lane: PDJ66 +1089 x 10).

(2) Secondary amplification: This could be achieved by utilising primers internal to both the plasmid and pYAC4 vector primers used in the first round of PCR. This would be expected to re-select the required amplification products using an aliquot of the first PCR reaction products as the template. Inherent in this would be the ability to favour the annealing temperature, slightly, in favour of the pYAC4 primers. In this way, it was expected to greatly reduce the complexity of the resultant PCR products. Furthermore, this would also avoid the use of excessive cycling in the primary PCR reaction to generate sufficient product. Such cycling might be expected to allow the generation of non-specific artifacts.

The positions of the primers in both the pBluescript (the plasmid used in these exeriments) and pYAC4 are illustrated in figure 7.4. A schematic illustration of the method is depicted in figure 7.5.

7.4. Results 2. 'Junction-trapping' of YAC termini in yeast B0624.

To illustrate this method, the isolation of YAC insert termini for each of the two YACs in the B0624 yeast host cell is described in detail (refer to table 5.2, Chapter 5 for details of the YACs in ICRFy900B0624).

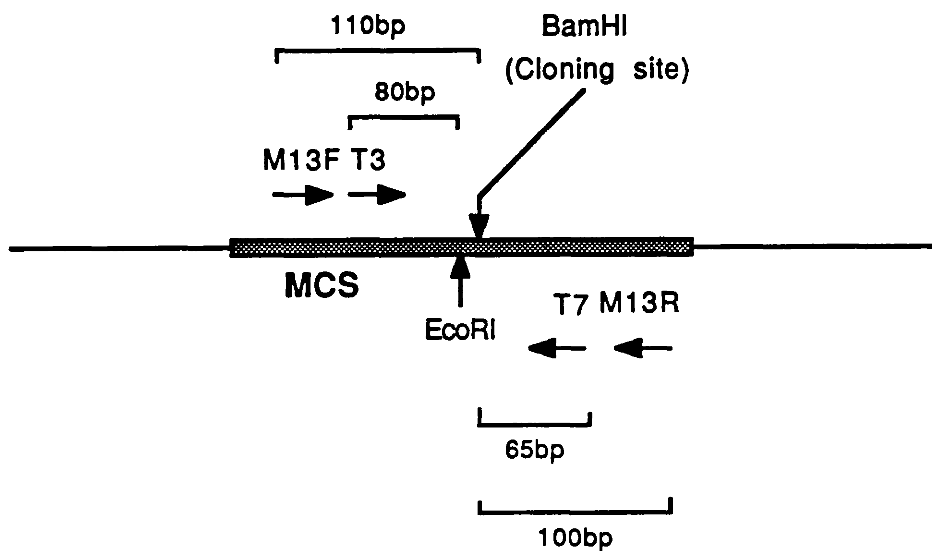
7.4.1. Preparation of insert DNA.

The B0624 yeast DNA was digested with a constant amount of Sau3A1 enzyme (20U) for varying times to generate partial digests products (see Chapter 2). Inspection of the digested DNA on 0.8% agarose gels generally demonstrated that one or more digestion conditions were ideal (in the 500 to 2000bp range). Due to the predicted specificity of the PCR reactions, it was thought that the presence of totally digested YAC DNA would not interfere with amplification reactions and thus, the partial digests were pooled and the DNA purified. During partial digestions, calf intestinal phosphatase (CIP) was added to 5'-dephosphorylate the insert DNA (see above).

<u>Figure 7.4.</u> Positions of pBluescript- and pYAC4-specific amplification primers.

(a) This diagram shows a portion of the pBluescript plasmid around the multiple cloning site (**MCS**). The positions of the 4 pBluescript-specific primers are indicated with the distances of each (except T3) to the BamHI cloning site in **MCS**. The distance of the **T3** to the EcoRI site in the **MCS** is shown (refer to **figure** 7.5).

(b) This diagram shows a portion of both the '**left**' (**TRP1** gene) and '**right**' (**URA3** gene) pYAC4 vector arms with a portion of a human insert. The positions of the 4 pYAC4-specific primers are indicated with the distances of each to the EcoRI cloning site at the insert-vector junction (refer also to figure 7.5)


a
pBluescript
110bp
BamHI
(Cloning site)
80bp
M13F T3
MCS
EcoRI
T7 M13R
65bp
100bp

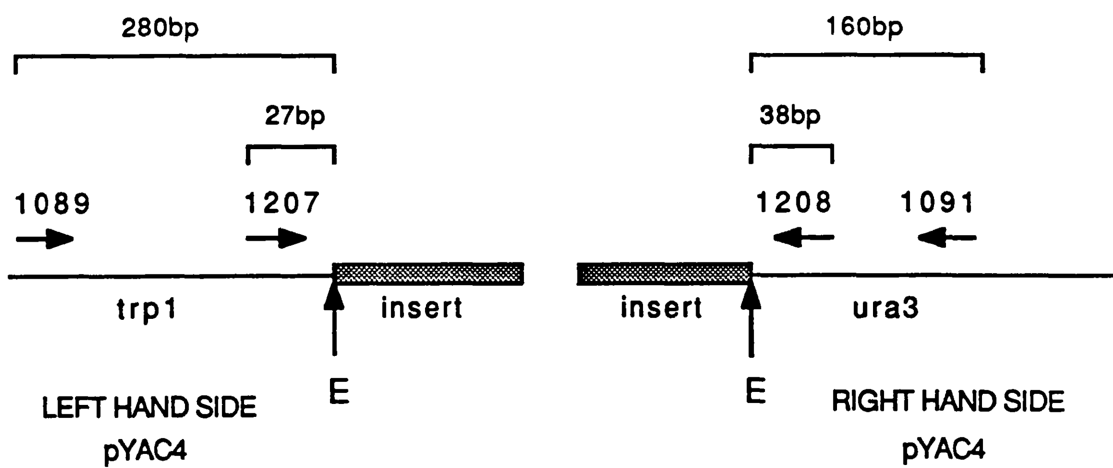
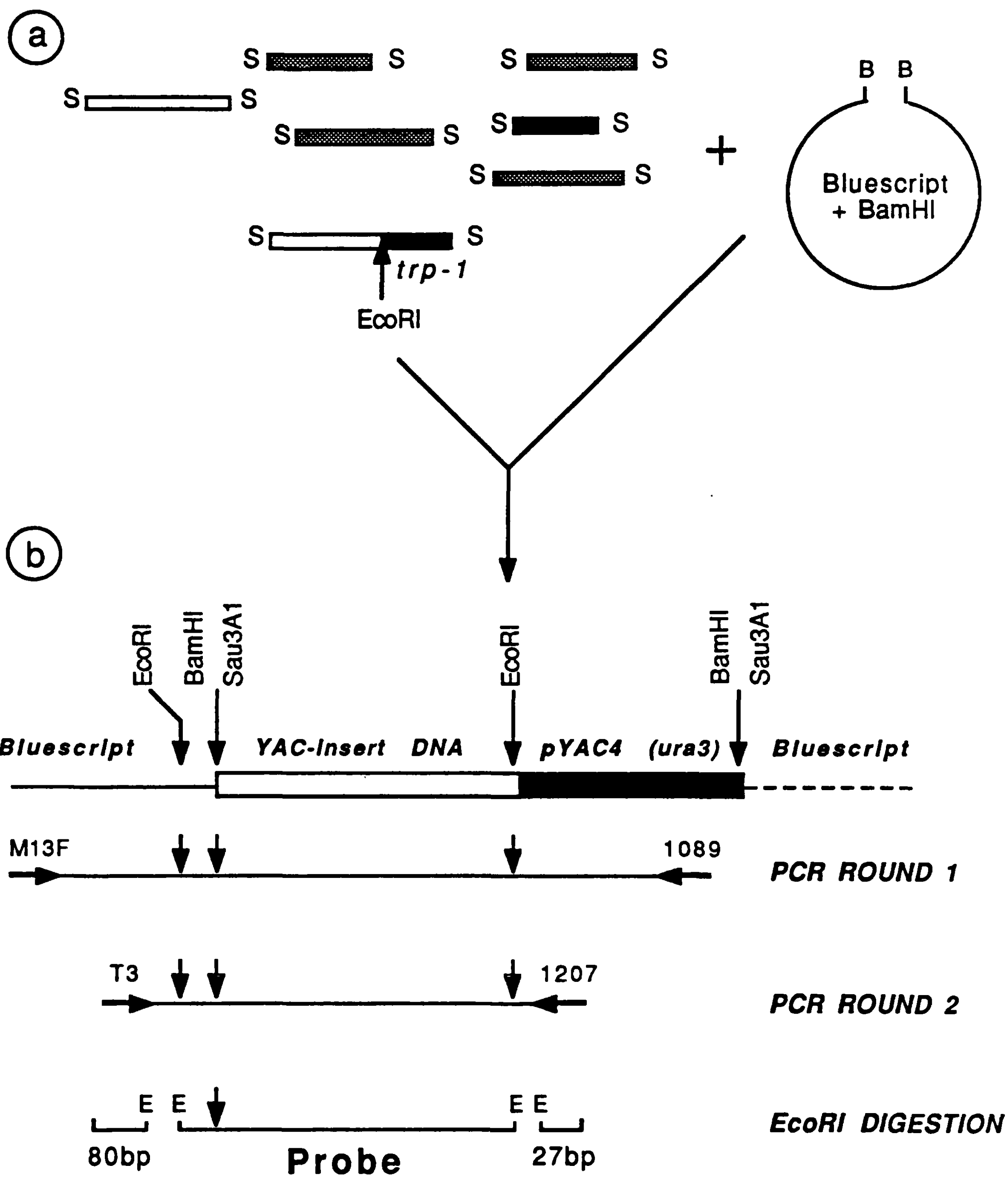

b
pYAC4
280bp
27bp
1089
1207
trp1
insert
E
LEFT HAND SIDE
pYAC4
160bp
38bp
1208
1091
insert
ura3
E
RIGHT HAND SIDE
pYAC4

Figure 7.5. **Schematic illustration of the 'junction-trapping' method**

(a) Partial Sau3A1 digestion products of whole yeast DNA (**S-S**) ligated to BamHI (**B-B**) digested pBluescript plasmid DNA. Stippled boxes represent random yeast DNA fragments, black boxes represent fragments of pYAC4 vector DNA and unfilled boxes represent human YAC-insert DNA fragments.

(b) An example of specific amplification of a junction-fragment containing a portion of the left arm (TRP1 gene) of pYAC4 and the terminal fragment of a human YAC insert. In this example the junction-fragment is amplified firstly by the **M13F** (forward) primer and the left arm pYAC4-specific primer; **1089**, and subsequently using the appropriate internal primers, **T3** and **1207**. EcoRI digestion of the products from the second round of PCR would be expected to generate 3 fragments of 80bp, 38bp and a variably sized YAC insert fragment(s) (see **figure 7.5**, lane **4**). Amplification of this junction fragment with the **M13R** (reverse) primer, and subsequent amplification with the internally nested **T7** primer would be expected to produce a similarly sized product but would release only 2 fragments upon EcoRI digestion (38bp and a variably sized YAC insert fragment) since no equivalent EcoRI sites in the pBluescript polylinker exists on the M13R/T7 side (see **figure 7.5**, lane **2**).

If a right-hand end-specific fragment was to be amplified, similar results would be expected except that a 27bp fragment would be produced as opposed to a 38bp fragment reflecting the difference in the distances from the pYAC4 primer binding sites to the EcoR1 cloning site (see **figure 7.5**, lanes **6** and **8**).


a
S
S
S
S
S
S
S
S
S
S
S
S
S
S
trp-1
EcoRI
+
B B
Bluescript
+ BamHI
b
EcoRI
BamHI
Sau3A1
EcoRI
BamHI
Sau3A1
Bluescript
YAC-Insert DNA
pYAC4 (ura3)
Bluescript
M13F
1089
PCR ROUND 1
T3
1207
PCR ROUND 2
E E
E E
EcoRI DIGESTION
80bp
Probe
27bp

7.4.2. Ligations.

The ligation conditions were chosen to maximise circularisation. This was achieved by ligating an excess of dephosphorylated insert DNA (in the order of 100 to 200ng) to a small amount of BamHI digested pBluescript plasmid (approximately 2.5 to 5ng) in a final volume of 10µl.

The ligation reactions were assessed by amplification using both of the M13 forward and reverse primers in the same reaction. Figure 7.6(a), lane 5, shows an example of this test. It can be seen that a smear was generated in the range of 250bp to at least 3kb with a peak intensity around 600bp; this is approximately twice the size of fragments expected from complete Sau3A1 digestion. The band at 220bp is due to the amplification of self-ligated pBluescript molecules; that is, the sum of the distances between the M13 priming sites and the BamHI cloning site (refer to figure 7.4).

7.4.3. Generation of candidate termini PCR fragments.

Subsequently, aliquots of the ligations were amplified using each of four possible primer combinations; M13 reverse with primer 1091 or 1089 and M13 forward with primer 1091 or 1089 (see figure 7.4). These PCR reactions (figure 7.6(a), lanes 1-5) gave rise to a large number of products varying in size from 100bp up to approximately 2.5kb, some of which were thought to represent progressively distant Sau3A1 sites in the YAC insert.

The second round of PCR utilised primers internal to the primers used in both the pYAC4 vector (1207 for 1089 and 1208 for 1091; Riley et al., (1990); figure 7.4) and the pBluescript plasmid vector (T3 for M13R and T7 for M13F; figure 7.4). The appropriate primer combinations were used for amplification. As predicted, the complexity of the initial PCR reaction products was greatly reduced such that only two or three major products were observed per primer combination (figure 7.6(b), lanes 1, 3, 5 and 7).

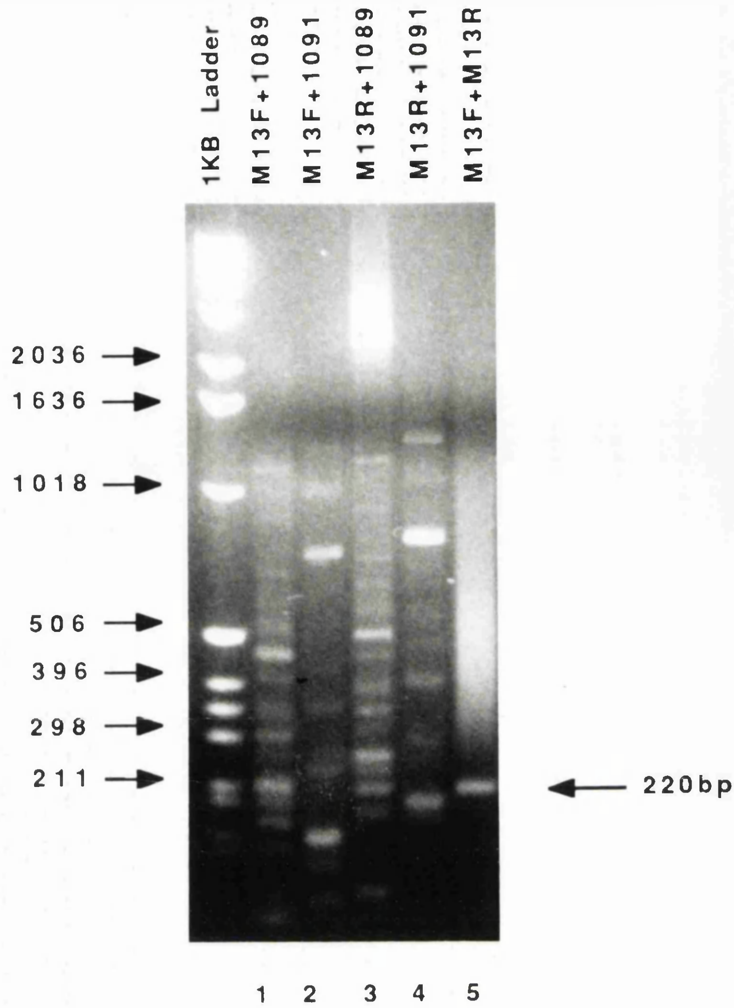
7.4.4. Assessment of candidate termini fragments.

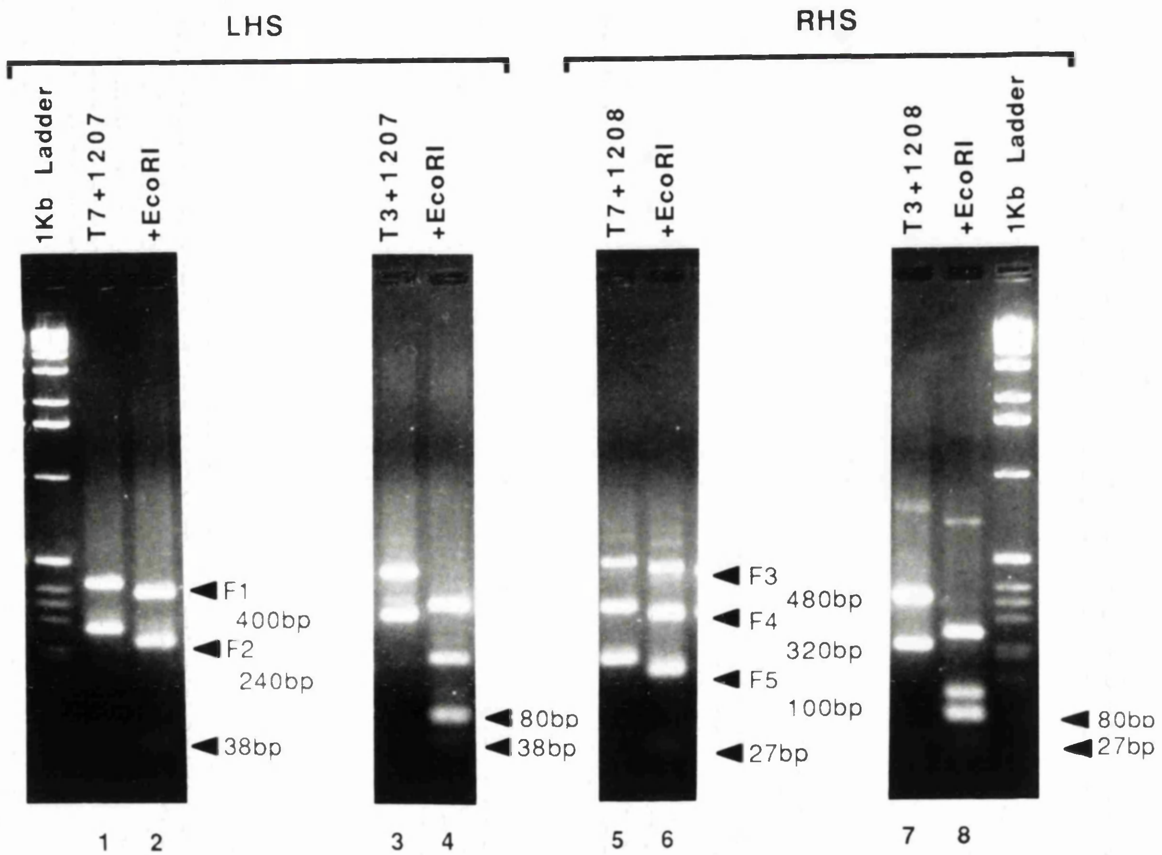
As in the *alu*-vector and 'vectorette' systems, the potentially positive YAC termini products could be further investigated by digestion at appropriate restriction enzyme sites that exist in the pYAC4 vector arms and, in this case, the pBluescript plasmid vector. Internally primed PCR reactions greatly

Figure 7.6. Generation and primary characterisation of YAC insert termini from two YACs in yeast ICRFy900B0624.

(A) First round PCR: Gel electrophoresis (1.5%) of the amplification products generated from 1µl aliquots of the ligation reactions (see text). The primer combinations used for left and right arm-specific PCR reactions are indicated above the figure (**1089** = left; **1091** = right). The **220bp** band representing amplification across self-ligated pBluescript molecules in lane **5** is indicated to the right of the figure. Note the similarities in banding patterns in lanes **2** and **4**; **3** and **5**. The lane containing the IKb ladder DNA size markers is indicated above the figure.

(B) Second round PCR: Gel electrophoresis (1.8%) of the amplification products generated from the left hand side (**LHS**) and the right hand side (**RHS**) of the YAC inserts using 1µl aliquots of the first round PCR products. The appropriate nested primer combinations are indicated above the figure (i.e. **1207** for 1089 and **1208** for 1091; **T3** for M13F and **T7** for M13R). The sizes of the EcoRI (**+ EcoRI**) digested fragments (**F1-F5**) generated in the T7+1207 and T7+1208 primer combination reactions (lanes **2** and **6**; used for assessment of YAC specificity) are indicated to the right of the figures. The **27bp** and **38bp** pYAC4 vector fragments, in addition to the **80bp** pBluescript-specific fragments from the T3 amplification products released after EcoRI digestion are denoted by arrowheads in lanes **2, 4, 6** and **8** (refer also to **figure 7.5 (b)**). Lanes containing the 1Kb ladder DNA size markers are indicated above the figure. Fragment F5, specific for the right-hand-specific terminus of 624-575, was isolated and used as a ^{32}P-labelled hybridisation probe in **figure 7.7**.


A
1KB Ladder
M13F+1089
M13F+1091
M13R+1089
M13R+1091
M13F+M13R
2036
1636
1018
506
396
298
211
220bp
1
2
3
4
5


B
LHS
RHS
1Kb Ladder
T7+1207
+EcoRI
T3+1207
+EcoRI
T7+1208
+EcoRI
T3+1208
+EcoRI
1Kb Ladder
F1
400bp
F2
240bp
38bp
80bp
38bp
F3
480bp
F4
320bp
F5
100bp
27bp
80bp
27bp
1
2
3
4
5
6
7
8

reduced the distances between the primer binding sites and the pYAC4 EcoRI cloning site (refer to the schematic illustrations in figure 7.4 and figure 7.5). However, digestion with EcoRI was still expected to produce a shift in the size of *bona fide* termini products by 38 and 27bp (including the primer sequences) from the EcoR1 site on the left and right sides of pYAC4, respectively. Further, in the plasmid vector an EcoRI site should be present 80bp distant to the T3 primer.

Figure 7.6(b) shows the products generated after EcoRI digestion ('+ EcoRI') of the second round PCR products (lanes 2, 4, 6 and 8). Careful estimation of the fragment sizes clearly indicates that all of the PCR products have been reduced by the appropriate size, whether they were amplified by the T7 or T3 plasmid primer combinations. The fragments of the pYAC4 vector representing the distance from primers 1207 and 1208 to the EcoRI cloning site can be seen clearly as approximately 40bp (lanes 2 and 4) and 30bp (lanes 6 and 8) products, respectively. The pBluescript fragments representing the 80bps of DNA from the T3 primer site to the BamHI site can be seen in addition to these fragments in lanes 4 and 8.

As predicted, other less intense products did appear in the secondary PCR reactions. These presumably represent amplification products from the pYAC4 vector to more distant Sau3A1 sites in the insert. All of those that can be seen are reduced by the appropriate size after EcoRI digestion (for example compare lanes 5 and 6; 7 and 8). These fragments are almost certainly real but were not been further investigated.

Comparison of the patterns generated by either T7 or T3, especially in combination with primer 1207 (lanes 1 and 3), shows that they are extremely similar, but shifted with respect to each other by a size exactly corresponding to the difference in the distances between each primer and the BamHI cloning site in the pBluescript plasmid. This implied that ligation of partially digested DNA to pBluescript was efficient enough to give essentially equal numbers of ligated molecules in either orientation. Although the use of either T3 or T7 alone with the pYAC4 primers should be sufficient for termini generation, it has been observed in subsequent reactions that the products appearing in one combination do not always appear consistently with the other. It is, in any case, not any more labour intensive to amplify the ligation material with all of the possible primer combinations.

7.4.5. Mapping of the candidate insert termini fragments.

Each of the EcoRI-digested PCR products resulting from the amplification using the T7-1207 or T7-1208 combinations were then used for mapping studies (that is, F1 to F5). To differentiate the YAC of origin for each of these fragments, the products were used as probes on PFGE blots containing separated B0624 yeast chromosomes. Products F1 and F2 (lane 2) mapped to YACs 624-D and 624-575, respectively. Products F3, F4 and F5 (lane 6) mapped to YACs 624-D, 624-D and 624-575, respectively. Thus, each end of each of the YACs appeared to have been isolated. Fragments F1 to F5 were then used as probes on SacII and MluI partial digestions of this yeast. Figure 7.7 shows the results of an experiment using fragment F5 (that is, the right end of 624-575) which is compared to a hybridisation experiment using a pYAC4 right-arm specific probe. It can be seen that fragment F5 hybridised only to the 624-575 YAC (compare the 'undigested' lanes).

Each of the fragments were then used as probes on the chromosome 5 mapping panels. Only products F1, F2 and F3 could be unequivocally mapped. As expected, neither products F1 or F3 mapped to chromosome 5, while product F2 clearly mapped to interval II. Hybridisation of this product to EcoRI digest panels of the interval II YACs identified A1010 as well as B0624 confirming the suspected overlap of these two YACs at the *MCC* gene. Isolation of these termini was particularly useful for the construction of accurate restriction maps of the 624-575 YAC (figure 6.2(b), Chapter 6) and for assessing the degree of inter-YAC overlap on multiple digests of the A1010 YAC (figure 8.3, Chapter 8).

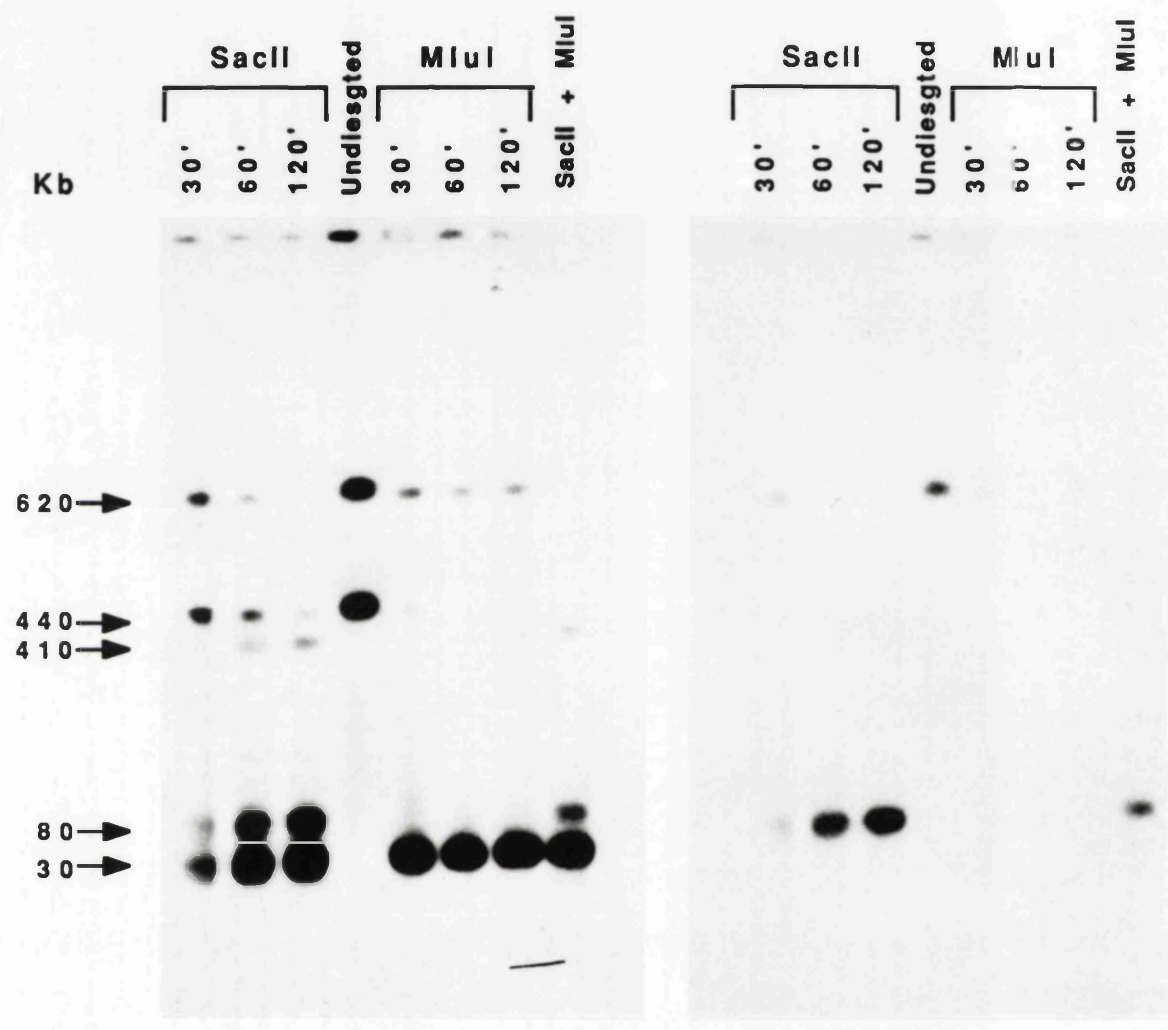
7.4.6. 'Junction-trapping' as a useful strategy

Using this method, it has been possible to isolate candidate fragments for all of the YAC termini investigated (summarised in table 7.1 at the end of this chapter). Each of these termini behaved exactly as predicted after EcoRI digestion and were investigated as described for the B0624 YACs to establish the candidates as *bona fide* insert termini. In all cases, the fragments hybridised to the YACs in question and to the appropriate SacII and MluI terminal restriction fragments. However, unequivocal chromosomal mapping of some of the candidate fragments was difficult due to their small size. Of the 4 termini of interest that could be mapped; the right end of YAC G016 (in addition to the left end of 624-575; see above) clearly mapped to

Figure 7.7. **'Fragment-5' maps to YAC 624-575.**

(A) Hybridisation of a pulsed field gel (PFG) filter containing **SacII** and **MluI** partial digests of yeast ICRFy900B0624 with a ^{32}P-labelled probe specific for the right-hand pYAC4 vector arm (**pBR322-RHS**). One-third agarose plugs were digested with a constant amount of enzyme (20U SacII; 15U MluI) for 30, 60 and 120 minutes as indicated above the figure. A one-third block digested with 20U of SacII and 15U of MluI for 4 hours is denoted as **SacII + MluI** above the figure. The lane marked **undigested** represents a one-half block incubated with no enzyme. The gel (1%) was electrophoresed at 5V/cm for 20 hours with a linear switching time of 20 to 84 seconds and processed according to the methods outlined in Chapter 2. Note that YAC 624-575 does not contain an MluI restriction site whereas YAC 624-D does. The sizes of the hybridisation signals shown to the left of the figure were calculated by co-electrophoresis of λ multimers on the agarose gel (not shown).

(B) The filter in **(A)** above was stripped and hybridised with EcoRI digested and ^{32}P-labelled **fragment-'5'**. Note that fragment 5 only recognises the undigested 624-575 YAC (at 620kb). Partial SacII fragments of YAC 624-575, which have been lost during reproduction, could be observed after extended exposure of this filter (see **figure 6.2(b)**).



A

probe: pBR322-RHS

B

probe: Fragment-'5'

chromosome 5 in interval II. Neither of the D027 insert termini, however, mapped to chromosome 5. This was not unexpected in view of the physical mapping and FISH analyses (see table 5.3, Chapter 5).

7.5. Results 3. Coincidence mapping with YAC insert termini.

All of the insert termini isolated by the methods described in this chapter (including those which were known not to map to chromosome 5; see table 7.1) were used as probes on the coincident mapping filters (see Chapter 6). Each of the probes, except the right end of y12.75, detected the YAC of origin but none of the YACs isolated with independent probes. The overlaps inferred by physical mapping of the G016 and A0714 YACs (figure 6.2(a), Chapter 6) were confirmed by hybridisation of the internal YAC insert termini (that is, the left end A0714 and the right end of G016) as was the overlap predicted for YACs A1010 and 624-575 (see above). However, no internal contig overlaps could be inferred for the D073 YAC, which, as commented above, was not unexpected.

7.6. Conclusions.

The method chosen to assess potential inter-YAC overlap was based on the isolation of YAC insert termini, particularly for those YACs that were thought to represent the most distal points of contigs inferred from YAC hybridisations and physical mapping. That no coincident overlaps were detected was disappointing but not surprising based on the apparently high degree of chimerism in the YACs. In total, it was possible to map 8 of the 14 termini investigated. Of these, only 4 mapped to chromosome 5 (50%). It is probably not appropriate to compare statistics of this sort; it would require the successful generation of termini fragments from a much larger random collection of YACs and, unequivocal mapping all of these either by hybridisation or PCR.

However, the percentage of chimerism observed in this study using the ICRF YAC library (that is 50% of the termini investigated) is in line with detailed investigations of 6 different YACs localised to the Ewing's Sarcoma breakpoint region on chromosome 11 (Dr. K. Patel, ICRF, personal communication).

Table 7.1. Isolation and characterisation of YAC insert termini.

(1) *Alu*-vector PCR

YAC	Terminus	EcoRI digestion[a]	Size (bp)	PFGE[b]	Ch. 5 ?	Overlap[c]
A1010	'Left' (LHE)	+	850	+	+	-
A0714	LHE	+	450	+	+	+; G016

(2) Vectorette-PCR

YAC	Terminus	EcoRI digestion	Size (bp)	PFGE	Ch. 5	Overlap
A10109	LHE (a)	+	1500	+	-	-
	LHE (b)	+	1000	+	-	-
	RHE (a)	+	1000	+	?	-
	RHE (b)	+	700	+	?	-
D073	LHE	+	350	+	-	-

(3) Junction-trapping

YAC	Terminus	EcoRI digestion	Size (bp)	PFGE	Ch. 5	Overlap
A0714	RHE	+	200	+	?	-
G016	LHE	+	100	+	?	-
	RHE	+	300	+	+	+; A0714
D027	LHE	+	1000	+	-	-
	RHE	+	800	+	-	-
y12.75	LHE	+	300	nt[d]	?	-
	RHE	+	125	nt	?	?
B0624 (575)	LHE	+	240	+	+	+; A1010
	RHE	+	100	+	?	-

a + indicates candidate products that released the appropriate sized pYAC4 vector fragments after EcoRI digestion

b + indicates candidate products that recognised the appropriate SacII and Mlu1 restriction DNA fragments in the YAC of origin

c Overlap was detected by hybridisation to EcoRI digested YACs (see Chapter 6, table 6.1)

d nt; not tested

The results of the studies described here led to the following conclusions: Firstly, isolation of YAC insert termini may be an effective method for the detection of coincident overlap. However, such a high level of YACs containing sequences non-contiguous in the genome interfered with the expected success of this approach. Nonetheless, in contrast to the pooled *alu*-PCR experiments commented upon in the introduction, the overlap of the B0624 and A1010 YACs was easily confirmed by insert termini isolation.

Results pertaining to one YAC (A10109) suggested that, in general, terminus isolation may be most effective if complemented by other methods for overlap detection. Sequences from the 'left' end of this YAC do not map to chromosome 5 but, interestingly, the probe with which the YAC was isolated, i.e., cL5.4.s-1, maps to the terminal 50kb SacII restriction fragment at the 'left' end of the insert (see figure 6.2(e), Chapter 6). Thus, a rearrangement must have occurred somewhere within this small 50kb region and distant to the cL5.4.s-1 probe. This type of YAC insert containing a sequence non-contiguous with chromosome 5 was unexpected but similar types of YAC insert have been observed a number of times during the construction of a YAC contig across the human HLA class II gene cluster (Ragoussis et al., 1991, Dr. I. Ragoussis, ICRF, personal communication).

The experience gained with a number of methods used for the isolation of YAC insert termini led to the successful development of a new technique that greatly simplifies this process and largely eliminates most of the disadvantages of other methods. The technique proved to be especially favourable since the reagents used are common to most laboratories and are relatively inexpensive. It was possible in all of the cases described here and in each of 6 six cases examined by others (Dr. K. Patel, ICRF, personal communication) to generate candidates for each YAC insert termini investigated.

For the purposes of this study it was intended to investigate YAC insert termini candidates, initially to the level of coincident YAC overlap detection, although it was anticipated that some genomic 'walking' would be initiated. This was not pursued, however, as the construction of a YAC contig across Interval II became irrelevant for the identification of the *APC* gene (see Chapter 8). A number of the products initially examined were small in size and therefore not suitable for rigorous genomic mapping. For a number of the YAC inserts other larger fragments, most likely representing more distant

Sau3A1 sites in the YAC insert, appeared in addition to the intense products (see figure 7.6(b), lanes 5 and 6; 7 and 8). These can be isolated and re-amplified for the purposes of chromosomal and physical mapping experiments (Mr. T. Ward, in progress). In two cases (the right end of y12.75 and the right end of G016), however, it was impossible to detect such fragments even after repeating the experiment. Modification of the PCR reactions to cope with longer DNA fragments may alleviate this problem (e.g., see Ponce and Micol, 1992).

Appendix to Chapter 7.

Physical mapping summaries.

Appendix to Chapter 7.

Physical mapping summaries.

A7.1. Introduction.

The information described in this sub-section is intended to provide an integrated summary to the physical distributions of the microclones and YACs described in this thesis. The localisation of the microclone inserts and YACs has been derived from experiments aimed at the construction of a genomic physical map of the *APC* gene region (kindly provided by Mr. T. Ward) and on the basis of the coincident YAC mapping data presented in Chapter 6. At the time of writing, a nearly complete genomic physical map of the minimal region (Interval II) described in Chapter 4 has been completed (see figure A7.1) facilitating this retrospective summary. However, at this time, two breaks in the physical map exist. It is not known how much DNA is involved because of the relative paucity of markers (and identified large restriction fragments) at these positions.

A7.2. Distribution of microclones.

Twelve of the 14 microclones mapped to interval II (see figure 4.3, Chapter 4) have been physically localised; 8 on the basis of YAC isolation or coincident hybridisation assays and 4 by genomic physical mapping using cosmid sub-fragments obtained for each of these microclones. The distribution of these probes is depicted at the top of figure A7.1.

It can be seen that 8 of these microclones are localised to the telomeric region of interval II (and therefore distal to *APC*) defined by the break in the physical map. All of these probes, except mc411, are distinctly clustered:

(1) Clones mc545 and mc449 have been found to reside on a single cosmid, although an additional cosmid has been isolated for mc449. The overlap between these two cosmids has suggested that the maximal distance between these two clones is about 30kb.

(2) As described in Chapter 6, three independent microclones; mc289, mc404 and mc440 were found to cluster on the YACs identified by probe YN5.48. Hybridisation of these three clones to SacII and MluI total digests of these

YACs has shown that they are all sub-localised to an internal 50kb SacII fragment.

(3) Two microclones - mc14 and mc451, reside on the same two YACs (G016 and A0714; contig 1, figure 6.2(a), Chapter 6). Interestingly, these two clones bound the site of a major CpG island in the genomic physical map (also detected in the YACs). These clones were isolated independently from libraries I and II, respectively, and their relative distribution was unexpected.

Possible reasons for why such clusters of microclones should exist are addressed in detail in Chapter 9.

The other 4 physically localised microclones; mc575, mc241, mc534 and mc434 map to the the first half of the proximal region of interval II and do not appear to be clustered. Interestingly, two of these clones, mc575 and mc241 flank the *APC* gene at a maximal distance of about 1Mb; the other two clones being separated by approximately 600kb. Clone mc434 maps to the same 120kb internal SacII fragment as probe p3.1 (Spurr et al., 1991) in YAC D027. This other independent probe; p3.1 also appears associated with a major CpG island which has been detected in unmethylated cosmids clones around this region (Mr. T. Ward, personal communication).

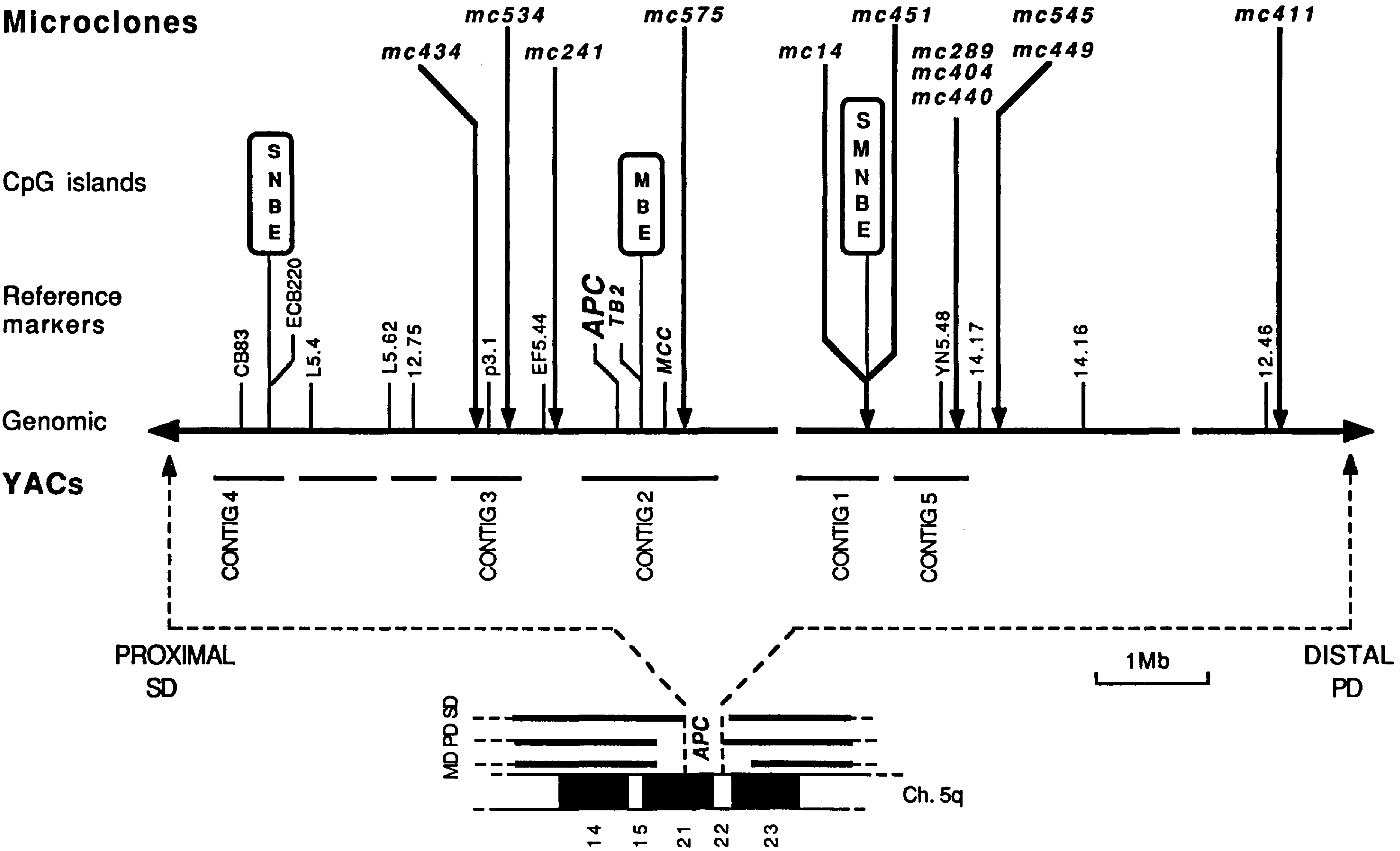
A7.2. Distribution of YACs.

Figure A7.1 also shows the distribution of the YACs isolated in this and other studies in the *APC* gene region (depicted in the lower part of the figure). It has been possible to localise the two YAC contigs represented by G016 and A0714 (Contig 1), in addition to B0624 and A1010 (Contig 2) on the genomic physical map by virtue of CpG islands detected in the YACs and the genome of the cell lines used.

However, it has not been possible to rigorously localise most of the other YACs, either because of chimerism (and the failure to align the YACs in the first instance) or because of the genomic methylation of distinct restriction sites that are present in the YACs. Thus for the remaining YACs, the physical localisations are based solely on the position of the probes with which they were selected. The failure to detect coincident overlap in the YACs (apart from chimerism), is not surprising with reference to the genomic physical map. It is apparent from this map that as much as 40 to 50% of Interval II has

Figure A7.1. Summary of physical mapping in Interval II.

The map depicted in this figure illustrates the physical distributions of the microclones and YACs that have been mapped to interval II. The thick black line, demarcated by the arrows in the centre of the figure, represents the genomic DNA in interval II (centromere to telomere) as defined by the proximal breakpoint of the SD deletion and the distal breakpoint of the PD deletion (refer to the idiogram at the bottom of the figure). The map was constructed by PFGE analysis of large restriction fragments (not shown for reasons of clarity), principally recognised by the **reference makers** indicated at the top of the line (Mr. T. Ward, personal communication), as well as by coincidence hybridisation assays on YACs mapped to this interval (see Chapter 6). The two gaps in the physical map are illustrated by breaks in the continuous genomic line. The three key **CpG islands** recognised in genomic DNA are indicated in the boxes, where E = EagI; B = BssHII; M = MluI; S = SacII, N = NruI. The positions of the **microclone** sequences, derived either by hybridisation of the insert sequences or cosmid sub-fragments are shown by the pointed heavy lines. The **YACs** and **YAC contigs** are depicted by the lines underneath the genomic DNA. For reasons of clarity, individual YACs within the contigs are not shown. Instead, the overall distance represented by the YACs, taking into account the known chimeric parts, are shown as a continuous line. The YACs belonging to each contig are as described in section 6.2.3., Chapter 6. YAC contigs 1 and 2 could be aligned by virtue of common restriction sites in cloned YAC and genomic DNA. The centromere/telomere orientations have been ascertained by PFGE hybridisations using YAC insert termini in both cases (Mr. T. Ward). The positions of the other YACs and YAC contigs are approximate.



Microclones
CpG islands
Reference markers
Genomic
YACs
mc434
mc534
mc241
mc575
mc14
mc451
mc289
mc404
mc440
mc545
mc449
mc411
S N B E
M B E
S M N B E
CB83
ECB220
L5.4
L5.62
12.75
p3.1
EF5.44
APC
TB2
MCC
YN5.48
14.17
14.16
12.46
CONTIG 4
CONTIG 3
CONTIG 2
CONTIG 1
CONTIG 5
PROXIMAL
SD
1Mb
DISTAL
PD
MD PD SD
APC
Ch. 5q
14
15
21
22
23

been cloned in YACs (with chimerism taken into account) and, that the original estimate of 5Mb for this region was incorrect. The region, in fact, is probably closer to 8 to 9Mb in size.

Chapter 8.

Identification and characterisation of the *Adenomatous Polyposis Coli* (*APC*) gene-region.

Chapter 8

Identification and characterisation of the *Adenomatous Polyposis Coli* (*APC*) gene-region.

8.1. Introduction.

This chapter describes a number of experiments undertaken for the identification of the genomic region containing the *APC* gene.

One of the global strategies undertaken by the ICRF Polyposis Group was the identification of small sub-microscopic deletions in the *APC* gene region; either in the constitutional DNA of about 50 independent FAP patient cell lines or as somatic alterations (for example, homozygous deletions) in a collection of 34 colorectal tumour derived cell lines. At the time of these studies, no such alterations had been detected in any of the cell lines either by PFGE or standard Southern blot analyses using a large number of marker loci sub-localised to Interval II (my own studies and Ms. Sally Cottrell, unpublished).

The supposition of detecting such alterations was based on the high mutation rate at the *APC* locus (Reed and Neal, 1955; Veale, 1965; see also, Chapter 1) which suggested that the *APC* gene might cover a large genomic region or was a target of mutation by virtue of some sort of sequence 'hot-spot', exactly analogous to the hypotheses put forward prior to the cloning of the Duchenne Muscular Dystrophy gene (Monaco and Kunkel, 1986). In line with the types of sub-microscopic alterations found in the Duchenne (Monaco and Kunkel, 1986) and Retinoblastoma (Goodrich and Lee, 1990) genes, the same variety of mutational types were expected in the *APC* gene region. At the time only large constitutional deletions had been observed (Herrera et al., 1986; Varesco et al., 1989; Hockey et al., 1990) but, the detection of smaller deletions at the molecular level was simply thought to be dependent on having markers close enough, or actually within the gene (e.g. see Koenig et al., 1987).

These suppositions were strengthened by the detection of such deletions (albeit at low frequencies) as somatic alterations at the Wilm's tumour (WT-1) locus (Gessler et al., 1990) and as constitutional alterations at the

Neurofibromatosis type-1 (NF-1) locus (Viskochil et al., 1990). Such alterations greatly facilitated the recent positional cloning of these genes.

Although *MCC* represented a strong candidate for *APC* at the time of publication (Kinzler et al., 1991a), indirect evidence, in addition to the lack of mutations described in FAP kindreds, suggested that the *APC* gene had yet to be identified:

Firstly, cDNAs from the *MCC* gene (provided by Prof. Bert Vogelstein; Kinzler et al., 1991a) were used to screen standard Southern blot filters containing the 50 independent FAP patients and 34 colorectal cell lines described above (my own studies and Ms. Sally Cottrell). Only one apparent alteration was identified with a probe representing nucleotides 1679-1862 in an FAP patient of Swedish origin. This was thought to be a rare polymorphic variant[1] after examination of the patient's DNA with a variety of other restriction enzymes. As discussed above, it was proposed that a lack of rearrangement at the *APC* locus would be unlikely.

A personal communication from Dr. Yusuke Nakamura (Howard Hughes Medical Institute, Utah; presently at Cancer Research Institute, Tokyo) to Sir Walter Bodmer (ICRF) suggested that a constitutional alteration had been observed in a patient with polyposis. This alteration was thought to represent a 260 kb deletion in one allele of chromosome 5, region 5q21-q22, detected on the basis of variantly sized PFGE fragments using a random marker, L5.71 (marker cited in Dunlop et al., 1990; Kinzler et al., 1991a). This putative deletion was shown to be inherited from the father in a two generation FAP pedigree and observed with a number of other infrequently cutting restriction enzymes. Lymphoblastoid cells from this patient (3214), were fused to rodent cells and both homologs of chromosome 5 segregated in two human-rodent somatic cell hybrids; HHW1155 and HHW1159 (Dr. J. Wasmuth, cited in Joslyn et al., 1991; see Chapter 2, appendix 3). Screening of the *MCC* cDNAs on Southern blots containing these hybrids (kindly provided by Dr. J. Wasmuth) failed to detect this putative deletion which, although by no means definitive, did suggest that a search for other genes should be undertaken. The L5.71 probe that detected the putative deletion in patient 3214, also revealed a dramatically increased frequency of allele-loss around

---

1 allele frequencies were calculated at 0.99 and 0.01 in a sample of 100 chromosomes from 50 independent FAP patients. This can be assumed to be an essentially random population sample

this local region relative to other closely spaced markers (Ashton-Rickardt, 1991; Kinzler et al., 1991). This therefore suggested that genes should be sought in the genomic region close to *MCC*.

Initial physical analysis of the YACs identified by *MCC* exon fragments 'a' and 'b' (see figure 6.2(b), Chapter 6), demonstrated that the 1.1Mb contig represented by YACs A1010 and B0624 overlapped at the *MCC* gene itself, with approximately 500kb of DNA flanking the gene. This large amount of uninvestigated DNA was thought to represent a reasonable sample of the genomic region around *MCC* for the detection of new genes to assess as candidates for the *APC* gene. No evidence was available at the time to indicate in which relative direction to search preferentially. Thus, the strategies considered were dependent on the efficient random sampling of *both* YACs.

The strategy chosen was to use the YACs as hybridisation probes on the gridded chromosome 5 cosmid library filters (see Chapter 3, part II). This would avoid any of the YAC artifacts such as chimerism and yeast co-transformation that would certainly compound the analysis of bacteriophage or cosmid libraries constructed from whole YAC-containing yeast DNA, and, in addition, provide an immediate resource of usable clones solely derived from chromosome 5. Unpublished data concerning the use of this method for sampling YACs in the Huntington's disease region suggested that a reasonable coverage of the YACs could be expected (Dr. G. Bates; Ms. S. Baxendale, Genome Analysis Laboratory, ICRF, personal communications). Cosmids revealing any potential alterations on chromosome 5 in FAP patient genomes would then be used for isolation of sequences conserved across species (Monaco et al., 1986; Rommens et al., 1990; Fearon et al., 1990), and ultimately to select cDNA clones representing genes in the region that could be assessed as potential candidates.

8.2. Results.

8.2.1. Direct YAC screening on chromosome 5-specific cosmid libraries.

The experiments described in this sub-section were carried out jointly with Mr. T. Ward and Ms. K. Howe. The B0624 and A1010 YACs were isolated from pulsed field gels and the DNA purified. The YAC DNAs were then labelled with 50μCi of [α^{32}P]-dCTP and extensively competed with sheared

human placental DNA. These probes were hybridised to the gridded chromosome 5-specific cosmid library essentially as described by Baxendale et al., (1991). YAC B0624 was screened once over a total of about 9000 cosmid clones while A1010 was screened twice over a total of about 18000 cosmid clones. The results of these experiments are summarised in table 8.1(a).

Hybridisations of YACs B0624 and A1010 resulted in the identification of 25 and 29 cosmids, respectively. Both of these YACs identified an identical sub-set of 7 cosmids, presumably from the region of inter-YAC overlap. In addition, 2 of these 7 seven cosmids from each YAC had been identified by a cDNA probe; SW15, representing nucleotides 133 to 1918 of the *MCC* gene. A further 4 cosmids identified by YAC A1010 were found to be in common with those seen by cDNA probe MCC40cI (nucleotides 1634 to 3969). These results agreed well with the predicted overlap between the two YACs determined by PFGE analysis (see Chapter 6). Subsequently, YAC A1010 was re-screened over an additional 9000 cosmid clones (see table 8.1(a)). Of 14 identified cosmids, 6 were also identified by sequences from the *MCC* gene. It was not possible to determine whether any of the cosmids were common to the B0624 YAC as the filters screened represented an entirely different set of cosmids. The cosmids from the YACs outside the regions of overlap (and not in common with the *MCC* gene region) would provide a resource for the isolation of sequences that could be used to identify the *APC* gene, viz-a-viz the identification of sub-microscopic genomic alterations in FAP patients.

8.2.2. Direct cosmid screening.

The experiments described in this sub-section were carried out by Ms. S. Cottrell and myself. It was decided to screen the cosmids, initially at least, across a somatic cell hybrid panel containing the segregated chromosomes 5 from FAP patients suspected to contain sub-microscopic deletions in the *APC* gene region:

(1) Hybrids HHW1155 and HHW1159 from patient 3214 who was suspected to harbour a 260kb interstitial deletion in the 5q21-q22 region on the basis of variantly sized PFGE fragments (Joslyn et al., 1991; see also introduction to this chapter). Prior to these studies, none of the probes available to the ICRF Polyposis Group, including sequences from the *MCC* gene had been found to identify or map into this putative region by Southern blot analysis (Ms. S.

Cottrell, unpublished). The two hybrids were known to segregate one of each of the chromosome 5 homologs (Dr. J. Wasmuth, personal communication)

(2) Hybrids JT/Ts-1 and JT/Ts-3 from patient JT (Thomas, 1991; Hampton et al., submitted) exhibiting FAP and mental retardation (Chiba et al., 1990). The association of mental impairment and FAP, seen previously in three Caucasian FAP patients (Herrara et al., 1986; Varesco et al., 1989; Hockey et al., 1990) suggested that this patient may also have a sub-microscopic deletion of chromosome 5.

(3) The somatic cell hybrid, SD/Ts-1 (see Chapter 4), was also included to assess each cosmid for appropriate genomic localisation.

In total, 44 cosmids were used for direct screening; 36 and 8 from the first and second YAC hybridisation experiments, respectively. The results of these experiments are summarised in table 8.1(b). From the first 36 cosmids, only 17 were found to be absent in the SD deletion (8 and 9 from YACs B0624 and A1010, respectively). This result was unexpected since some of the cosmids found to map outside the SD deletion were common to both YACs (see conclusions to this chapter). However, each of the 8 cosmids from the second A1010 hybridisation experiment were found to be absent in the SD deletion.

8.2.3. Identification of an interstitial deletion.

A total of 25 cosmids from both YACs were shown to be absent in the SD deletion and were therefore presumed to represent sequences specific to the YACs (see table 8.1(b)). No aberrant patterns were detected in either of the somatic cell hybrids from patient J.T., raising a question at to the nature of the mutation in this patient (see Chapter 9). Six of the cosmids derived from the A1010 YAC, however, were found to exhibit aberrant hybridisation patterns in hybrid HHW1155 (see figure 8.1). One of these 6 cosmids, ym75, appeared to be totally absent in the HHW1155 hybrid and represented, therefore, a very strong candidate for containing sequences of the *APC* gene. The other five cosmids, ym8, ym21, ym64, ym72 and ym73, exhibited hybridisation patterns suggestive of partial deletion in that most of the bands revealed in the normal homolog of chromosome 5 from patient 3214 were absent in hybrid HHW1155 DNA. (One of these cosmids, ym73, has since been interpreted as recognising a polymorphism with at least two variant alleles of 6.5 and 4.4kb).

Table 8.1(a). **Hybridisation of YACs on chromosome 5-specific cosmid libraries.**

YAC probe	Nos. cosmids	Nos. common to both YACs	Nos. cosmids positive for *MCC*
A1010	29 [a]	7	6
	14 [b]	nd [c]	6
B0624	25	7	2

Table 8.1(b). **Direct screening of YAC-identified cosmids on hybrid panels.**

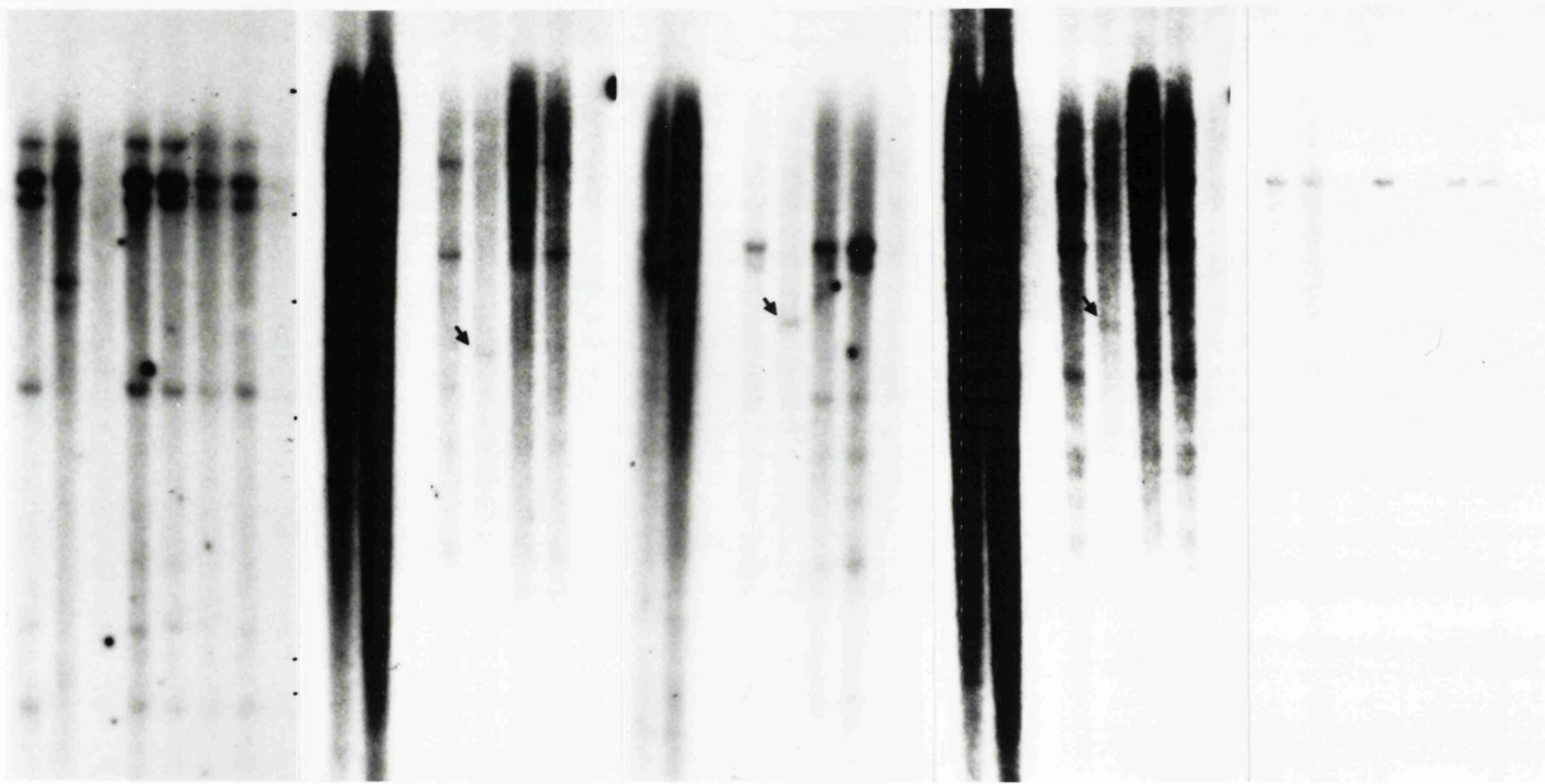
YACs	Nos. cosmids screened	Nos. deleted in hybrid SD/Ts-1	Nos. deleted in in hybrid HHW1155
A1010	21 [d]	8 (40%)	2
	8 [e]	8 (100%)	4
B0624	15	9 (60%)	0

a Derived from screening ≈9000 clones (1–≈9000)
b Derived from screening an additional ≈9000 clones (≈9000–≈18000)
c nd; not determined. The B0624 YAC was not screened on this section of the library

d From the first cosmid library screen
e From the second library screen

Figure 8.1. **Identification of an interstitial deletion in patient 3214.**

This figure shows examples of whole cosmid and cosmid sub-fragment hybridisation onto panels **(1 to 5)** containing somatic cell hybrids segregating chromosomes 5 from FAP patients JT and 3214. The identities of the EcoRI digested genomic DNAs used in the panels are indicated above each figure: Human **(Hu)**, Old World Monkey **(Mo)**, Hamster **(Ha)**, Hybrid HHW1159 **(1159)**, HHW1155 **(1155)**, JT/Ts-1 **(JT1)**, JT/Ts-3 **(JT3)** and SD/Ts-1 **(SD)**. Panels were hybridised with cosmids ym2 **(1)**, ym8 **(2)**, ym21 **(3)**, ym72 **(4)** and a subclone derived from cosmid ym72 **(5)**. The sizes of bacteriophage lambda DNA digested with HindIII (co-electrophoresed with the digested DNAs) are indicated to the left of the figure. The 'altered' sized restriction fragments revealed in hybrid HHW1155 by cosmids ym8 (5.1kb), ym21 (5.9kb) and ym72 (5.9kb) are indicated by the **arrows**. Cosmids and the cosmid sub-fragment were pre-annealed with human placental DNA prior to hybridisation. The filters were washed to a final stringency of 0.2 x SSC/0.1% SDS at 65^0C and autoradiographs were exposed overnight to 3 days.


1
2
3
4
5
Kb
Hu
Mo
Ha
1159
1155
JT1
JT3
SD
23.1
9.4
6.5
4.3
2.3
2.0

Subclones from ym72, in addition to one of the 'partially' deleted cosmids, ym8, were both found to be entirely absent from HHW1155, confirming the presence of a deletion which overlaps with YAC A1010. Since neither of the *MCC* cDNA probes (SW15 and MCC40cI) were found to be visibly rearranged in the HHW1155 hybrid cell line, it was reasoned that the deletion must map distant to the *MCC* gene and toward the 'left' pYAC4 vector arm of the A1010 YAC.

Comparison of the hybridisation patterns from the 'partially' deleted cosmids suggested that three of the probes, ym21, ym64 and ym72, were overlapping as they shared some of the 'normal' bands seen in the other homolog of chromosome 5 from patient 3214 (i.e., HHW1159) and in the J.T.-derived cell lines. Each of these probes also revealed a single altered sized restriction fragment of 5.9kb in hybrid HHW1155, suggesting, *a priori*, that they recognised the breakpoint of this deletion from one, or other 'side'. Cosmid ym8, however, did not share any of the restriction fragments in common to the other cosmids, and, in addition, revealed an altered sized restriction fragment of 5.1kb, shown repeatedly to be different from the 5.9kb band seen by the overlapping set of cosmids (see conclusions to this chapter).

8.2.4. Identification of the *APC* gene.

At this time, a number of genes in the immediate *APC* gene region were identified by B. Vogelstein, Y. Nakamura, R. White and colleagues (Kinzler at al., 1991b; Joslyn et al., 1991). Each of three genes, TB2 (or DP.1), SRP19 and *APC* (or DP2.5), identified by both groups, were assessed for mutations segregating with FAP. One of these genes, *APC* (or DP2.5), was found to exhibit mutations segregating with polyposis in a number of FAP kindreds (Nishisho et al., 1991; Groden et al., 1991, respectively), confirming it as the *APC* gene responsible for the polyposis phenotype. The identification of these genes is discussed further in Chapter 9.

The coding region of the *APC* gene is 8972 nucleotides in length and contains 15 exons (Kinzler et al., 1991b; Joslyn et al., 1991). This gene was thought to encompass some 125 kb of genomic DNA on the basis of physical mapping (Joslyn et al., 1991). Except for nucleotides representing the first exon, the entire gene (including TB2 and SRP19) was reported as deleted in hybrid HHW1155 from patient 3214 (Joslyn et al., 1991).

8.2.5. A detailed YAC-based physical map of the *APC* gene region.

The cosmid mapping experiments clearly demonstrated that at least part of the 3214 deletion must overlap the A1010 YAC. It was not clear, however, whether the entire *APC* gene (in addition to the other genes mapped close to *APC*) would be contained within the YAC.

The restriction map of the A1010 YAC presented in figure 6.2(b), Chapter 6 was thought to be too sparse in terms of infrequently cutting restriction sites to place cosmid fragments or cDNA probes except to within large, and therefore crude physical distances. Thus, further digestions were performed with this YAC to provide a sufficient density of sites for regional mapping. Two approaches were taken to achieve this. Firstly, the YAC was digested to completion with SacII, MluI, NruI and BssHII individually and in all possible combinations. Secondly, extensive partial digestions were performed with each of these enzymes, as described in Chapter 6.

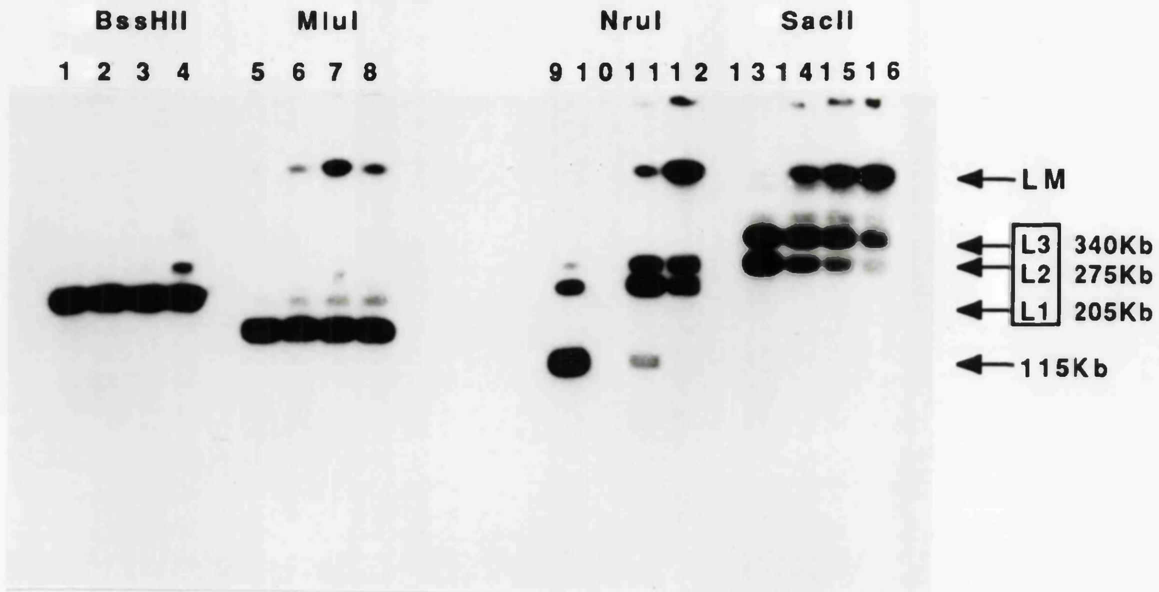
Southern blots of these yeast digestions were then hybridised with left- and right-arm specific pYAC4 probes. The partial digests allowed the construction of a detailed and continuous map across the A1010 YAC as determined from both of the pYAC4 vector-specific probes (see figure 8.2 for the derivation of the basic map). The filters were then sequentially hybridised with the following probes (kindly provided by Prof. B. Vogelstein; see Chapter 2, appendix 2): (1) SW15 and MC40cI, representing nucleotides 133 to 1918 and 1634 to 3969 of the *MCC* gene, respectively, (2) a cDNA representing the entire coding sequence of the TB2 gene, and (3) FB9A and FB54D, representing nucleotides -22 to 2705 and 6640 to 8954 of the *APC* gene, respectively.

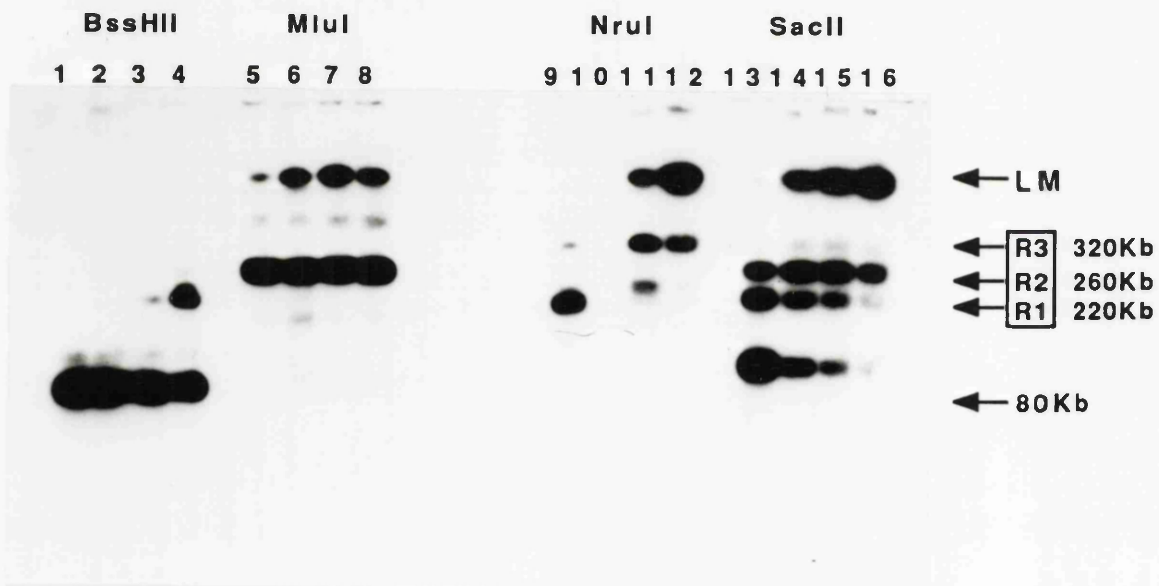
For each of these probes, detailed maps were constructed using the filters containing the single and double total restriction enzyme digestions. The probes were also hybridised to the partial digest filters but, since the probes represent cDNAs (where exons will not necessarily be contiguous across sequential restriction fragments) these results were used only to confirm the total digest maps. Each of the maps constructed in this way were then aligned with the partial restriction map. In all cases, the 'sub'-maps agreed well with the overall map in terms of genomic distance. This complete map is shown in figure 8.3.

Figure 8.2. Construction of a detailed physical map of YAC A1010.

Hybridisation of a filter containing partial digests of YAC A1010 with probes specific for the 'left' **(A)** and 'right' **(B)** pYAC4 vector arm. The enzymes used for partial digestions are indicated above the autoradiographs. One third blocks were digested with varying amounts of enzyme for 1 hour. **BssHII** - 4U **(1)**, 0.2U **(2)**, 0.4U **(3)** and 0.04U **(4)**; **MluI** - 15U **(5)**, 1.5U **(6)**, 0.15U **(7)** and 0.015U **(8)**; **NruI** - 10U **(9)**, 1U **(10)**, 0.1U **(11)** and 0.01U **(12)**; **SacII** - 20U **(13)**, 2U **(14)**, 0.2U **(15)** and 0.02U **(16)**. The partial digests were electrophoresed in a 1% gel at 5V/cm for 14 hours with a linear switching time of 1 to 35 seconds. The limiting mobility **(LM)** of this gel (≈ 520kb) is indicated to the right of the figures. The positions and sizes of coincident restriction sites detected by the 'left' **(L1-L3)** and 'right' **(R1-R3)** pYAC4 vector arms are indicated to the right of the figures (see also **(C)** below). In each autoradiograph the size of the smallest hybridising fragment is also indicated.

(C) The inferred restriction map from the hybridisation experiments. The 'left' and 'right' pYAC4 vector arms are denoted by the arrow and the arrow with the circle, respectively. The restriction sites are depicted above the line (**B** = BssHII; **M** = MluI; **N** = NruI and **S** = SacII). The positions of the coincident sites are depicted below the line. The coincident sites denoted by the asterix mark the position of a central CpG island that is also cut in genomic DNA (see **figure 8.3**).


A
pBR322 LHE
BssHII
MluI
NruI
SacII
1 2 3 4
5 6 7 8
9 10 11 12
13 14 15 16
LM
L3 340Kb
L2 275Kb
L1 205Kb
115Kb


B
pBR322 RHE
BssHII
MluI
NruI
SacII
1 2 3 4
5 6 7 8
9 10 11 12
13 14 15 16
LM
R3 320Kb
R2 260Kb
R1 220Kb
80Kb

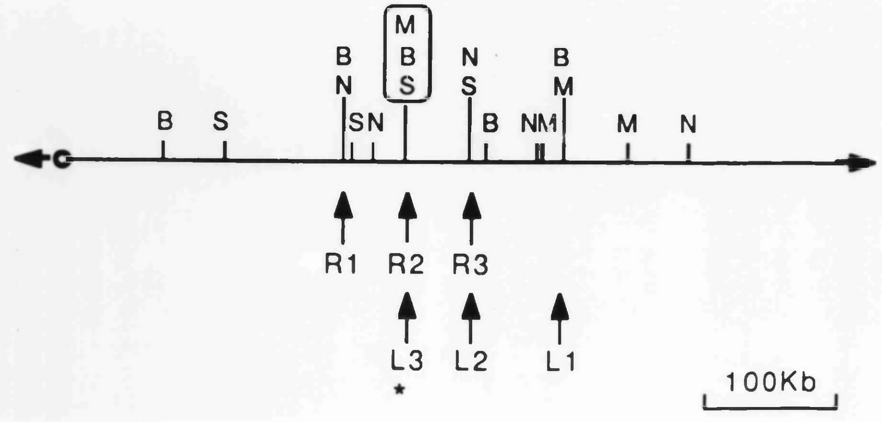
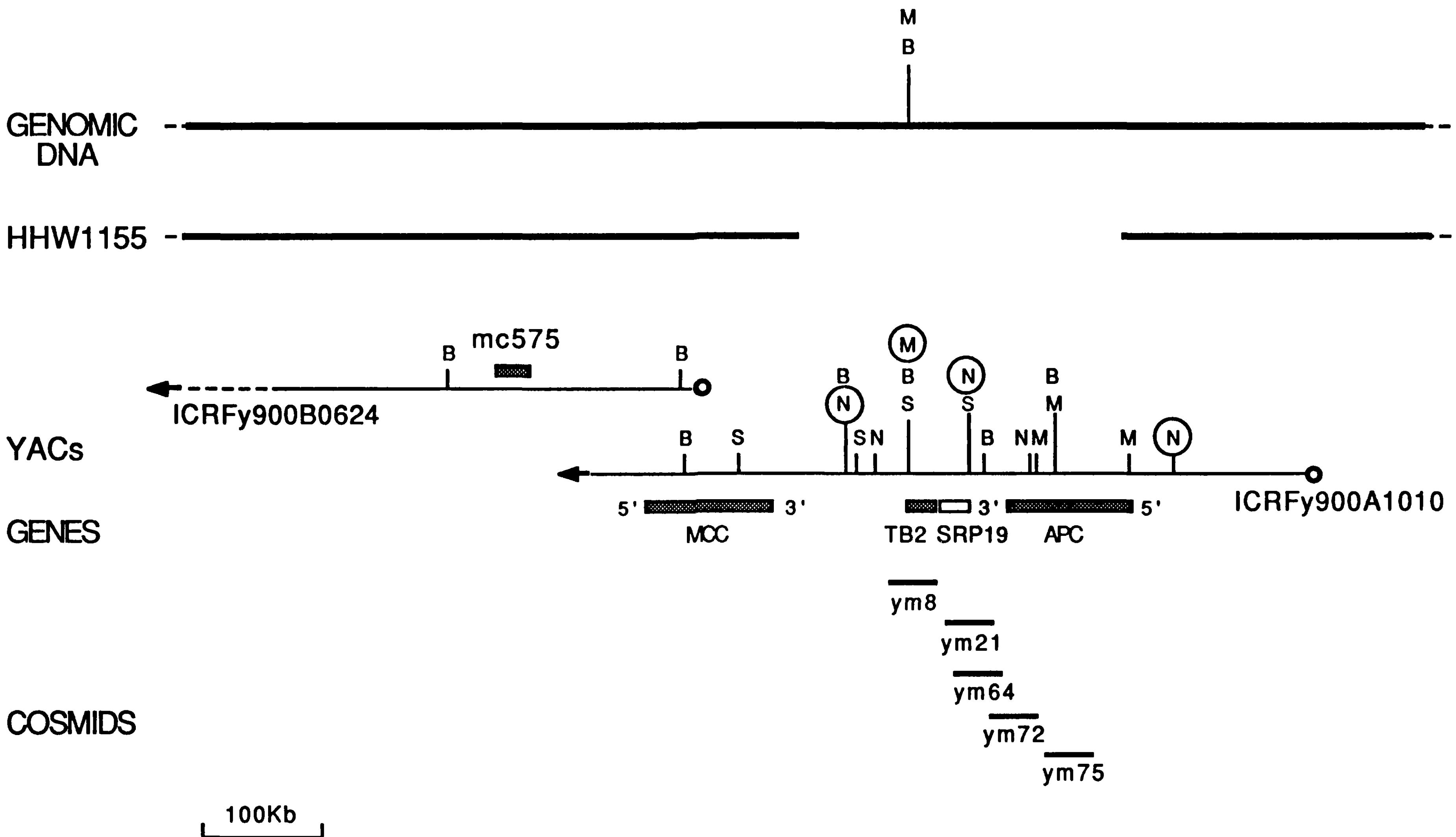

C
M
B
S
B
N
N
S
B
M
B
S
SN
B
NM
M
N
C
R1
R2
R3
L3
L2
L1
*
100Kb

Figure 8.3. A detailed physical map of the *APC* gene-region.

This figure depicts a detailed YAC-based physical map of the *APC* gene-region showing the relative location of YACs and cosmids to the deletion in patient 3214. **Human genomic DNA** (4X) and hybrid **HHW1155** DNA are depicted at the top of the figure by the thick black lines. The break in the **HHW1155** line indicates the most likely position and extent (260kb) of the deletion based on the physical map reported by Joslyn et al., (1991). The YACs are depicted by the thin black lines bounded by the circles ('left' arm) and arrows ('right' arm). Restriction sites are denoted above these lines: **B** = BssHII, **M** = MluI, **N** = NruI, **S** = SacII. The restriction sites bounded by circles are those present in genomic DNA according to Joslyn et al., (1991). The map of YAC **B0624** is only partially illustrated; the hatched line represents continuity in the YAC in the direction of the arrow (total size = 620kb). The position of the ***MCC***, **TB2** and ***APC*** genes are illustrated below YAC **A010**. **SRP19** (which was not localised on this map) is depicted in the most likely position according to Joslyn et al., (1991). The position of microclone insert mc575 is shown on the 200kb BssHII fragment in YAC **B0624**. Cosmids are depicted by the thick black lines (average size = 35kb) and positioned on the map on the basis of ***APC*** and **TB2** cDNA hybridisation analyses. The position of cosmid **ym8** was confirmed by hybridisation of a sub-fragment on the A1010 digests, in addition to the presence of rare-cutting restriction sites in the cosmid DNA.



GENOMIC DNA
HHW1155
YACs
GENES
COSMIDS
M
B
mc575
ICRFy900B0624
S
N
SN
NM
ICRFy900A1010
5'
3'
MCC
TB2
SRP19
APC
ym8
ym21
ym64
ym72
ym75
100Kb

It is clear that the entire *APC* gene is contained within this YAC with approximately 160 kb of flanking DNA at the 5' end, relative to the 'left' pYAC4 vector arm. The positions of the central NruI and MluI sites (at the 5' end of the TB2 gene and at the 3' and 5' ends of the *APC* gene; outlined by circles in the figure) agree well with the sites detected in genomic DNA by Joslyn et al., (1991). In addition, it has been possible to align this YAC-based physical map with the genomic PFGE map constructed at ICRF (Mr. T. Ward). The YACs align at the central coincident BssHII and MluI sites that divide the genomic regions containing the *MCC* and *APC* genes.

8.2.6. Relative transcriptional orientation of the *MCC* and *APC* genes.

The cDNA probes provided for the *MCC* and *APC* genes were essentially divided into 5' and 3' portions and, by hybridisation, it was possible to determine the relative transcriptional orientations of each of these genes on the restriction map (see figure 8.3). It was found that the *MCC* and *APC* genes are transcribed in inverse orientations with respect to each other as described by Joslyn et al., (1991) and that the 3' ends of the genes are about 150kb distant from each other.

8.2.7. TB2 is associated with a CpG island.

The position of TB2 is shown in figure 8.3. (SRP19 was not available at the time of these studies and therefore not placed on the YAC map; the most likely position of this gene is, however, indicated by the unfilled box). According to the physical map constructed by Joslyn et al., (1991), which appears to be in agreement with all aspects of the map described here, the TB2 and SRP19 genes are located entirely within a 60 to 70kb NruI-MluI fragment. These genes are transcribed in opposite orientations relative to each other. The 5' end of TB2 bounds the MluI site of this fragment (Joslyn et al., 1991 and this thesis) and is clearly associated with a CpG island cut in both the cloned YAC and genomic DNAs. It appears that the 5' ends of the *APC* and *MCC* genes are not associated with such islands either in the cloned YAC or genomic DNAs.

8.2.8. Deletion breakpoints of 3214.

At the time of writing, the 3214 deletion breakpoints have not been cloned and the sites have not therefore been accurately localised on the YAC-based

physical map. Thus, the approximate position of the deletion shown in figure 8.3 has been inferred from the genomic map constructed by Joslyn et al., (1991).

8.2.9. The maximal overlap of YACs A1010 and B0624 is 110kb.

The overlap between the A1010 and B0624 YACs (i.e., at the *MCC* gene) was determined by hybridisation of the end-clone 624-575, fragment '2' (see Chapter 7) to the partial and total digest filters of the A1010 YAC. This 'end-probe' recognises the restriction fragment between the first BssHII and SacII restriction sites relative to the right end of A1010 (figure 8.3). The size range of overlap is therefore between 80 and 110kb, which largely agrees with the estimate made during the original PFGE analysis (see Chapter 6).

8.2.10. Cosmids ym72 and ym75 contain portions of the *APC* gene.

In independent hybridisation experiments carried out by Ms. S. Cottrell and Mr. T. Ward, exons from both the *APC* and TB2 genes were found on some of the 6 cosmids that exhibited the aberrant hybridisation patterns in the HHW1155 hybrid. Specifically, exons from the 5' and 3' regions of the *APC* gene (i.e. probes FB9A and FB54D), were found to be present in cosmids ym75 and ym72, respectively. Exons from the TB2 gene were found in cosmid ym8. Cosmids ym21 and ym64, however, (which had been shown to overlap with ym72) did not appear to contain any exons from either of these two genes. On the basis of these hybridisation experiments, the likely positions of these cosmids are indicated on the map in figure 8.3. The positions of cosmids ym8 and ym75 were, in addition, confirmed by the presence in the cosmids of rare-cutting restriction sites localised to the same regions in the YAC and also by hybridisation of cosmid sub-fragments on the A1010 digests.

These results, however, presented some interpretive difficulties in that none of the cosmids should exhibit altered sized restriction fragments as they are predicted to be contained entirely within the 3214 deletion. Possible explanations for these discrepancies are discussed below.

8.3. Conclusions.

The use of YACs B0624 and A1010 for direct screening on the gridded chromosome 5-specific library allowed the rapid detection of cosmids specific

to the region of interest and provided an essentially random sample of about 1Mb of cloned DNA. However, in the initial YAC screens, a large number (about 50%) of cosmids did not map back to the YACs (table 8.1(b)). These included cosmids that had been independently identified by the region of overlap between A1010 and B0624. The exact reason for this is unknown but may have been due to insufficient blocking of low- or moderate-copy number repeat elements possibly present in the YACs. Cosmids containing these repeats would therefore appear as representing sequences specific to the YACs. Further attempts using the A1010 YAC as probe on the library gave rise to a set of 14 cosmids, all of which mapped to the region of interest (that is, by mapping analysis or by cross-hybridisation to *MCC* cDNA probes). It is presumed that more efficient quenching of repeat sequences occurred in this case.

Sufficient numbers of cosmids were identified such that a 260kb sub-microscopic deletion could be identified. The genomic region covering the deleted DNA is entirely contained in A1010. This YAC identified a total of 16 *bona fide* cosmids representing a coverage of about 650kb assuming an average insert size of 40kb for each cosmid. Not all of the cosmid hybridisation patterns were rigorously compared and it is therefore not known what degree of redundancy exists in the set used. However, restriction enzyme digestion analysis on a large number of these cosmids has suggested that most are likely to be independent (Mr. T. Ward, personal communication).

Analysis of cosmid hybridisation patterns on the HHW1155 somatic cell hybrid, in addition to hybridisation of cDNAs representing the *APC* and TB2 genes on the cosmids exhibiting the aberrant hybridisation patterns present some difficulties in fully understanding the nature of the 3214 deletion. It was at first assumed that the group of three overlapping cosmids (ym21, ym64 and ym72) recognised one 'side' of the deletion. However, since cosmid ym72 contains a portion of the 3' end of the *APC* gene, this cosmid (and by extension, the other two cosmids) must be present entirely within the deletion, assuming that this genomic alteration arose by simple excision of DNA.

That ym8 contains a portion of the TB2 gene and also exhibits an altered sized restriction fragment in hybrid HHW1155 is also difficult to explain (in particular because this altered band is different from that of the other three

overlapping cosmids). The breakpoint of the 260kb deletion relative to the 'right' pYAC vector arm of this YAC must extend at least 80-100kb distant to the position of this cosmid since the first exon of *APC* is not deleted (see figure 8.3). Thus, it is unlikely that ym8 reveals the deletion breakpoint. The difficulty, therefore, lies in accounting for the recognition of altered sized restriction fragments in hybrid HHW1155.

One explanation is that the deletion is complex, resulting in the retention of rearranged DNA fragments from this, or, indeed, from another genomic region by an insertion/deletion mechanism. However, it is unlikely that all of the four cosmids would recognise a small segment(s) of foreign DNA or even a small segment(s) of the *APC* gene-region remaining after deletion had occurred. Another, and perhaps more likely explanation, is that these cosmids recognise an homologous locus or loci present on a marker chromosome (other than chromosome 5) segregated in only the HHW1155 hybrid. In this respect, it is interesting that each of the cosmids revealing altered bands are in a genomic location likely to harbour the SRP19 gene. Cross-homology to related SRP genes (on different sized exon-containing restriction fragments for the overlapping and ym8 cosmids, respectively) might reveal such bands upon whole cosmid hybridisation. At the time of writing, these and other possibilities are under investigation. Regardless of the exact reason for this anomaly, two of the cosmids found by this approach recognised exons from the *APC* gene, showing that this strategy was successful.

Extensive physical mapping of A1010 showed that this YAC contains all of the genes recently identified in the vicinity of *MCC* (Joslyn et al., 1991; Kinzler et al., 1991b). Although the map correlates extremely well to the genomic physical maps constructed by Joslyn et al., (1991b) and at ICRF (Mr. T. Ward, unpublished), some discrepancies exist between this map and the YAC-based physical map of Kinzler et al., (1991). This is thought to be due to a mis-alignment of one of their YACs at the 5' end of the *APC* gene. The A1010 YAC isolated in this project provides an ideal tool for studying the *APC* gene (as well as the *MCC* gene) in a biological context by YAC transfer to cell lines derived from colorectal tumours (see Chapter 9 for further discussion).

Chapter 9.

Discussions.

Chapter 9

Discussions.

9.1. Isolation of DNA probes in the *APC* gene-region.

9.1.1. Interspersed repetitive sequence (IRS)-PCR.

Studies presented in Chapter 3 of this thesis demonstrated that the use of just a single *alu*-directed primer, more highly conserved in the human genome (primer IV), could generate highly specific and reproducible banding patterns from a somatic cell hybrid containing one copy of chromosome 5 as the only human constituent (i.e. PN/Ts-1). The 'inter-*alu*' sequences generated from this hybrid cell DNA were shown to be truly human in origin on the basis of mapping studies from cloned material.

The human DNA content of PN/Ts-1 can be estimated to represent some 2-3% of the total hybrid cell DNA content. Since none of the probes tested from the 'inter-*alu*' PCR product library appeared to be hamster in origin, a very high degree of enrichment for human sequences was attained. It is, in addition, likely that most of these sequences are actually derived from chromosome 5 on the basis of cytogenetic characterisation of the PN/Ts-1 hybrid (Varesco et al., 1989). Isolation of chromosome-specific probes from complex sources such as somatic cell hybrids usually involves the construction of highly representative cosmid or bacteriophage libraries from which a proportionally small number of human sequences have to be isolated and processed. The *alu*-PCR method demonstrated that human chromosome-specific probes could be generated after a simple PCR reaction and rapid cloning procedure. Moreover, on the basis of the library clone *in situ* hybridisation (in addition to the chromosomal hybridisation studies), a large proportion of these probes would appear to be useful for mapping studies in containing little or no additional repeat sequences; at least no more than would be expected by cloning similar sized fragments in a plasmid based library from a human source (Ludecke et al., 1989; 1990).

The method developed to efficiently clone the inter-*alu* PCR products has the following advantages: (1) No digestion of the PCR products are required and thus clones with internal restriction sites for the same enzyme will not be

reduced in size and, (2) none of the inserts can be co-ligated together before insertion into the vector since the PCR products are not 5'-phosphorylated.

It was expected that the amplification and cloning procedures would prove useful for the production of libraries from selected or non-selected human chromosome fragments generated by irradiation and fusion (e.g. Goodfellow and Pritchard, 1988; Cox et al., 1990; Benham et al., 1989) or from small numbers of flow sorted chromosomes. In general, these goals have been realised by independent studies (Brookes-Wilson et al., 1990; Cotter et al., 1991). Prior to the development of this technique, the construction of libraries from flow sorted chromosomes had been arduous with respect to the numbers of chromosomes required for a single cloning experiment (at least 5 x 10^5; e.g. see Krimlauf et al., 1982; Nizetic et al., 1991). It is now possible to generate useful libraries from as few as 100-200 flow sorted chromosomes (Cotter et al., 1991).

More recently, the basic *alu*-PCR method has been used to generate specific banding patterns from somatic cell hybrids that retain non-selected segments of human chromosomal DNA thought to overlap at regions of interest. PCR products in common between the chromosomal segments, which are assumed to come from the region of overlap, can be directly isolated and mapped, thus avoiding any cloning procedures (e.g. see Benham and Rowe, 1992)

The observation of highly reproducible and discrete banding patterns demonstrated for the PN/Ts-1 hybrid led to the hypothesis that homologs of a chromosome exhibiting interstitial deletions might show differences in banding patterns as compared to the normal chromosome. Although primer IV did not show such alterations in pattern, amplification using other *alu*-based primers alone or in combination with a primer based on the LIHs repeat family (Ledbetter et al., 1990) led to the identification of a number of fragments; one of which, L5.4, was localised to the minimal region near the *APC* gene.

Although this comparative technique has been used to a greater effect on the X-chromosome at Xq28 (Ledbetter et al., 1990), a major limitation in their studies, and the studies described here, has been the compounding background smears observed in the PCR reactions, particularly when using *alu*-based primers in combination with the LIHs primer. These smears make

identification of potential difference products difficult, especially in the higher molecular weight ranges. Some of the more highly conserved *alu*-based primers (e.g. primer IV and 559), in addition, also suffer from this effect.

The least contaminating smears are seen in the inter-LIHs PCR products and this probably reflects the lower number of these repeat sequences in the genome. Ledbetter et al., (1990) have shown that no additional products were observed in a comparison of hybrids containing Xq28 with hybrids containing the whole X chromosome which suggested a lower template competition effect with this primer. Thus, the pattern generated for chromosome 5 using the LIHs primer shown in this thesis probably reflects the actual number of repeat-pairs that meet all of the criteria of inverse orientation and distance. Indeed, comparison of this pattern to that generated from an independent chromosome-5 only hybrid by Ledbetter et al., (1991), are extremely similar, if not identical.

In contrast, comparison of chromosome X with region Xq28 using *alu*-based primers shows that additional products appear in the latter with an associated reduction in background smear (Ledbetter *et al.*, 1990). This argues that other less favourable products will be generated under reduced competitive IRS-PCR reactions.

Recently, methods have been developed to counter the background effects. In principal the most effective has been the prior digestion of the hybrid DNA samples with frequently cutting restriction enzymes. This pre-digestion strategy restricts the genomic DNA between a number of the potential repeat pairs and therefore leads to greatly reduced banding patterns (Guzzetta et al., 1991). Interestingly, this also results in a considerably reduced background smear with the apparent generation of entirely new, or previously 'hidden', inter-*alu* products. This can be attributed, again, to a reduction of the competitive template effect.

Further adaptations of the basic IRS-PCR method have led to other strategies for the identification of inter-hybrid difference products. For example, chromosome-specific genomic cosmid or bacteriophage libraries can be screened sequentially with the IRS-PCR products of a deleted and then normal chromosome (Bernard et al., 1991) or, chromosomes with a number of different translocations (Cotter et al., 1991). In the latter approach, genomic clones can be identified in the region between two translocation breakpoints

(Cotter et al., 1991). Conversely, non-subtractive enrichment methods such as coincidence cloning, where probes from a region of overlap between human chromosomal fragments retained in somatic cell hybrids, have also proved successful (Aslanadis and deJong, 1991). In sum, these methods illustrate the power of IRS-PCR-based approaches in generating markers for a region of interest, depending on the resources available.

9.1.2. Microdissection and Microcloning.

The use of genomic libraries constructed from DNA derived by the physical dissection of chromosomal sub-region 5q21-q22 proved to be a valuable source of probes mapping near the *APC* gene, as defined by the breakpoints of two, and later three interstitial deletions. However, as discussed in Chapter 4, the use of these microclone inserts was generally difficult because of the small insert sizes, thus requiring the optimisation of the hybridisation procedures.

It is clear also that even with these modifications (see Chapter 4), application of the microclone inserts to other elements in the positional cloning strategy aimed at the *APC* gene were found to be difficult. For example, experiments aimed at the at the construction of genomic physical maps have most recently been carried out entirely with sub-fragments of cosmid inserts isolated with the microclone sequences, rather than with the microclone inserts themselves (Mr. T. Ward, personal communication). In addition, the failure to detect YACs with a significant number of the microclone inserts (Chapter 4) supports this observation (see part II of this discussion).

Nonetheless, 14 microclone inserts were mapped to the minimal region harbouring the *APC* gene (i.e., interval II) which has provided a valuable resource, in conjunction with other markers, for the construction of physical and YAC-based genomic maps of the region. Indeed, two of these microclones have been confirmed to flank the *APC* gene at a maximal distance of between 900kb to 1Mb by a combination of YAC coincidence and genomic physical mapping (see figure A7.1, appendix to Chapter 7). Thus, even in the absence of the cloned *MCC* gene, it may have been the case that a YAC contig seeded with microclone insert mc575 would have progressed to include mc241, and therefore have encompassed the *APC* gene leading to its identification (see below).

One of the key questions that could be addressed was the distribution of the 12 physically localised microclone inserts produced by microdissection and microcloning around the *APC* gene-region. On a gross chromosomal level, extensive deletion mapping studies with a large number of library II inserts suggested that the probes were evenly distributed across the dissected area; in line with published reports on similar libraries (Davis et al., 1990; MacKinnon et al., 1990; Hirst et al., 1991a). At the level of cosmids, YACs and genomic physical maps of large restriction fragments, however, the clones were found to be tightly clustered in a number of cases.

This distribution, which may be due to preferential restriction site clustering (see below), or perhaps due to the selective cloning of DNA least physically damaged by acid depurination during chromosome preparation (Dr. B. Horsthemke, personal communication; see also Brown and Greenfield, 1987), may have precluded the construction of a complete YAC contig across Interval II in the absence of other sub-localised probes. In retrospect, this became irrelevant for cloning the *APC* gene on the basis of mapping microclones that flanked this region, but does warrant caution in assuming a totally random genomic distribution.

Except for the Wilm's tumour region at 11p13 (Davis et al., 1989) and, to some extent, the fragile-X region, few published reports exist concerning the localisation of microclone inserts from these library types at the level of pulsed field gel (PFG) maps. In the case of the Wilm's tumour gene-region, six microclone inserts assisted in the construction of a 7.5Mb genomic PFG map, although some suggestion of clustering is apparent from their report. At the level of PFG and YAC-based maps, Hirst et al., (1991b) have shown the localisation of two microclone inserts flanking the FRAXA mutation at Xq27.3, analogous to the case of the *APC* gene-region described in this thesis. However, apart from an apparently even distribution of microclone inserts across an extensive panel of somatic cell hybrids retaining X chromosome variants (Hirst et al., 1991a), there is no published data to ascertain the actual physical distributions of these Xq-derived microclones.

An interesting feature of these and other studies, has been the observation of cross-species sequence conservation in a significant proportion number of microclone inserts examined. Approximately 25% of the microclones characterised in the large scale mapping analysis described here were found to

be conserved in rodent DNA; some displaying very strong cross-hybridisation signals (e.g. see figure 4.2(b), panel 2).

In two independent studies at least, this conservation has been shown to represent the identification of gene-regions. Sequencing of a conserved microclone; EAN04, derived from microdissection of chromosome region 22q11.2-q13.1, established that it is entirely homologous to the leukemia inhibitory factor gene (LIF) previously mapped to 22q12-q13.1. Interestingly, the RsaI restriction site at the 5' end of the clone is part of a splice acceptor site in the LIF gene (Fiedler et al., 1991). In addition, the microclone - mc1 (D15S62), characterised in microdissection library I (Chapter 4) appears to recognise a splice donor site in an unreported gene mapping to 15q11.2-q12.2 (K. Buiting; B. Horsthemke, personal communication).

From the coincident YAC hybridisation analysis and genomic PFG mapping studies presented in Chapters 6 and the appendix to Chapter 7, it appears that a number of these clones may also be associated with CpG islands cut in genomic and/or cloned DNA. Taken together with the observation of cross-hybridisation in rodent DNA, this suggests that many of the microclones studied may be part of, or close to genes at 5q21-q22. It is unclear whether any of the microclones clustered at the YN5.48 site or clustered distal to this locus recognise other gene-containing regions. If one could generalise on the observation of preferential splice-site recognition (on the basis of the two studies outlined above), then other genes not associated with CpG islands might also be recognised. Perhaps such genes are present in these regions.

The bias in distribution toward genes in the microdissection libraries constructed by this method must be a result of RsaI restriction digestion, since human genomic microclone libraries constructed in other ways do not show such features (Bates et al., 1986; Kaiser et al., 1987; Weber et al., 1990; Kao and Yu, 1991). These libraries may, therefore, provide a distinct advantage in reverse genetic strategies where genes are being sought in particular genomic regions. However, if one were to rely entirely on this strategy, a very large number of clones must be mapped to ensure, at least, sufficient 'gene-coverage'.

9.2. Identification and characterisation of YACs in Interval II.

9.2.1. YAC isolation.

The isolation of YACs by microclone hybridisation was only partially successful. In total, 4 of 10 microclones studied in detail gave rise to positive signals on the high density gridded YAC library filters. Poisson statistics calculated for YAC libraries similar to the one used in these studies predict that some 86.5% of probes should recognise at least one cognate YAC in a 2-hit library[1] (Abidi et al., 1990; Schlessinger et al., 1991). The identification of YACs for 23 of 30 autosomal human probes (76%) in the ICRF two-genome equivalent filters largely agrees with these expectations (Larin et al., 1991). A 'null' class will always be observed representing sequences for which no YAC exists in the library (e.g. 14% in the case of the ICRF library; Larin et al., (1991)). This is not thought to represent 'non-clonability' in the YAC system as 'null' probes have been shown to detect cognate YACs in other independent libraries (Schlessinger et al., 1991). The 'null' class is therefore due to statistical variation. That 60% of the microclones did not detect a cognate YAC suggests a signal-to-noise problem rather than a 'null' class. Moreover, the microclones that did identify positive YACs recognised an average of 2 per probe (see table 5.1, Chapter 5); exactly as would be predicted for a two-genome equivalent library.

These results do support the use of alternative detection methods for small insert probes; particularly by PCR selection (e.g. see Hirst et al., 1991a; Hirst et al., 1991b) where the signal-to-noise ratios can be equalised by changing the probe format to primer-pairs. However, at the time of these studies YAC DNA pools were not available. Thus, a more relevant conclusion here would be that a greater number of microclones could have been investigated by somatic cell hybrid mapping to increase the pool of 'usable' probes.

In retrospect, however, little information would have been gained by the detection of further YACs with the microclones that had been localised to interval II in this study. In total, 8 of the 14 clones in this interval have been localised to one or more YACs, either by primary hybridisation or coincidental mapping. A further 3 have been localised distal to *APC* by genomic physical mapping and only 2 (mc5 and mc452) remain unlocalised at

---

[1] $1/e^n$ is the fraction of YACs expected to be missing from the library where n = the number of genome equivalents; in this case 2 (Abidi et al., 1990).

the time of writing. Probe mc241, which flanks the *APC* gene proximal to mc575 (see above), may have been useful, however, for linking the YAC contigs around the *APC* gene-region and the mc434-p3.1-mc534 probe region. However, identification of YACs for mc241 (in addition to the unlocalised microclones mc5 and mc452) was not possible on the filters used.

9.2.2. YAC overlap detection.

Initial attempts to detect overlap between YACs or sets of YACs isolated by independent probes was based on regional coincident mapping (Abidi et al., 1990). In a study of chromosome Xq24-q28, Abidi et al., (1990) used 70 localised probes in hybridisation assays across a collection of 467 YACs. A reasonable number of the probes detected coincident overlap as expected from essentially random probe distributions and random cloning in YACs. Although this study dealt with only 17 YACs and 17 probes, these YACs were expected to be limited to a distance of some 5Mb (compared with about 50Mb in Xq24-q28). This did not result in the linking of any YACs or YAC contigs but did help to sub-localise 6 (33%) interval II markers (see above).

After coincident mapping, a detailed consideration of all methods available for overlap detection was made before the initiation of a major study. For the reasons outlined in the introduction to Chapter 7, it was decided to isolate the terminal sequences from the YACs of interest; particularly from the termini thought to represent the distant extremes of the YAC contigs.

Several methods for insert termini isolation were assessed. *Alu*-vector PCR gave rise to sequences from only 20% of termini investigated. This is in contrast to the frequencies reported by Nelson et al., (1991) where 60% of termini were isolated in a set of 110 YACs on the X chromosome. However, their experiments were carried out with an *alu*-based primer; 278, which is localised to the most 5' end of the consensus sequence, in addition to 517 and 559. This primer appears to be highly suitable for use on yeast backgrounds and, because of its position on the *alu* element, may allow *alu*-vector amplification from a greater number of *alu* repeats in either orientation. The frequency of termini generated by the 'vectorette' method (Riley et al., 1990) was also found to be too low for rigorous analysis (i.e., 37%). Comparative statistics have not been reported for this system but it has been favoured to library construction or subcloning in YAC-based genome projects

(Schlessinger et al., 1991) and has proved rapid in attempts at genomic walking (Butler et al., 1992).

Analysis of the limitations of both of the methods discussed above led to the formulation of a novel protocol for terminal sequence isolation, termed 'junction-trapping'. The two principal features of this system were (1) the use of Sau3A1 partial yeast digestions to create sufficiently close points in the insert DNA to be within a distance amenable to PCR (in combination with pYAC4-specific primers) and, (2) the manipulation of ligation conditions to favour circularisation such that junction sequences would be preferentially amplified with specific primer pairs. Other selection mechanisms, such as secondary PCR and altered primer ratios were also employed to increase the fidelity of the system. The isolation of both YAC insert termini from both of the YACs present in the B0624 yeast demonstrates the power of this approach.

In general, junction-trapping proved to be useful for the generation of candidate insert termini for all of the YACs investigated. These candidate fragments were shown to be *bona fide* in all but one case by rigorous analysis of the sequence origins. This single case (that is, the right hand end of YAC y12.75) proved difficult to assess because of the small insert size.

Hybridisation of the insert termini on EcoRI digests panels of the interval II YACs did not demonstrate overlap between YACs isolated with independent loci. In retrospect, this is not surprising based on the frequency of YAC chimerism (see below) and the actual physical distribution of the YACs on the genomic physical map. Now that a near complete physical map is available, it is clear that none of the YACs or YAC contigs in interval II are expected to demonstrate overlap.

In conclusion, experience with the *APC* gene-region YACs has suggested that isolation of terminal YAC insert sequences provided a potentially effective method for overlap detection. However, the finding of a small non-contiguous stretch of DNA in the cL5.4.s-1 YAC (see conclusions to Chapter 7) has suggested that repetitive sequence approaches, such as *alu*-PCR fingerprinting or fingerprinting by Southern blotting and repeat sequence hybridisations, may have been useful. Thus, overlap detection should probably be carried out by a combination of both of these approaches. Similar conclusions have been recently reported for the Xq24-q28 YAC-based genome project (Schlessinger et al., 1991; Little et al., 1992).

9.2.3. YAC chimerism.

Crude physical mapping and YAC insert termini isolation demonstrated that about 50% of the YACs investigated were chimeric; that is, containing sequences non-contiguous in the genome. Investigation of YAC contigs around the Duchenne Muscular Dystrophy (DMD) gene region initially suggested that chimerism may be only about 20% in the ICRF library (Monaco et al., 1992). However, in this case, aberrant YACs were only inferred to exist, largely on the basis of DMD cDNA hybridisation and crude SfiI PFG analysis. YAC insert termini were not isolated and their estimate (22%) is likely to be under the real level since small non-contiguous pieces of YAC insert, like that found in the L5.4 YAC here, would not have been detected. Other more recent investigations on YAC contigs around the Ewing's Sarcoma breakpoint region on chromosome 11q and the the HLA class II gene-region constructed with YACs from the ICRF library have, in addition to this thesis, suggested that chimeric YACs should, in fact, be expected about 50% of the time (Drs. K. Patel and J. Ragoussis, personal communications).

The origin of chimeric YACs is still debatable (Abidi et al., 1990; Schlessinger et al., 1991) but may be due to recombination between large DNA fragments at homologous sequences in the same yeast cell *in vivo*, co-cloning of fragmented DNA in the ligation mixes or a combination of both. Evidence for intra-chromosomal recombination is, however, increasing. Green et al., (1991a) have investigated the chimeric junction of a single YAC which contains sequences from chromosomes 7 and 10. The junction appears to occur at an *alu* sequence and recombination would be a likely explanation for this.

Studies of a YAC library constructed from a somatic cell hybrid containing Xq24-q28 as the only human component have also provided some possible insights into chimerism (Abidi et al., 1990; Schlessinger et al., 1991). The frequency of rodent-human chimeras in this library was found to be about 15%; much lower than the frequency of inter-human chimeras found in the St. Louis human YAC library which is about 50% (Green and Olson, 1990b). Schlessinger et al., (1991) have proposed that this might be a direct result of the lower degree of homology between the rodent and human genomes such that recombination would be less frequent in the hybrid YAC library. Moreover, in their studies 3 of 820 characterised YAC clones have been found

with sequences from different locations on human Xq. The frequency of co-cloning Xq sequences should be about 1 in 10^5 YAC clones, since the rodent DNA represents a 300-fold excess over the human DNA content, suggesting, again, a recombination mechanism in the origin of these YAC clones (Schlessinger et al., 1991).

Most recently, Nussbaum and colleagues have shown that the chimeric YAC population in a library constructed from a somatic cell hybrid retaining human Xpter to Xq27.3 represents only 11%, in line with Abidi et al's., study, and further supporting the idea of recombination as the origin of these YACs. However, even in the presence of chimeric YACs, it should still be possible to construct YAC-based physical maps providing that sufficient numbers of YACs can be isolated per probe and that the region of interest is saturated to a sufficient extent with random probes (for example, see Little et al., 1992).

9.3. Identification of genes at 5q21-q22.

9.3.1. The *MCC* gene.

Studies on sporadic colorectal carcinomas around region 5q21-q22, revealed a small region, centred on a random polymorphic probe, L5.71, that showed a dramatically increased frequency of allele-loss (Kinzler et al., 1991a). The finding of a somatic alteration in a single tumour, representing a heterozygous 100kb deletion, led to the identification of a candidate gene, termed *MCC* (for mutated in colorectal carcinoma; Kinzler et al., 1991). This gene was predicted to code for a protein of 829 amino acids with a noted 19 amino acid homology to the G protein-activating region of the rat m3 muscarinic acetylcholine receptor (mAChR). PCR-based 'exon-connection' assays (see Fearon et al., 1990) on rat tissue cDNAs demonstrated that the gene is expressed in most normal organs (including the colon), with the highest expression in the brain.

Comparative rodent-human sequencing identified six exons, and analysis of these exons in tumours identified a further two mutations; both of which were found to be conservative changes in different amino acids. These results did, initially at least, suggest a role for this gene in colorectal tumorigenesis.

9.3.2. YACs for the *MCC* gene.

For the reasons outlined in the introduction to Chapter 5, YACs were selected for the *MCC* gene. Three YACs were identified and, on the basis of standard Southern blotting and PFG analyses, two of these were found to be useful, covering a contiguous stretch of 1.1Mb of DNA centred on the *MCC* gene. The circumstantial evidence presented in the introduction to Chapter 8 prompted an analysis of this contig in an attempt to identify other potential candidates for the *APC* gene.

The strategy chosen, that is, direct cosmid library screening, proved to be a useful approach, and, given sufficient coverage of the YACs, allowed the identification of a 260kb interstitial deletion in patient 3214, distant from the *MCC* gene and overlapping with the A1010 YAC. A number of the cosmids representing sequences from the deleted region, and in particular the ym75 cosmid, were thought to represent strong candidates for containing sequences of the polyposis gene. Similar rearrangements were not detected in either of the somatic cell hybrids segregating each of the homologs of chromosome 5 from patient J.T. The association of mental retardation with polyposis, in this patient at least, may have been coincidental to a mutation in the *APC* gene.

9.3.3. The *APC* gene.

At the same time as these studies, a further three genes were identified in the immediate genomic region near the *MCC* gene (Kinzler et al., 1991b; Joslyn et al., 1991). In exactly analogous approaches to those described in this thesis, both groups used YAC-based approaches to identify these genes, seeding each of the contigs with the *MCC* gene. White and colleagues (Joslyn et al., 1991) identified putative deletions in two FAP patients (including the 3214 deletion) on the basis of variantly sized pulsed field restriction fragments detected by probe L5.71 (see above). Subclones from YACs known to span both of these deletions were used to identify cDNAs representing two genes; DP1 and DP2.5, in addition to one gene, SRP19, previously reported but unlocalised at the time (Joslyn et al., 1991). All of these genes were found to be deleted (except the most 5' end of DP2.5) in patient 3214, but only parts of DP1 and DP2.5 were found to be deleted in patient 3824 (Joslyn et al., 1991).

Vogelstein and colleagues constructed a 1.6Mb YAC contig with 9 overlapping YACs around the *MCC* gene and used each of these to screen, directly, cDNA

libraries from normal colonic mucosa. This also led to identification of the same three genes; TB2 (or DP1); SRP19 and *APC* (or DP2.5) (Kinzler et al., 1991b). These three genes were found to be expressed at reasonable levels in the colon, judged by identification of transcripts on Northern blots (Kinzler et al., 1991b).

On the basis of a gross somatic alteration found in a single tumour using cDNA probes from the *APC* gene (Kinzler et al., 1991b), and the relative positions of the three genes with respect to the constitutional deletions in patients 3214 and 3824 (Joslyn et al., 1991), extensive mutation studies were carried out. Constitutional mutations in FAP patients were only found in the *APC* (or DP2.5) gene, segregating precisely with the polyposis phenotype in each of 10 cases (Nishisho et al., 1991; Groden et al., 1991).

Of these 10 *APC*-specific sequence variations, the majority (80%) were predicted to result in premature termination of the *APC* transcript (Nishisho et al., 1991; Groden et al., 1991). One of these premature termination mutations was found to be a heterozygous 2bp deletion in a patient shown by VNTR analysis to be the true sibling of two aged and unaffected parents. Two of three of this patient's offspring have polyposis concurrent with inheritance of this 2bp deletion. This elegantly demonstrated the transmission of a new mutation at the DP2.5 locus; perhaps the most convincing confirmation of this candidate gene as *APC*.

Interestingly, two of the mutations described by Nishisho et al., (1991) were found to be identical in two patients clinically diagnosed as exhibiting FAP and Gardner's Syndrome (GS), respectively. The GS patient was 46 years of age and had, at the time, no evidence of extra-colonic manifestations which is not an uncommon observation in GS kindreds (see Chapter 1). Although no genetic heterogeneity has been found between FAP and GS patients (Nakamura et al., 1988), it is apparent that the specific *APC* gene mutation is not sufficient to give rise to the extra-colonic manifestations. The exact phenotype is therefore likely to be a result of other genetic or environmental influences.

A similarly high frequency (75%) of potentially inactivating mutations has also been observed in primary colorectal tumours. These mutation-types support the hypothesis of the *APC* gene as coding for a tumour suppressor product, analogous in its formal genetics to Retinoblastoma (Knudson, 1985).

It may be that more conservative changes, such as missense amino acid substitutions commonly found in the p53 gene (reviewed in de Fromentel and Soussi, 1992), are not sufficiently damaging to the *APC* gene product.

Nonetheless, more conservative changes may occur in the *APC* gene, giving rise to milder phenotypes not currently recognised as familial polyposis. In this respect, one large family, called kindred 353, has recently been identified in which individuals have a high risk of colorectal cancer but exhibit relatively few adenomatous polyps. This phenotype, however, clearly segregates with polymorphic alleles flanking the *APC* gene (Leppert et al., 1990). If a threshold effect exists for the *APC* gene-product in colorectal epithelial cells (Bodmer et al., 1987), where fluctuations of the gene product in heterozygotes is sufficient to allow occasional deregulated cellular proliferation, functionally abrogating mutations, such as premature termination, would be more likely to give rise to the greater numbers of polyps observed in polyposis patients. If, however, more 'simple' amino acid substitutions could be considered phenotypically milder in effect (on the basis of the mutation types so far described), then these aberrant gene products may be more tolerated with respect to this threshold effect, giving rise to fewer polyps overall. The hypothesis is, of course, dependent on mutations in the *APC* gene being dominant with respect to the initiation of polyp formation, which is almost certainly the case in polyposis individuals. It will be interesting to see if inheritance of small polyp numbers in families, by virtue of allelic heterogeneity at the *APC* locus, is responsible for some of the risk of colon cancer in the general population (Burt et al., 1985).

The frequencies of somatic or constitutional gross alterations at the *APC* locus appears to be quite low: only 2 of 40 FAP patients have shown alterations at the *APC* locus judged by PFGE analysis (Joslyn et al., 1991). Work in the Cancer Genetics Laboratory, ICRF, extends this observation. Of 50 independent FAP patients analysed in a similar fashion by PFGE with similar probes, no gross alterations have been detected (Ms. S. Cottrell, unpublished). This suggests a frequency in the range of 2%. However, it must be noted that such a frequency may be an underestimate since detection will be highly dependent on the restriction enzymes used and the PFGE conditions under which the restriction fragments are resolved (this is discussed further in Viskochil et al., 1990). Gross somatic alterations have been judged from screening cDNAs from the *APC* gene across tumour samples. Since only one alteration was detected in this way (Kinzler et al., 1991b), which has recently

been shown to be caused by the insertion of a LIHs sequence into the gene (Miki et al., 1992), the frequency may be in the range of 0.5%.

The general frequencies of gross somatic or constitutional alterations observed at the *APC* locus are certainly much lower than the frequencies observed at the Rb locus (about 12-30%; see Goodrich and Lee, 1991 for a review) on which our original expectations were based. They are, however, essentially in line with the frequency of constitutional aberrations reported at the NF-1 locus (about 5%; Viskochil et al., 1990) and somatic rearrangements at the WT-1 locus (about 3%; Gessler et al., 1990).

Gross structural analyses of the predicted *APC* gene product have provided relatively little suggestion of a particular function for the protein (Nishisho et al., 1991; Groden et al., 1991). There are no indications of any signal peptides, transmembrane regions, or nuclear targeting signals which, together with a noted hydrophilicity, has suggested a cytoplasmic localisation for the gene product. Multiple serine phosphorylation motifs, glycosylation motifs and myristoylation sites have been detected but the functional significance of these remains to be established (Groden et al., 1991). To detect similarities with other reported gene motifs, Kinzler et al., (1991b) have carried out similarity searches with short stretches of the predicted *APC* protein. The most suggestive has been a similarity to the *ral2* gene in yeast implicated with regulation of *ras* gene activity, but the matches do not appear to be highly statistically significant.

9.3.4. A possible role for *MCC* in colorectal tumorigenesis ?

The studies of Vogelstein and colleagues on the *MCC* gene raise important questions regarding a possible role for this gene in colorectal tumorigenesis (Kinzler et al., 1991a; Kinzler et al., 1991b). Firstly, detailed allele-loss studies on chromosome 5q have shown that a number of tumours exhibit proximal allele-loss up to, and possibly including the *MCC* locus, with some other tumours, however, exhibiting the reverse pattern (Kinzler et al., 1991b). These allele-loss patterns would suggest, *a priori*, that there are two targets of somatic alteration in colorectal tumours; that is, *MCC* and *APC*.

Secondly, the frequency of non-functional mutations in tumour cells has been calculated at $<5 \times 10^{-6}$ per base pair per cell generation, on the basis of observing two mutations in one allele of the p53 gene (where one of the

mutations is expected to be inconsequential; cited in Kinzler et al., 1991b). This figure apparently agrees with the frequency of loss or gain of restriction sites detected by anonymous probes in colorectal tumours (Baker et al., 1989; Vogelstein et al., 1989). Somatic mutation rates for the *APC* and *MCC* genes in tumour samples have been estimated at 4×10^{-5} and 6×10^{-5}, respectively, correcting for the sensitivity of the ascertainment methods and the numbers of nucleotides sequenced (Kinzler et al., 1991b). Thus, the frequency of mutation at the *MCC* locus is about an order of magnitude above that expected for 'random' or inconsequential mutation. One would expect that such a high mutation rate would be reflective of fixation in the clonally expanding tumour because of a selective growth advantage.

Furthermore, the types of mutations so far observed at this locus are suggestive of functional effects: Although 4 of the mutations represent conservative amino acid substitutions, 2 of the mutations are predicted to affect either a splice donor or splice acceptor site. If mutations in the *MCC* gene were truly inconsequential, one would expect to observe an approximately equivalent number of 'silent changes'. All of those described affect the second base of each of the relevant codons. As Kinzler et al., (1991b), point out, "The demonstration of somatic mutations in such (tumour-specific) genes generally provides, at least initially, the most cogent evidence in favour of their involvement (in tumorigenesis). If a substantial proportion of such mutations turn out to be inconsequential, the search for such genes will be considerably complicated". However, that such mutations are random cannot be ruled out.

Kinzler et al., (1991b) and Groden et al., (1991) have proposed a two-gene model to encompass these observations. Comparison of short stretches of the two predicted gene-products has revealed heptad repeat elements in both which have a strong potential for coiled-coil formation; such structures are known to mediate intra- and inter-protein multimers. If these two gene-products were to interact by the formation of hetero-dimers, then somatic mutations in one or other of these genes might be expected to give rise to similar phenotypic effects.

9.4. Future studies.

The arguments set out above pose a critical question of the role of *MCC*, in addition to *APC*, in colorectal tumorigenesis. These questions may be

answered, in part, by functional analysis of these genes in combination or isolation. To this end, the A1010 YAC which contains, entirely, both of these genes (see Chapters 5 and 8), provides an ideal tool for future studies. Functional assessment of tumour suppressor genes on chromosome 5 has, so far, only been addressed at the level of chromosome transfer. Although microcell mediated transfer of chromosome 5 to colorectal tumour derived cell lines shows suppression of tumorigenicity in athymic nude mice with associated alterations in cellular morphology (Tanaka et al., 1991; Hoshino et al., 1991; Goyette et al., 1992), this and related methods can not differentiate which, or how many genes, are required for the effect.

At the time of writing the A1010 YAC has been 'retrofitted' with a novel vector (termed pRAN4; Markie et al., in preparation) containing genes that confers resistance to drug selection in mammalian cells (my own studies). This modified YAC (A1010-pRAN4) is being transferred to a variety of colorectal cell lines where the mutation in the *APC* gene have been characterised (Dr. D. Markie, Cancer Genetics Laboratory, ICRF, and G. Hampton, in progress). The exact roles of both the *MCC* and *APC* genes in colorectal tumorigenesis should then be amenable to investigation by the establishment of cell lines containing the A1010 YAC (on the assumption that this YAC contains all of the appropriate cellular controlling sequences at the 5' ends of the *MCC* and *APC* genes). Analysis of one or other of the genes in isolation may then be achieved either by gene-specific YAC fragmentation, for example at the TB2 locus, or by functional inactivation of either of the *APC* or *MCC* genes by mutatgenesis using yeast integration vectors containing sequences from the gene of interest.

10.

Bibliography.

10

Bibliography.

Abidi, F. E., Wada, M., Little, R. and Schlessinger, D. (1990) Yeast artificial chromosomes containing human Xq24-Xq28 DNA: Library construction and representation of probe sequences. *Genomics* **7**: 363-376.

Aldred, M. A., Rees, M., Tsioupra, K., Leigh, S. E. A., Neale, K. F. and Delhanty, J. D. A. (1988) Familial polyposis coli. *Lancet* **2**: 565-565.

Alitalo, K., Winqvist, R., Lin, C. C., de la Chapelle, A., Schwab, M. and Bishop, J. M. (1984) Aberrant expression of an amplified c-*myb* oncogene in two cell lines from a colon carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **81**: 4534-4538.

Alm, T. and Licznerski, G. (1973) The intestinal polyposes. *Clin. Gastroenterol.* **2**: 577-602.

Anand, R., Riley, J. H., Butler, R., Smith, J. C. and Markham, A. F. (1990) A 3.5 genome equivalent, multiaccess YAC library: construction, characterisation, screening and storage. *Nucl. Acids Res.* **18**: 1951-1958.

Anand, R., Ogilvie, D. J., Butler, R., Finnear, R. S., Powell, S. J., Smith, J. C. and Markham, A. F. (1991) A yeast artificial chromosome contig encompassing the cystic fibrosis locus. *Genomics* **9**: 124-130.

Armitage, P. and Doll, R. (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**: 1-12.

Armstrong, B. and Doll, R. (1975) Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *Int. J. Cancer* **15**: 617-631.

Ashley, D. J. B. (1969a) The two 'hit' and multiple 'hit' theories of carcinogenesis. *Br. J. Cancer* **23**: 313-328.

Ashley, D. J. B. (1969b) Colonic cancer arising from polyposis coli. *J. Med. Genet.* **6**: 376-378.

Ashton-Rickhardt, P. G., Dunlop, M. G., Nakamura, Y., Morris, R. G., Purdie, C. A., Steel, C. M., Evans, H. J., Bird, C. C. and Wyllie, A. H. (1989) High frequency of *APC* loss in sporadic colorectal carcinoma due to breaks clustered in 5q21-q22. *Oncogene* **4**: 1169-1174.

Ashton-Rickhardt, P. G., Wyllie, A. H., Bird, C. C., Dunlop, M. G., Steel, C. M., Morris, R. G., Piris, J., Romanowski, P., Wood, R., White, R. and Nakamura, Y. (1991) *MCC*, a candidate familial polyposis gene in 5q21, shows frequent allele loss in colorectal and lung cancer. *Oncogene* **6**: 1881-1886.

Aslanidis, C. and deJong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucl. Acids Res.* **18**: 6069-6074.

Aslanidis, C. and deJong, P. J. (1991) Coincidence cloning of *Alu* PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 6765-6769.

Augenlicht, L. H., Wahrman, M. Z., Halsey, H., Anderson, L., Taylor, J. and Lipkin, M. (1987) Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate *in vitro*. *Cancer Res.* **47**: 6017-6021.

Bains, W. (1986) The multiple origins of human *alu* sequences. *J. Mol. Evol.* **23**: 189-199.

Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., vanTuinen, P., Ledbetter, D. H., Barker, D. F., Nakamura, Y., White, R. and Vogelstein, B. (1989) Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**: 217-221.

Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. V. and Vogelstein, B. (1990) Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science* **249**: 912-915.

Balaban, G., Gilbert, F., Nichols, W., Meadows, A. T. and Shields, J. (1982) Abnormalities of chromosome no. 13 in retinoblasts from individuals with constitutional karyotypes. *Cancer Genet. Cytogenet.* **6**: 213-221.

Barlow, D. P. and Lehrach, H. (1987) Genetics by gel electrophoresis: the impact of pulsed field gel electrophoresis on mammalian genetics. *Trends Genet.* **3**: 167-171.

Bates, G. P., Wainwright, B. J., Williamson, R. and Brown, S. D. M. (1986) Microdissection of and microcloning from the short arm of human chromosome 2. *Mol. Cell Biol.* **6**: 3826-3830.

Baxendale, S., Bates, G. P., MacDonald, M. E., Gusella, J. F. and Lehrach, H. (1991) The direct screening of cosmid libraries with YAC clones. *Nuc. Acids Res.* **19**: 6651-6651.

Benedict, W. F., Murphee, A. L., Banerjee, A., Spina, C. A., Sparkes, M. D. and Sparkes, R. S. (1983) Patient with 13 chromosome deletion: evidence that the retinoblastoma gene is a recessive cancer gene. *Science* **219**: 973-975.

Benham, F., Hart, K., Crolla, J., Bobrow, M., Francavilla, M. and Goodfellow, P. N. (1989) A method for generating hybrids containing non-selected fragments of chromosomes. *Genomics* **4**: 509-517.

Benham, F. and Rowe, P. (1992) Use of *Alu*-PCR to characterise hybrids containing multiple fragments and to generate new Xp21.3-p22.2 markers. *Genomics* **12**: 368-376.

Berg, P., Baltimore, D., Boyer, H. W., Cohen, S. N., Davis, R. W. and Hogness, D. S. (1974) Letter: Potential biohazards of recombinant DNA molecules. *Science* **185**: 303.

Bernard, L. E., Brooks-Wilson, A. and Wood, S. (1991) Isolation of DNA fragments from a human chromosomal sub-region by *Alu*-PCR differential hybridisation. *Genomics* **9**: 241-246.

Bicknell, D. C., Markie, D., Spurr, N. K. and Bodmer, W. F. (1991) The human chromosome content in human x rodent somatic cell hybrids analysed by a screening technique using *Alu*-PCR. *Genomics* **10**: 186-192.

Bird, A. P. (1987) CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**: 342-347.

Bishop, D. T. and Thomas, H. J. W. (1991) The genetics of colorectal cancer. *Cancer Surveys* **9**: 585-604.

Bishop, J. M. (1987) The molecular genetics of cancer. *Science* **235**: 305-311.

Blair, N. P. and Trempe, C. L. (1980) Hypertrophy of the retinal pigment epithelium associated with Gardner's syndrome. *Amer. J. Opthalmol.* **90**: 661-667.

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

Bodmer, W. F. Hereditary Colorectal Cancer: A commentary. pp. 37-42 in *Hereditary colorectal cancer*. Edited by Utsunomiya, J. and Lynch, H. T., Springer-Verlag, Tokyo, 1989.

Borrow, J., Goddard, A. D., Sheer, D. and Solomon, E. (1990) Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. *Science* **249**: 1577-1580.

Bos, H. L., Fearon, E. R., Hamilton, S. R., Verlaan-de Vries, M., van Boom, J. H., van der Eb, A. J. and Vogelstein, B. (1987) Prevalence of *ras* gene mutations in human colorectal cancers. *Nature* **327**: 293-297.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980) Construction of a linkage map using restriction fragment length polymorhisms. *Am. J. Hum. Genet.* **32**: 314-331.

Boyle, P., Zairde, D. G. and Smans, M. (1985) Descriptive epidemiology of colorectal cancer. *Int. J. Cancer* **36**: 9-18.

Bratthauer, G. L. and Fanning, T. G. (1992) Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* **7**: 507-510.

Breukel, C., Wijnen, J., Tops, C., v.d.-Klift, H., Dauwerse, H. and Khan, P. M. (1990) Vector-*alu* PCR: a rapid step in mapping cosmids and YACs. *Nucl. Acids. Res.* **18**: 3097-3097.

Breukel, C., Tops, C., Ras, E. and Kahn, P. M. (1991) MspI RFLP at the D5S122 locus tightly linked to *APC*. *Nucl. Acids Res.* **19**: 685.

Britten, R. J., Baron, W. F., Stout, D. B. and Davidson, E. H. (1988) Sources and evolution of human *alu* repeated sequences. *Proc. Natl. Acad. Sci. U.S.A.* **85**: 4770-4774.

Brookes-Wilson, A., Goodfellow, P. N., Povey, S., Nevanlinna, H. A., deJong, P. and Goodfellow, P. J. (1990) Rapid cloning and characterisation of new chromosome 10 DNA markers by *alu* element-mediated PCR. *Genomics* **7**: 614-620.

Brown, S. D. M. and Greenfield, A. J. (1987) A model to describe the size distribution of mammalian genomic fragments recovered by microcloning. *Gene* **55**: 327-332.

Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessinger, D. and Olson, M. V. (1989) Isolation of single-copy genes from a library of yeast artificial chromosome clones. *Science* **244**: 1348-1351.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, D., Haber, D. A., Sharp, P. and Housman, D. E. (1991) Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 4005-4009.

Buiting, K., Neumann, M., Ludecke, H.-J., Senger, G., Claussen, U., Antich, J., Passarge, E. and Horsthemke, B. (1990) Microdissection of the Prader-Willi syndrome chromosome region and identification of potential gene sequences. *Genomics* **6**: 521-527.

Bulow, S., Sondergaard, J. O., Witt, I. N., Larsen, E. and Tetens, G. (1984) Mandibular osteomas in familial polyposis coli. *Dis. Colon Rectum* **27**: 105-108.

Bulow, S., Lauritsen, K. B., Johansen, A. A., Svedsen, L. B. and Sondergaard, J. O. (1985) Gastroduodenal polyps in familial polyposis coli. *Dis Colon Rectum* **28**: 90-93.

Bulow, S. (1986) Results of the Danish polyposis register. *Dis. Colon Rectum* **29**: 102-107.

Bulow, S., Holm, N. V. and Hauge, M. (1986a) The incidence and prevalence of polyposis coli in Denmark. *Scand. J. Soc. Med.* **14**: 67-74.

Bulow, S., Holm, N. V., Sondergaard, J. O., Witt, I. N. and Tetens, G. (1986b) Mandibular osteomas in unaffected sibs and children of patients with familial polyposis coli. *Scand. J. Gastroenterol.* **21**: 744-748.

Bulow, S. (1987) Familial polyposis coli. *Dan. Med. Bulletin* **34**: 1-14.

Buluwela, L., Forster, A., Boehm, T. and Rabbitts, T. H. (1989) A rapid procedure for colony screening using nylon filters. *Nucl. Acids Res.* **17**: 452-452.

Burgers, P. M. J. and Percival, K. J. (1987) Transformation of yeast spheroplasts without cell fusion. *Analyt. Biochem.* **163**: 391-397.

Burke, D. T., Carle, G. F. and Olson, M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**: 806-812.

Burke, D. T. (1990) YAC cloning: Options and problems. *Gen. Anal. Tech. Appl.* **7**: 94-99.

Burt, R. W., Bishop, D. T., Cannon, L. S., Dowdell, M. A., Lee, R. G. and Skolnick, M. H. (1985) Dominant inheritance of adenomatous colonic polyps and colorectal cancer. *New Eng. J. Med.* **312**: 1540-1544.

Burt, R. and Samowitz, W. (1989) The adenomatous polyp and the hereditary polyposis syndromes. *Gastroenterology Clinics North Amer.* **17**: 657-678.

Bussey, H. J. R. Familial polyposis coli. Family studies, histopathology, differential diagnosis and results of treatment. Johns Hopkins University Press, Baltimore, 1975.

Bussey, H. J. R., Veale, A. M. O. and Morson, B. C. (1978) Genetics of gastrointestinal polyposis. *Gastroenterology* **74**: 1325-1330.

Bussey, H. J. R. Polyposis coli: Genetic factors. pp. 45-58 in *Recent results in cancer research*. Springer-Verlag, Berlin, 1982.

Butler, R., Ogilvie, D. J., Elvin, P., Riley, J. H., Finnear, R. S., Slynn, G., Morten, J. E. N., Markham, A. and Anand, R. (1992) Walking, cloning, and mapping with yeast artificial chromosomes: A contig encompassing D21S13 and D21S16. *Genomics* **12**: 42-51.

Cannon-Albright, L. A., Skolnick, M. H., Bishop, D. T., Lee, R. G. and Burt, R. W. (1988) Common inheritance of susceptibility to colonic adenomatous polyps and associated colorectal cancers. *New Eng. J. Med.* **9**: 533-537.

Cavenee, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphee, A. L., Strong, L. C. and White, R. (1983) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**: 779-784.

Cavenee, W. K. (1991) Recessive mutations in the causation of human cancer. *Cancer* **67**: 2431-2435.

Chapman, P. D., Church, W., Burn, J. and Gunn, A. (1989) Congenital hypertrophy of retinal pigment epithelium: a sign of familial adenomatous polyposis. *British Med. J.* **298**: 353-354.

Chiba, M., Ito, R., Masamune, O., Waga, T. and Sato, S. Two cases of familial adenomatous polyposis associated with mental retardation. pp. 417-422 in *Hereditary colorectal cancer*. Edited by Utsunomiya, J. and Lynch, H. T., Springer-Verlag, Tokyo, 1989.

Church, G. and Gilbert, W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **81**: 1991-1995.

Claussen, U., Klein, R. and Schmidt, M. (1986) A pipette method for rapid karyotyping in prenatal diagnosis. *Prenatal Diagnosis* **6**: 401-408.

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

Coffey, A. J., Roberts, R. G., Green, E. D., Cole, C. G., Butler, R., Anand, R., Giannelli, F. and Bently, D. R. (1992) Construction of a 2.6-Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS-based approach. *Genomics* **12**: 474-484.

Cohen, S. B. (1982) Familial polyposis coli and its extracolonic manifestations. *J. Med. Genet.* **19**: 193-203.

Collins, F. S. (1992) Positional cloning: Let's not call it reverse anymore. *Nature Genet.* **1**: 3-6.

Comings, D. E. (1973) A general theory of carcinogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **70**: 3324-3328.

Cotter, F. E., Hampton, G. M., Nasipuri, S., Bodmer, W. F. and Young, B. D. (1990) Rapid isolation of human chromosome-specific DNA probes from a somatic cell hybrid. *Genomics* **7**: 257-263.

Cotter, F. E., Das, S., Douek, E., Carter, N. P. and Young, B. D. (1991) The generation of DNA probes to chromosome 11q23 by *Alu* PCR on small numbers of flow sorted 22q- derivative chromosomes. *Genomics* **9**: 473-480.

Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986) Towards a physical map of the nematode *Ceanorhabditis elegans*. *Proc. Natl. Acad Sci. U.S.A.* **83**: 7821-7825.

Coulson, A., Waterson, R., Kiff, J., Sulston, J. and Kohora, Y. (1988) Genome linking with yeast artificial chromosomes. *Nature* **335**: 184-186.

Cox, D. R., Burmeister, M., Price, E. R., Kim, S. and Myers, R. M. (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**: 245-250.

Cox, D. R. and Shimizu, N. (1992) Report of the committee on the genetic constitution of chromosome 21. Human Gene Mapping 11. *Cytogenet. Cell Genet.* **58**: 800-826.

Cross, I., Delhanty, J., Chapman, P., Bowles, L. V., Griffin, D., Wolstenholme, J., Bradburn, M., Brown, J., Wood, C., Gunn, A. and Burn, J. (1992) An

intrachromosomal insertion causing 5q22 deletion and familial adenomatous polyposis coli in two generations. *J. Med. Genet.* in press.

D'Emilia, J., Bulovas, K., D'Erole, K., Wolf, B., Steele, G. and Sumerhayes, I. C. (1989) Expression of the c-*erb-2* gene product (P185) at different stages of neoplastic progression in the colon. *Oncogene* **4**: 1233-1239.

Davis, L. M., Senger, G., Ludecke, H.-J., Claussen, U., Horsthemke, B., Zhang, S. S., Metzroth, B., Hohenfellner, K., Zabel, B. and Shows, T. B. (1990) Somatic cell hybrid and long range physical mapping of 11p13 microdissected genomic clones. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 7005-7009.

de Fromentel, C. C. and Soussi, T. (1992) TP53 tumour suppressor gene: A model for investigation human mutagenesis. *Genes, Chromosomes Cancer* **4**: 1-15.

Deininger, P. L., Jolly, D. J., Rubin, C. M., Freidmann, T. and Schmid, C. W. (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* **151**: 17-33.

DeLellis, R. A., Dayal, Y., Tischler, A. S., Lee, A. K. and Wolfe, J. (1986) Multiple endocrine neoplasia type-2. *Intl. Rev. Expt. Path.* **28**: 163-215.

Domizio, P., Talbot, I. C., Spiegelman, A. D., Williams, C. B. and Phillips, R. K. S. (1990) Upper gastrointestinal pathology in familial adenomatous polyposis: results from a prospective study of 102 patients. *J. Clin. Pathol.* **43**: 738-743.

Drmanac, R., Petrovic, N., Gilsin, V. and Crkvenjakov, R. (1986) A calculation of fragment lengths obtainable from human DNA with 78 restriction enzymes: An aid for cloning and mapping. *Nucl. Acids Res.* **14**: 4691-4699.

Dukes, C. E. (1952) Familial intestinal polyposis. *Ann. Roy. Coll. Surg.* **10**: 293-304.

Dunlop, M. G., Steel, C. M., Wyllie, A. H., Bird, C. C. and Evans, H. J. (1989) Linkage analysis in familial adenomatous polyposis: order of C11P11 (D5S71) and pi-227 (D5S37) loci at the *APC* gene. *Genomics* **5**: 350-353.

Dunlop, M. G., Wyllie, A. H., Nakamura, Y., Steel, C. M., Evans, H. J. and White, R. (1990) Genetic linkage map of six polymorphic DNA markers around the gene for familial adenomatous polyposis on chromosome 5. *Amer. J. Hum. Genet.* **47**: 982-987.

Duyk, G. M., Kim, S., Myers, R. M. and Cox, D. R. (1990) Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 8995-8999.

Eliceiri, B., Labella, T., Hagino, Y., Srivastava, A., Schlessinger, D., Pilia, G., Palmieri, G. and D'Urso, M. (1991) Stable integration and expression in mouse cells of yeast artificial chromosomes harbouring human genes. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 2179-2183.

Elvin, P., Kerr, I. B., McArdle, C. S. and Birnie, G. D. (1988) Isolation and preliminary characterisation of cDNA clones representing mRNAs associated with tumour progression and metastasis in colorectal cancer. *Br. J. Cancer* **57**: 36-42.

Elvin, P., Kerr, I. B., McArdle, C. S. and Birnie, G. D. (1988) Isolation and preliminary characterisation of cDNA clones representing mRNAs associated with tumour progression and metastasis in colorectal cancer. *Br. J. Cancer* **57**: 36-42.

Elvin, P., Slynn, G., Black, D., Graham, A., Butler, R., Riley, J., Anand, R. and Markham, A. F. (1990) Isolation of cDNA clones using yeast artificial chromosome probes. *Nuc. Acids Res.* **18**: 3913-3917.

Evans, E. P., Burtenshaw, M. D., Brown, B. B., Hennion, R. and Harris, H. (1982) The analysis of malignancy by cell fusion: IX. Re-examination and clarification of the cytogenetic problem. *J. Cell Sci.* **56**: 113-130.

Evans, G. A., Lewis, K. and Rothenberg, B. E. (1989) High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* **79**: 9-20.

Evans, G. A. and Wahl, G. M. Cosmid vectors for genomic walking and rapid restriction mapping. pp. 604-611 in *Methods in Enzymology*. Edited by Berger, S. L. and Kimmel, A. R., Academic Press, London, 1987.

Farr, C. J., Marshall, C. J., Easty, D. J., Wright, N. A., Powell, S. C. and Paraskeva, C. (1988) A study of *ras* gene mutations in colonic adenomas from familial polyposis coli patients. *Oncogene* **3**: 673-678.

Fearon, E. R., Hamilton, S. R. and Vogelstein, B. (1987) Clonal analysis of human colorectal tumours. *Science* **238**: 193-197.

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

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

Feinberg, A. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction fragments to a high specific activity. *Anal. Biochem* **132**: 6-13.

Feinberg, A. P., Gehrke, C. W., Kuo, K. C. and Ehrlich, M. (1988) Reduced genomic 5-methylcytosine content in human colonic neoplasia. *Cancer Res.* **48**: 1159-1161.

Fenoglio, C. M. and Pascal, R. R. (1982) Colorectal adenomas and cancer. Pathologic Relationships. *Cancer* **50**: 2601-2608.

Fiedler, W., Claussen, U., Ludecke, H.-J., Senger, G., Horsthemke, B., van Kessel, G., Goertzen, W. and Fahsold, R. (1991) New markers for the neurofibromatosis-2 region generated by microdissection of chromosome 22. *Genomics* **10**: 786-791.

Fisher, E. M. C., Cavanna, J. S. and Brown, S. D. M. (1985) Microdissection and microcloning of the mouse X-chromosome. *Proc. Natl. Acad. Sci. U.S.A.* **82**: 5846-5849.

Fisher, J. C. (1958) Multiple-mutation theory of carcinogenesis. *Nature* : 651-652.

Forrester, K., Almoguera, C., Han, K., Grizzle, W. E. and Perucho, M. (1987) Detection of high incidence of K-*ras* mutations during human colon tumorigenesis. *Nature* **327**: 298-303.

Francke, U. and Kung, F. (1976) Sporadic bilateral retinoblastoma and 13q-chromosomal deletion. *Med. Pediatr. Oncol.* **2**: 379-385.

Friedberg, E. DNA repair. Freeman, San Francisco, 1985.

Friend, S. H., Bernards, R., Rojeli, S., Weinberg, R. A., Rapaport, J. M., Albert, D. M. and Dryja, T. P. (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**: 643-646.

Fung, Y. K. T., Murphee, A. L., T'Ang, A., Qian, J., Hinrichs, S. H. and Benedict, W. F. (1987) Structural evidence for the authenticity of the human retinoblastoma gene. *Science* **236**: 1657-1661.

Garber, J. E., Li, F. P., Kingston, J. E., Krush, A. J., Strong, L. C., Finegold, M. J., Bertario, L., Bulow, S., Filippone, A., Gedde-Dahl, T. and Jarvinen, H. J. (1988) Hepatoblastoma and familial adenomatous polyposis. *J. Natl. Cancer Inst.* **80**: 1626-1628.

Gardner, E. J. (1951) A genetic and clinical study of intestinal polyposis, a predisposing factor for carcinoma of the colon and rectum. *Am. J. Hum. Genet.* **3**: 167-176.

Gardner, E. J. and Plenk, H. P. (1952) Hereditary pattern for multiple osteomas in a family group. *Amer. J. Hum. Genet.* **5**: 31-36.

Gardner, E. J. and Richards, R. C. (1953) Multiple cutaneous and subcutaneous lesions occurring simultaneously with hereditary polyposis and osteomatosis. *Amer. J. Hum. Genet.* **5**: 139-147.

Gardner, E. J. (1962) Follow-up study of a family group exhibiting dominant inheritance for a syndrome including intestinal polyps, osteomas, fibromas and epidermal cysts. *Amer. J. Hum. Genet* **14**: 376-390.

Gardner, E. J., Rogers, S. W. and Woodward, S. (1982) Numerical and structural chromosome aberrations in cultured lymphocytes and cutaneous fibroblasts of patients with multiple adenomas of the colorectum. *Cancer* **49**: 1413-1419.

Gessler, M., Poustka, A.-M., Cavenee, W. K., Neve, R. L., Orkin, S. H. and Burns, G. A. P. (1990) Homozygous deletion in Wilm's tumour of a zinc-finger gene identified by chromosome jumping. *Nature* **343**(774-778):

Gnirke, A., Barnes, T. S., Patterson, D., Schild, D., Featherstone, T. and Olson, M. V. (1991) Cloning and *in vivo* expression of the human GART gene using yeast artificial chromosomes. *EMBO J.* **10**: 1643-1634.

Godbout, R., Dryja, T. P., Squire, J., Gallie, B. L. and Phillips, R. A. (1983) Somatic inactivation of genes on chromosome 13 is a common event in retinoblastoma. *Nature* **304**: 451-453.

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

Goodfellow, P. N. and Pritchard, C. A. (1988) Chromosome fragmentation by chromosome mediated gene transfer. *Cancer Surveys* **7**: 251-266.

Goodrich, D. W. and Lee, W.-H. (1990) The molecular genetics of retinoblastoma. *Cancer Surveys* **9**: 529-524.

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

Green, E. D. and Olson, M. V. (1990a) Systematic screening of yeast artificial-chromosome libraries by the use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 1213-1217.

Green, E. D. and Olson, M. V. (1990b) Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. *Science* **250**: 94-98.

Green, E. D., Riethman, H. C., Dutchik, J. E. and Olson, M. V. (1991a) Detection and characterisation of chimeric yeast artificial-chromosome clones. *Genomics* **11**: 658-669.

Green, E. D., Mohr, R. M., Idol, J. R., Jones, M., Buckingham, J. M., Deaven, L. L., Moyzis, R. K. and Olson, M. V. (1991b) Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: Application to the mapping of human chromosome 7 using yeast artificial chromosomes. *Genomics* **11**: 548-564.

Griffiths, D. F. R., Davies, S. J., Williams, D., Williams, G. T. and Williams, E. D. (1988) Demonstration of somatic mutation and colonic crypt clonality by X-linked enzyme histochemistry. *Nature* **333**: 461-463.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Steven, J., Spirio, L., Robertson, M., Sargent, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J. P., Warrington, J., McPherson, J., Wasmuth, J., Le Paslier, D., Abderrahim, H., Cohen, D., Leppert, M. and White, R. (1991) Identification and characterisation of the familial adenomatous polyposis coli gene. *Cell* **66**: 589-600.

Grundy, P., Koufos, A., Morgan, K., Li, F. P., Meadows, A. T. and Cavenee, W. K. (1988) Familial predisposition to Wilm's tumour does not map to the short arm of chromosome 11. *Nature* **336**: 374-376.

Guzman, P. and Ecker, J. (1988) Development of large DNA methods for plants: molecular cloning of large segments of Arabadopsis and carrot DNA into yeast. *Nucl. Acids Res.* **16**: 11091-11105.

Guzzetta, V., Montes de Oca-Luna, R., Lupski, J. R. and Patel, P. I. (1991) Isolation of region-specific and polymorphic markers from chromosome 17 by restricted *alu* polymerase chain reaction. *Genomics* **9**: 31-36.

Hampton, G., Leuteritz, G., Ludecke, H.-J., Senger, G., Trautmann, U., Thomas, H., Solomon, E., Bodmer, W. F., Horsthemke, B., Claussen, U. and

Ballhausen, W. G. (1991a) Characterisation and mapping of microdissected genomic clones from the adenomatous polyposis coli (*APC*) region. *Genomics* **11**: 247-251.

Hampton, G., Howe, K., Leuteritz, G., Thomas, H., Bodmer, W. F., Solomon, E. and Ballhausen, W. G. (1991b) Regional mapping of 22 microclones to the adenomatous polyposis coli (*APC*) locus on chromosome 5q. *Hum. Genet.* **88**: 112-114.

Hampton, G. M., Ward, J. T. R., Cottrell, S., Howe, K., Thomas, H. J. W., Jones, T., Sheer, D., Ballhausen, W. G., Frischauf, A.-M., Solomon, E. and Bodmer, W. F. (1992) Yeast artificial chromosomes (YACs) for the molecular analysis of the familial polyposis *APC* gene region. *Proc. Natl. Acad. Sci.*, submitted.

Hanahan, D. Techniques for transformation of *E. coli*. pp. 109-135 in *DNA cloning. A practical approach*. Edited by Glover, D. M., IRL Press, Oxford, 1985.

Harris, H. and Klein, G. (1969) Malignancy of somatic cell hybrids. *Nature* **224**: 353-357.

Harris, H., Miller, O. J., Klein, G., Worst, P. and Tachibana, T. (1969) Suppression of malignancy by cell fusion. *Nature* **223**: 363-368.

Harris, H. and Watkins, J. F. (1965) Hybrid cells derived from mouse and man: artificial heterokaryons of mammalian cells from different species. *Nature* **205**: 640-646.

Hayashi, K., Nakazawa, M., Ishizaki, Y., Hiraoka, N. and Obayashi, A. (1986) Regulation of inter- and intra-molecular ligation with T4 DNA ligase in the presence of polyethylene glycol. *Nucl. Acids Res.* **14**: 7617-7630.

Hayashi, K. (1991) PCR-SSCP: A simple method for detection of mutations in the genomic DNA. *PCR Meths. Appls.* **1**: 34-38.

Hermann, B., Bucan, M., Bains, P. E., Frischauf, A.-M., Silver, L. M. and Lehrach, H. (1986) Genetic analysis of the proximal region of the mouse t-complex: evidence for a second inversion within t-haplotypes. *Cell* **44**: 469-476.

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

Hirst, M. C., Roche, A., Flint, T. J., MacKinnon, R. N., Bassett, J. H. D., Nakahori, Y., Watson, J. E. V., Bell, M. V., Patterson, M. N., Boyd, Y., Thomas, N. S. T., Knight, S. J. L., Warren, S. T., Hors-Cayla, M., Schmidt, M., Sutherland, G. R. and Davies, K. E. (1991a) Linear order of new and established DNA markers around the fragile site Xq27.3. *Genomics* **10**: 243-249.

Hirst, M. C., Rack, K., Nakahori, Y., Roche, A., Bell, M. V., Flynn, G., Christadoulou, Z., MacKinnon, R. N., Francis, M., Littler, A. J., Poustka, A.-M., Lehrach, H., Schlessinger, D., D'Urso, M., Buckle, V. J. and Davies, K. E. (1991b) A YAC contig across the fragile X site defines the region of fragility. *Nucl. Acids Res.* **19**: 3283-3298.

Hockey, K. A., Mulcahy, M. T., Montgomery, P. and Levitt, S. (1989) Chromosome deletion of 5q and familial adenomatous polyposis. *J. Med. Genet.* **26**: 61-62.

Holm, N. V., Hauge, M. and Jensen, O. M. (1982) Studies of cancer aetiology in a complete twin population: Breast cancer, colorectal cancer and leukemia. *Cancer Surveys* **1**: 17-32.

Hoshino, Y., Horikawa, I., Oshimura, M. and Yuasa, Y. (1991) Normal human chromosome 5, on which a familial adenomatous polyposis gene is located, has tumour suppressive activity. *Biochem. Biophys. Res. Comm.* **174**: 298-304.

Huff, V., Compton, D. A., Chao, L.-Y., Strong, L. C., Geiser, C. F. and Saunders, G. F. (1988) Lack of linkage of familial Wilm's tumour to chromosomal band 11p13. *Nature* **336**: 377-378.

Huw, H. R., Roberts, J. W., Davidson, E. H. and Britten, R. J. (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. *Proc. Natl. Acad. Sci. U.S.A.* **83**: 3875-3879.

Huxley, C. and Gnirke, A. (1991) Transfer of yeast artificial chromosomes from yeast to mammalian cells. *Bioessays* **13**: 545-550.

Huxley, C., Hagino, Y., Schlessinger, D. and Olson, M. V. (1991) The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion. *Genomics* **9**: 742-750.

Itoh, H., Ohsato, K., Yao, T., Iida, M. and Watanabe, H. (1979) Turcot's syndrome and its mode of inheritance. *Gut* **20**: 414-419.

Jelinek, W. R. and Schmid, C. W. (1982) Repetitive sequences in eukaryotic DNA and their expression. *Ann. Rev. Biochem* **51**: 813-844.

Jonasson, J., Povey, S. and Harris, H. (1977) The analysis of malignancy by cell fusion. *J. Cell Sci.* **24**: 217-254.

Joslyn, G., Carlson, M., Thliveris, A., Albertsen, H., Gelbert, L., Samowitz, W., Groden, J., Stevens, J., Spirio, L., Robertson, M., Sargent, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J. P., Warrington, J., McPherson, J., Wasmuth, J., Le Paslier, D., Abderrahim, H., Cohen, D., Leppert, M. and White, R. (1991) Identification of deletion mutations and three new genes at the familial polyposis locus. *Cell* **66**: 601-613.

Kaiser, R., Weber, J., Grzeschik, K.-H., Edstrom, J. E., Driesel, A., Zengerling, S., Buchwald, M., Tsui, L.-C. and Olek, K. (1987) Microdissection and microcloning of the long arm of human chromosome 7. *Mol. Biol. Rep.* **12**: 3-6.

Kan, Y. W. and Dozy, A. M. (1978) Polymorphism of DNA sequence adjacent to the beta-globin gene region: relationship to sickle mutation. *Proc. Natl. Acad. Sci. U.S.A.* **75**: 5631-5635.

Kao, F.-T. and Yu, J.-W. (1991) Chromosome microdissection and cloning in the human genome and genetic disease analysis. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 1844-1848.

Kariya, K., Kato, K., Hayashizaki, Y., Himeno, S., Tarui, S. and Matsubara, K. (1987) Revision of consensus sequence of human *alu* repeats. *Gene* **53**: 1-10.

Kelly, T. J. and Smith, H.O. (1970) A restriction enzyme from Haemophilus influenzae. 11-base sequence and recognition site. *J. Mol. Biol.* **51**: 393-409.

Kern, S. E., Fearon, E. R., Tersmette, K. W. F., Enterline, J. P., Leppert, M., Nakamura, Y., White, R., Vogelstein, B. and Hamilton, S. R. (1989) Clinical and pathalogic associations with allelic loss in colorectal carcinoma. *JAMA* **261**: 3099-3103.

King, M.-C. (1990) Genetic analysis of cancer in families. *Cancer Surveys* **9**: 417-435.

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

Kinzler, K. W., Nilbert, M. C., Su, L.-K., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K. J., Preisinger, A. C., Hedge, P., McKechnie, D., Finnear, R., Markham, A., Groffen, J., Boguski, M. S., Altschul, S. F., Horii, A., Ando, H., Miyoshi, Y., Miki, Y., Nishisho, I. and Nakamura, Y. (1991b) Identification of FAP locus genes from chromosome 5q21. *Science* **253**: 661-665.

Knudson, A. G. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U.S.A.* **68**: 820-823.

Knudson, A. G. and Strong, L. C. (1972a) Mutation and cancer: a model for Wilm's tumour of the kidney. *J. Natl. Cancer Inst.* **48**(313-324):

Knudson, A. G. and Strong, L. C. (1972b) Mutation and cancer: neuroblastoma and pheochromocytoma. *Amer. J. Hum. Genet.* **24**: 514-532.

Knudson, A. G. (1973) Heredity and cancer in man. *Prog. Med. Genet.* **9**: 113-157.

Knudson, A. G. (1985) Genetics of human cancer. *Ann. Rev. Genet.* **20**: 231-251.

Koenig, M., Hoffman, E. P., Bertleson, C. J., Monaco, A. P., Feener, C. and Kunkel, L. M. (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organisation of the DMD gene in normal and affected individuals. *Cell* **50**: 509-517.

Korenberg, J. and Rykowski, M. C. (1988) Human genome organisation: Alu, Lines and the molecular structure of metaphase chromosome bands. *Cell* **53**: 391-400.

Krimlauf, R., Jeanpierre, M. and Young, B. D. (1982) The construction and characterisation of genomic libraries from specific human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **79**: 2971-2975.

Kunkel, L. M., Monaco, A. P., Middlesworth, W., Ochs, H. D. and Latt, S. A. (1985) Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. *Proc. Natl. Acad. Sci. U.S.A.* **82**: 4778-4782.

Kuspa, A., Volrath, D., Cheng, Y. and Kaiser, D. (1989) Physical mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 8917-8921.

Laird, C. D. (1990) Proposed genetic basis of Huntington's disease. *Trands Genet.* **6**: 242-247.

Landsvater, R. M., Mathew, C. G. P., Smith, B. A., Marcus, E. M., Te-Meerman, G. J., Lips, C. J. M., Geerdink, R. A., Nakamura, Y., Ponder, B. A. J. and Buys, C. H. C. (1989) Development of multiple endocrine neoplasia type 2A does not involve substantial deletions of chromosome 10. *Genomics* **4**: 246-250.

Larin, Z., Monaco, A. P. and Lehrach, H. (1991) Yeast Artificial Chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 4123-4127.

Lasko, D., Cavenee, W. and Nordenskjold, M. (1991) Loss of constitutional heterozygosity in human cancer. *Ann. Rev. Genet.* **25**: 281-314.

Lathrop, G. M., Lalouel, J.-M., Juliet, C. and Ott, J. (1984) Strategies for multi-locus linkage analysis in humans. *Proc. Natl. Acad. Sci. U.S.A.* **81**: 3443-3446.

Law, D. J., Olshwang, S., Monpezat, J.-P., Lefancois, D., Jagelman, D., Petrelli, N. J., Thomas, G. and Feinberg, A. P. (1988) Concerted nonsyntenic allelic loss in human colorectal carcinoma. *Science* **241**: 961-965.

Ledbetter, S. A., Nelson, D. L., Warren, S. T. and Ledbetter, D. H. (1990) Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. *Genomics* **6**: 475-481.

Ledbetter, S., Garcia-Heras, J. and Ledbetter, D. H. (1991) "PCR-karyotype" of human chromosomes in somatic cell hybrids. *Genomics* **8**: 614-622.

Lee, J. T., Murgia, A., Sosnoski, D. M., Olivos, I. M. and Nussbaum, R. L. (1992) Construction and characterisation of a yeast artificial chromosome library for Xpter-Xq27.3: A systematic determination of cocloning rate and X-chromosome representation. *Genomics* **12**: 526-533.

Lee, W.-H., Bookstein, Hong, F., Young, L. J., Shew, J.-Y., Donoso, L. A. and Lee, E. Y.-H. P. (1987) Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science* **235**: 1394-1399.

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

Leppert, M., Burt, R., Hughes, J. P., Samowitz, W., Nakamura, Y., Woodward, S., Gardner, E., Lalouel, J.-M. and White, R. (1990) Genetic analysis of an inherited predisposition to colon cancer in a family with a variable number of adenomatous polyps. *New Eng. J. Med.* **322**: 904-908.

Li, F. P. and Fraumeni, J. F. (1975) Familial breast cancer, soft tissue sarcomas, and other neoplasms. *Annals Intern. Med.* **83**: 833-834.

Litt, M. and White, R. L. (1985) A highly polymorphic locus in human DNA by cosmid-derived probes. *Proc. Natl. Acad. Sci. U.S.A.* **82**: 6206-6210.

Little, R., Porta, G., Carle, G., Schlessinger, D. and D'Urso, M. (1989) Yeast artificial chromosomes with 200- to 800-kilobase inserts of human DNA containing HLA, Vk, 5s and Xq24-Xq28 sequences. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 1589-1602.

Little, R. D., Pilia, G., Johnson, S., D'Urso, M. and Schlessinger, D. (1992) Yeast artificial chromosomes spanning 8 megabases and 10-15 centimorgans of human cytogenetic band Xq26. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 177-181.

Lovett, M., Kere, J. and Hinton, L. M. (1991) Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 9628-9632.

Ludecke, H.-J., Senger, G., Claussen, U. and Horstemke, B. (1989) Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. *Nature* **338**: 348-350.

Ludecke, H.-J., Senger, G., Claussen, U. and Horsthemke, B. (1990) Construction and characterisation of band-specific DNA libraries. *Hum. Genet.* **84**: 512-516.

MacKinnon, R. N., Hirst, M. C., Bell, M. V., Watson, J. E. V., Claussen, U., Ludecke, H.-J., Senger, G., Horsthemke, B. and Davies, K. E. (1990) Microdissection of the fragile X region. *Am. J. Hum. Genet.* **47**: 181-187.

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

Maniatis, T., Fritsch, E. F. and Sambrook, J. Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratories, New York, 1982.

Maxam, A. M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**: 560-564.

McAlpine, P. J., Van Cong, N., Boucheix, C., Pakstis, A. J., Doute, R. C. and Shows, T. B. (1987) The 1987 catalog of mapped genes and the report of the nomenclature committee, Ninth International Workshop on Human Genome Mapping. *Cytogenet. Cell Genet.* **46**: 29-101.

McCormick, M. K., Shero, J. H., Connelly, C., Antonarakis, S. E. and Hieter, P. A. (1990) Methods for cloning large DNA segments as artificial chromosomes in *S. cerevisiae*. *Technique* **2**: 65-69.

McKusick, V. A. Mendelian inheritance in man. John's Hopkins University Press, Baltimore, 1991.

Meera-Khan, P., Tops, C., van den Broek, M. H., Breukel, C., Oldenburg, M., Wijnen, J. T., Griffioen, G., Verspaget, H., Lamers, C. B. H. W., van Haeringen, A., Hampe, A., Galibert, G., Deitzsch, E. and Pearson, P. L. (1987) Linkage between familial polyposis coli (FPC), M4 (D5S6) and FMS on chromosome 5, Ninth International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet.* **46**: 661-661.

Menon, A. G., Anderson, K. M., Ricardi, V. M., Chung, R. Y., Whaley, J. M., Yandell, D. W., Farmer, G. E., Freiman, R. N., Lee, J. K., Li, F. P., Barker, D. F., Ledbetter, D. H., Kleider, A., Martuza, R. L., Gusella, J. F. and Seizinger, B. R. (1990) Chromosome 17p deletions and p53 gene mutations associated with the formation of malignant neurofibrosarcomas. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 5435-5439.

Michiels, F., Burmeister, M. and Lehrach, H. (1987) Derivation of clones close to met by preparative field inversion gel electrophoresis. *Science* **236**: 1305-1308.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., Vogelstein, B. and Nakamura, Y. (1992) Disruption of the *APC* gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52**: 643-645.

Miller, H. Practical aspects of preparing phage and plasmid DNA: Growth, maintenance, and storage of bacteria and bacteriophage. pp. 145-172 in *Guide to Molecular Cloning Techniques.* Edited by Berger, S. L. and Kimmel, A. R., Academic Press, Florida, 1987.

Mitelman, F. Catalog of chromosome aberrations in cancer. Liss, New York, 1988.

Miyaki, M., Seki, M., Okamoto, M., Yamanaka, A., Maeda, Y., Tanaka, K., Kikuchi, R., Iwama, T., Ikeuchi, T., Tonomura, A., Nakamura, Y., White, R., Miki, Y., Utsunomiya, J. and Koike, M. (1990) Genetic changes and

histopathological types in colorectal tumours from patients with familial adenomatous polyposis. *Cancer Res.* **50**: 7166-7173.

Monaco, A. P., Neve, R. L., Colletti-Feener, C., Bertelson, C. J., Kurnit, D. M. and Kunkel, L. M. (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323**: 646-650.

Monaco, A. P. and Kunkel, L. M. (1987) A giant locus for the Duchenne and Becker muscular dystrophy gene. *Trends Genet.* **3**: 33-39.

Monaco, A. P., Walker, A. P., Millwood, I., Larin, Z. and Lehrach, H. (1992) A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy gene. *Genomics* **12**: 465-473.

Monpezat, J.-P., Delattre, O., Bernard, A., Grunwald, D., Remvikos, Y., Muleris, M., Salmon, R. J., Frelat, G., Dutrillaux, B. and Thomas, G. (1988) Loss of alleles on chromosome 18 and on the short arm of chromosome 17 in polypoid colorectal carcinomas. *Int. J. Cancer* **41**: 404-408.

Morgan, T. H. (1911) Random segregation versus coupling in mendelian inheritance. *Science* **34**: 384-389.

Morson, B. C. (1966) Factors influencing the prognosis of early cancer of the rectum. *Proc. Roy. Soc. Med.* **59**: 607-608.

Morson, B. (1974a) The polyp-cancer sequence in the large bowel. *Proc. Roy. Soc. Med.* **67**: 451-457.

Morson, B. C. (1974b) Evolution of cancer of the colon and rectum. *Cancer* **34**: 845-849.

Morson, B. C., Bussey, H. J. R., Day, D. W. and Hill, M. J. (1983) Adenomas of large bowel. *Cancer Surveys* **2**: 451-477.

Morton, N. E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**: 277-318.

Morton, N. E. (1957) Further scoring types in sequential linkage tests with a critical review of autosomal and partial sex linkage in man. *Am. J. Hum. Genet.* **9**: 55-75.

Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J. R., Burks, C., Sirotkin, K. M. and Goad, W., B. (1989) The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* **4**: 273-289.

Muleris, M., Salmon, R. J., Zafrani, B., Girodet, J. and Dutrillaux, B. (1985) Consistent deficiencies of chromosome 18 and of the short arm of chromosome 17 in eleven cases of human large bowel cancer: a possible recessive determinism. *Ann. Genet. (Paris)* **28**: 206-213.

Murday, V. A. and Slack, J. (1989) Inherited disorders associated with colorectal cancer. *Cancer Surveys* **8**: 139-157.

Murray, A. W. and Szostac, J. (1983) Construction of artificial chromosomes in yeast. *Nature* **305**: 189-193.

Nabholz, M., Miggiano, V. and Bodmer, W. F. (1969) Genetic analysis using human-mouse somatic cell hybrids. *Nature* **223**: 358-363.

Nakamura, Y., Lathrop, M., Leppert, M., Dobbs, M., Wasmuth, J. J., Wolff, E., Carlson, M., Fujimoto, E., Krapcho, K., Sears, T., Woodward, S., Hughes, J., Burt, R., Gardner, E., Lalouel, J.-M. and White, R. (1988) Localisation of the genetic defect in familial adenomatous polyposis within a small region of chromosome 5. *Amer. J. Hum. Genet.* **43**: 638-644.

Needleman, S. W., Yuasa, Y., Srivastava, S. and Aaronson, S. A. (1983) Normal cells of patients with high cancer risk syndromes lack transforming activity in the NIH/3T3 transfection assay. *Science* **222**: 173-175.

Nelkin, B. D., Nakamura, Y., White, R., deBustros, A. C., Herman, J., Wells, S. A. and Baylin, S. B. (1989) Low incidence of loss of chromosome 10 in sporadic and hereditary human medullary thyroid carcinoma. *Cancer Res.* **49**: 4114-4119.

Nelson, D. L., Ledbetter, S. A., Corbo, L., Victoria, M. F., Ramirez-Solis, R., Webster, T. D., Ledbetter, D. H. and Caskey, C. T. (1989) *Alu*-polymerase chain

reaction: A method for the rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 6686-6690.

Nelson, D. L. (1990) Current methods for YAC clone characterisation. *Gen. Anal. Tech. Appl.* **7**: 100-106.

Nelson, D. L., Ballabio, A., Victoria, M. F., Pieretti, M., Bies, R. D., Gibbs, R. A., Maley, J. A., Chinault, A. C., Webster, T. D. and Caskey, C. T. (1991) *Alu*-primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: identification of clones associated with a disease locus. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 6157-6161.

Nigro, J. M., Baker, S. J., Preisinger, A. C., Jessup, J. M., Hostetter, R., Cleary, K., Bigner, S. H., Davidson, N., Baylin, S., Deville, P., Glover, T., Collins, F. S., Weston, A., Modali, R., Harris, C. C. and Vogelstein, B. (1989) Mutations in the p53 gene occur in diverse human tumour types. *Nature* **342**: 705-707.

NIH workshop on population screening for the cystic fibrosis gene (1990). *New Eng. J. Med.* **323**: 70-71.

Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., Hedge, P., Markham, A., Krush, A. J., Petersen, G., Hamilton, S. R., Nilbert, M. C., Levy, D. B., Bryan, T. M., Preisinger, A. C., Smith, K. J., Su, L.-K., Kinzler, K. W. and Vogelstein, B. (1991) Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* **253**: 655-670.

Nizetic, D., Zehetner, G., Monaco, A. P., Young, B. D. and Lehrach, H. (1991) Construction, arraying and high density display of large insert libraries of the human chromosomes X and 21; their potential use as reference libraries. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 3233-3237.

Nordling, C. O. (1953) A new theory on the cancer-inducing mechanism. *Br. J. Cancer* **7**: 68-72.

Okamoto, M., Sasaki, M., Sugio, K., Sato, C., Iwama, T., Ikeuchi, T., Tonomura, A., Sasazuki, T. and Miyaki, M. (1988) Loss of constitutional

heterozygosity in colon carcinoma from patients with familial polyposis coli. *Nature* **331**: 273-277.

Okamoto, M., Sato, C., Kohno, Y., Mori, T., Iwama, T., Tonomura, A., Miki, Y., Utsunomiya, J., Y., N., White, R. and Miyaki, M. (1990) Molecular nature of chromosome 5q loss in colorectal tumors and desmoids from patients with familial adenomatous polyposis. *Hum. Genet.* **85**: 595-599.

Pachnis, V., Pevney, L., Rothstein, R. and Constantini, F. (1990) Transfer of a yeast artificial chromosome carrying human DNA from *Saccharomyces cerevisiae* into mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 5109-5113.

Parada, L. F., Land, H., Weinberg, R. A., Wolf, D. and Rotter, V. (1984) Human EJ bladder carcinoma oncogene is a homolouge of Harvey sarcoma virus *ras* gene. *Nature* **297**: 474-478.

Parimoo, S., Patanjali, S. R., Shukla, H., Chaplin, D. D. and Weissman, S. M. (1991) cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* 88: 9623-9627.

Patel, P. I., Garcia, C., Montes de Oca-Luna, R., Malamut, R. I., Franco, B., Slaugenhaupt, S., Chakravarti, A. and Lupski, J. R. (1990) Isolation of a marker linked to the Charcot-Marie-Tooth disease Type 1A gene by differential *Alu*-PCR of human chromosome 17-retaining hybrids. *Am. J. Hum. Genet.* **47**: 926-934.

Pavan, W. J., Hieter, P. and Reeves, R. H. (1990) Modification and transfer into an embryonal carcinoma cell line of a 360-kilobase human-derived yeast artificial chromosome. *Mol. Cel. Biol.* **10**: 4163-4169.

Peto, R. and Doll, R. (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the USA today. *J. Natl. Cancer Inst.* **66**: 1192-1221.

Pohl, T. M., Zimmer, M., MacDonnald, M. E., Smith, B., Bucan, M., Poustka, A.-M., Voliana, S., Searle, S., Zehetner, G., Wasmuth, J. J., Gusella, J., Lehrach, H. and Frischauf, A.-M. (1988) Construction of a NotI linking library and isolation of new markers close to the Huntington's disease gene. *Nucl. Acids Res.* **16**: 9185-9197.

Polkinghorne, P. J., Ritchie, S., Neale, K., Schoeppner, G., Thomson, J. P. S. and Jay, B. S. (1990) Pigmented lesions of the retinal pigment epithelium and familial adenomatous polyposis. *Eye* **4**: 216-221.

Ponce, M. R. and Micol, J. L. (1992) PCR amplification of long DNA fragments. *Nucl. Acids Res.* **20**: 623-623.

Ponder, B. A. J., Schmidt, G. H., Wilkinson, M. M., Wood, M. J., Monk, M. and Reid, A. (1987) Derivation of mouse intestinal crypts from single progenitor cells. *Nature* **313**: 689-691.

Pontecorvo, G. (1975) Production of mammalian somatic cell hybrids by means of polyethylene glycol. *Somatic Cell Genet.* **1**: 397-400.

Pratt, C. B., Parham, D. M., Rao, B. N., Fleming, I. D. and Dilawari, R. (1988) Multiple colorectal carcinomas, polyposis coli, and neurofibromatosis. *J. Natl. Cancer Inst.* **14**: 1170-1171.

Pritchard, K., Casher, D., Bull, L., Cox, D. R. and Myers, R. M. (1990) A cloned DNA segment from the telomeric region of human 4p is not detectably rearranged in Huntington's disease patients. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 7309-7313.

Pullman, W. E. and Bodmer, W. F. (1992) Cloning and characterisation of a gene that regulates cell adhesion. *Nature* **356**: 529-532.

Ragoussis, J., Monaco, A. P., Mockridge, I., Kendall, E., Campbell, R. D. and Trowsdale, J. (1991) Cloning of the HLA class II region in yeast artificial chromosomes. *Proc. Natl. Acad. Sci.* **88**: 3753-3757.

Reed, T. E. and Neel, J. V. (1955) A genetic study of multiple polyposis of the colon (with an appendix deriving a method of estimating relative fitness). *Am. J. Hum. Genet.* **7**: 236-263.

Rees, M., Leigh, S. E. A., Delhanty, J. D. A. and Jass, J. R. (1989) Chromosome 5 allele loss in familial and sporadic colorectal adenomas. *Br. J. Cancer* **59**: 361-365.

Reeves, R. H., Pavan, W. J. and Hieter, P. (1990) Modification and manipulation of mammalian DNA cloned as YACs. *Gen. Anal. Tech. Appl.* **7**: 107-113.

Rhome, D., Fox, H., Herrmann, B., Frischauf, A.-M., Edstrom, J. E., Mains, P., Silver, L. M. and Lehrach, H. (1984) Molecular clones of the mouse t-complex region derived by microdissected metaphase chromosomes. *Cell* **36**: 783-788.

Richards, R. I. and Sutherland, G. R. (1992) Heritable unstable DNA sequences. *Nature Genet.* **1**: 7-9.

Rienhard, F. P., Ritch, T. G., Deininger, P. L. and Schmid, C. W. (1981) Renaturation rate studies of a single family of interspersed repeat sequences in human DNA. *Biochem.* **20**: 3003-3010.

Riley, J., Butler, R., Ogilvie, D., Finnear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C. and Markham, A. F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosomes. *Nuc. Acids Res.* **18**: 2887-2890.

Rodrigues, N. R., Rowan, A., Smith, M. E. F., Kerr, I. B., Bodmer, W. F., Gannon, J. V. and Lane, D. P. (1990) p53 mutations in colorectal cancer. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 7555-7559.

Rommens, J. M., Iannuzzi, M. C., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J. R., Tsui, L.-C. and Collins, F. C. (1990) Identification of the cystic fibrosis gene: Chromosome walking and jumping. *Science* **245**: 1059-1065.

Rubock, M. J., Larin, Z., Cook, M., Papalopulu, N., Krumlauf, R. and Lehrach, H. (1990) A yeast artificial chromosome containing the mouse homeobox cluster *Hox-2*. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 4751-4755.

Sager, R. (1986) Genetic suppression of tumor formation: A new frontier in cancer research. *Cancer Res.* **46**: 1573-1580.

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

Sasaki, M., Okamoto, M., Sato, C., Sugio, K., Soejima, J.-I., Iwama, T., Ikeuchi, T., Tonomura, A., Miyaki, M. and Sasazuki, T. (1989) Loss of constitutional heterozygosity in colorectal tumors from patients with familial polyposis coli and those with nonpolyposis colorectal carcinoma. *Cancer Res.* **49**: 4402-4406.

Saxon, P. J., Srivatsan, E. S. and Stanbridge, E. J. (1986) Introduction of human chromosome 11 via microcell mediated chromosome transfer controls tumorigenic expression of HeLa cells. *EMBO J.* **5**: 3461-3466.

Scalenghe, F., Turco, E., Edstrom, J. E., Pirrotta, V. and Melli, M. (1981) Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes. *Chromosoma* **82**: 205-216.

Schlessinger, D. (1990) Yeast artificial chromosomes: tools for mapping and analysis of complex genomes. *Trends Genet.* **6**: 248-258.

Schlessinger, D., Little, R. D., Freije, D., Abidi, F., Zucchi, I., Porta, G., Pilia, G., Nagaraja, R., Johnson, S. K., Yoon, J.-Y., Srivastava, A., Kere, J., Palmieri, G., Ciccodicola, A., Montanaro, V., Romano, G., Casamassimi, A. and D'Urso, M. (1991) Yeast artificial chromosome-based genome mapping: some lessons from Xq24-q28. *Genomics* **11**: 783-793.

Schmidt, M., Haaf, T. and Grunert, D. (1982) 5-azacytidine-induced undercondensations in human chromosomes. *Hum. Genet.* **67**: 257-263.

Schwartz, D. C. and Cantor, C. R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**: 67-75.

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

Seed, B. and Aruffo, A. (1987) Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. U.S.A.* **84**: 3365-3369.

Seizinger, B. R., Klinger, H. P., Junien, C., Nakamura, Y., Beau, M. M., Cavenee, W., Emmanuel, B., Ponder, B., Naylor, S., Mitelman, F., Louis, D., Menon, A., Newsham, I., Decker, J., Kaebling, M., Henry, I. and Deimling, A.

(1992) Report on the committee on chromosome and gene loss in human neoplasia. Human Gene Mapping 11. *Cytogenet. Cell Genet.* **58**: 1080-1096.

Senger, G., Ludecke, H.-J., Horsthemke, B. and Claussen, U. (1990) Microdissection of banded human chromosomes. *Hum. Genet.* **84**: 507-511.

Silverman, G. A., Ye, R. D., Pollock, K. M., Sadler, J. E. and Korsmeyer, S. J. (1989) Use of yeast artificial chromosome clones for mapping and walking human chromosome segment 18q21.3. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 7485-7489.

Silverman, G., Green, E. D., Young, R. L., Jockel, J. L., Domer, P. H. and Korsmeyer, S. (1990) Meiotic recombination between yeast artificial chromosomes yields a single clone containing the entire *BCL2* protooncogene. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 9913-9917.

Silverman, G., Jockel, J. I., Dohmer, P. H., Mohr, R. M., Taillon-Miller, P. and Korsmeyer, S. (1991) Yeast artificial chromosome cloning of a two-megabase-size contig within chromosomal band 18q21 establishes physical linkage between BCL2 and plasminogen activator inhibitor type-2. *Genomics* **9**: 219-228.

Smith, D. R., Smyth, A. P. and Moir, D. T. (1990) Amplification of large artificial chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 8242-8246.

Smith, H. O. and Wilcox, A. (1970) A restriction enzyme from Haemophilus influenzae. *J. Mol. Biol.* **51**: 379-391.

Solomon, E. and Bodmer, W. F. (1979) Evolution of the sickle cell variant. *Lancet* **1**: 923-923.

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

Solomon, E., Borrow, J. and Goddard, A. D. (1992) Chromosome aberrations in cancer. *Science* **254**: 1153-1160.

Sondergaard, J. O., Bulow, S. and Lynge, E. The incidence of colorectal cancer among first degree relatives of patients with colorectal cancer. pp. 157-158 in *Herediatry colorectal cancer.* Edited by Utsunomiya, J. and H.T., L., Springer-Verlag, Tokyo, 1989.

Southern, E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-512.

Sparkes, R. S., Sparkes, M., Wilson, M. G., Towner, J. W. and Benedict, W. F. (1980) Regional assignment of genes for esterase D and retinoblastoma to chromosome band 13q14. *Science* **208**: 1042-1044.

Sparkes, R. S., Murphee, A. L., Lingua, R. W., Sparkes, M. C., Field, L. L., Funderbuck, S. J. and Benedict, W. F. (1983) Gene for hereditary retinoblastoma assigned to chromosome 13 by linkage to esterase D. *Science* **219**: 971-973.

Spurr, N. K., Kelsell, D., Rooke, L., Cavalli-Sforza, L. L., Bowcock, A. and Feder, J. (1991) Assignment of eight human genomic DNA sequences detecting high frequency restriction fragment length polymorphisms. *Ann. Hum. Genet.* **55**: 141-150.

Stanbridge, E. J., Flandermeyer, R. P., Daniels, D. W. and Nelson-Rees, W. (1981) Specific chromosome loss associated with the expression of tumorigenicity in human cell hybrids. *Somatic Cell Mol. Genet.* **7**: 699-712.

Stanbridge (1990) Human tumor suppressor genes. *Ann. Rev. Genet.* **24**: 615-657.

Tabin, C. J., Bradely, S. M., Bargmann, C. I., Weinberg, R. A., Papageorge, A. G., Scolnick, E. M., Dhar, R., Lowy, D. R. and Chang, E. H. (1982) Mechanism of activation of a human oncogene. *Nature* **300**: 143-149.

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

Thomas, H. J. W. (1991) Ph.D., London University.

Thompson, J. P. S. (1988) Leeds Castle polyposis group meeting report. *Dis. Col. Rect.* **31**: 613-616.

Trautmann, U., Leuteritz, G., Senger, G., Claussen, U. and Balhaussen, W. G. (1991) Detection of *APC* region-specific signals by nonisotopic chromosomal in situ suppression (CISS)-hybridisation using a microdissection library as a probe. *Hum. Genet.* **87**: 495-497.

Traver, C. N., Klapholz, S., Hyman, R. W. and Davis, R. W. (1989) Rapid screening of a human genomic library in yeast artificial chromosomes for single copy sequences. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 5898-5902.

Utsunomiya, J. and Nakamura, T. (1975) The occult osteomatous changes in the mandible in patients with familial polyposis coli. *Br. J. Surg.* **62**: 45-51.

Utsunomiya, J. The concept of hereditary colorectal cancer and the implications of its study. pp. 3-16 in *Hereditary colorectal cancer*. Edited by Utsonomyia, J. and Lynch, H. T., Springer-Verlag, Tokyo, 1989.

van Arsdell, S. W., Denison, R. A., Berstein, L. B., Weiner, A. M., Manser, T. and Gestland, R. F. (1981) Direct repeats flank 3 small nuclear RNA pseudogenes in the human genome. *Cell* **26**: 11-17.

Varesco, L., Thomas, H. W. J., Cottrell, S., Murday, V., Fennell, S. J., Williams, S., Searle, S., Sheer, D., Bodmer, W. F., Frischauf, A.-M. and Solomon, E. (1989) CpG island clones from a deletion encompassing the gene for adenomatous polyposis coli. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 10118-10122.

Varmus, H. (1988) Retroviruses. *Science* **240**: 1427-1435.

Veal, A. M. O. Intestinal polyposis. Cambridge University Press, Cambridge, 1965.

Viskochil, D., Buchberg, A. M., Xu, G., Cawthorn, R. M., Stevens, J., Wolff, R. K., Culver, M., Carey, J. C., Copeland, N. G., Jenkins, N. A., White, R. and O'Connell, P. (1990) Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell* **62**: 187-192.

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. M. and Bos, J. L. (1988) Genetic alterations during colorectal-tumor development. *New Eng. J. Med.* **319**: 523-532.

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

Wallace, M., Marchuk, D. A., Andersen, L. B., Letcher, R., Odeh, H. M., Saulino, A. M., Fountain, J. W., Brereton, A., Nicholson, J., Mitchell, A. L., Brownstein, B. H. and Collins, F. S. (1990) Type 1 Neurofibromatosis gene: Identification of a large transcript disrupted in three NF1 patients. *Science* **249**: 181-186.

Washauer, M. E., Silverman, D. T., Schottenfeld, D. and Pollack, E. S. (1986) Stomach and colorectal cancers in Puerto Rican-born residents of New York City. *J. Natl. Cancer Inst.* **76**: 591-595.

Weatherall, D. J. (1990) Molecular genetics: mice, men and sickle cells. *Nature* **343**: 121.

Weber, J., Weith, A., Kaiser, R., Grzeschik, K.-H. and Olek, K. (1990) Microdissection and microcloning of human chromosome 7q22-32 region. *Somatic Cell Mol. Genet.* **16**: 123-128.

Weinberg, R. A. (1992) Tumour suppressor genes. *Science* **254**: 1138-1146.

Weissman, B. and Stanbridge, E. J. (1987) Complementation of the tumorigenic phenotype in human cell hybrids. *Natl. Cancer Inst.* **70**: 667-673.

White, R. L. and Lalouel, J.-M. (1988) Sets of linked genetic markers for human chromosomes. *Ann. Rev. Genet.* **22**: 259-279.

Wilkes, D., Shaw, J., Anand, R., Riley, J., Winter, P., Wallis, J., Driesel, A. G., Williamson, R. and Chamberlain, S. (1991) Identification of CpG islands in a physical map encompassing the Friedreich's Ataxia locus. *Genomics* **9**: 90-95.

Willett, W. (1989) The search for the causes of breast and colon cancer. *Nature* **338**: 389-394.

Williams, S., Jones, T. A., Cottrell, S., Zehetner, G., Varesco, L., Ward, T., Thomas, H., Lawson, P. A., Solomon, E., Bodmer, W. F., Frischauf, A.-M. and Sheer, D. (1991) Fine mapping of probes in the Adenomatous Polyposis Coli region of chromosome 5 by *in situ* hybridisation. *Genes, chromosomes and cancer* **3**: 382-389.

Yoshida, M. A., Ikeuchi, T., Iwama, T., Miyaki, M., Mori, T., Ushijima, Y., Hara, A., Miyakita, M. and Tonomura, A. (1991) Chromosome changes in desmoid tumours developed in patients with familial adenomatous polyposis. *Jpn. J. Cancer Res.* **82**: 916-921.

Zucchi, I. and Schlessinger, D. (1992) Distribution of moderately repetitive sequences pTR5 and LF1 in Xq24-q28 human DNA and their use in assembling YAC contigs. *Genomics* **12**: 264-275.

zur Hausen, H. (1992) Viruses in human cancer. *Science* **254**: 1167-1173.

Publications arising from this thesis.

Hampton, G., Leuteritz, G., Ludecke, H.-J., Senger, G., Trautmann, U., Thomas, H., Solomon, E., Bodmer, W. F., Horsthemke, B., Claussen, U. and Ballhausen, W. G. (1991a) Characterisation and mapping of microdissected genomic clones from the adenomatous polyposis coli (*APC*) region. *Genomics* **11**: 247-251.

Hampton, G., Howe, K., Leuteritz, G., Thomas, H., Bodmer, W. F., Solomon, E. and Ballhausen, W. G. (1991b) Regional mapping of 22 microclones to the adenomatous polyposis coli (*APC*) locus on chromosome 5q. *Hum. Genet.* **88**: 112-114.

Hampton, G. M., Ward, J. T. R., Cottrell, S., Howe, K., Thomas, H. J. W., Jones, T., Sheer, D., Ballhausen, W. G., Frischauf, A.-M., Solomon, E. and Bodmer, W. F. (1992) Yeast artificial chromosomes (YACs) for the molecular analysis of the familial polyposis *APC* gene region. *Proc. Natl. Acad. Sci.*, submitted.

Cotter, F. E., Hampton, G. M., Nasipuri, S., Bodmer, W. F. and Young, B. D. (1990) Rapid isolation of human chromosome-specific DNA probes from a somatic cell hybrid. *Genomics* **7**: 257-263.