A Molecular Genetic Analysis of the Chromosome 17q21 Region Containing the Familial Breast/Ovarian Cancer Susceptibility Gene, BRCA1

by

Karen A Jones

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

Somatic Cell Genetics Laboratory
Imperial Cancer Research Fund
44, Lincoln's Inn Fields
London

and

Department of Genetics and Biometry
University College
Gower Street
London

March 1995
ABSTRACT

Breast cancer is the most common malignancy in women, affecting about one in ten in the Western world. Although most cases are sporadic, around 5% are due to inherited susceptibility genes. One such gene, termed *BRCA1* may account for up to 1/3 of all cases diagnosed before 30 years and in addition appears to be important in the onset of some cases of ovarian cancer. In families in which both breast and ovarian cancer cases are inherited, *BRCA1* appears to be almost always responsible.

With the lack of positional, biochemical and functional clues to the identity of *BRCA1*, its isolation required a positional cloning strategy. The work described here commenced with the physical characterisation of the chromosome 17q21 region containing *BRCA1* and progressed through the various stages of the positional cloning strategy with the aim of isolating the responsible gene.

The initial part of this work was concerned with the construction of a detailed long-range restriction map spanning over 3.5 megabase pairs of the *BRCA1* region. This map detailed the precise location of previously isolated genes and markers in addition to the identification of the putative locations of other genes in the region. This information proved crucial to the construction of genomic clone contigs by orientating genomic clone walks and estimating the distances between them.

As part of the next stage in the positional cloning process, this project concentrated on a 1.0-1.5 megabase pair stretch of the region which appeared to be the most likely location of the *BRCA1* gene, as suggested by parallel genetic studies. This work involved the construction of a complete genomic clone contig spanning the distal half of this region to complement other work in the laboratory concerned with cloning the remaining half of the region. A thorough transcriptional analysis of this contig was carried out as part of this project and a number of new genes in the region were isolated with the aid of a variety of gene isolation techniques including direct selection, exon-trapping, surveying for evolutionary conservation and the identification of CpG islands. The work culminated in the mutational analysis of a strong candidate in the region.

With the recent identification of a novel gene bearing all the requirements for *BRCA1*, the detailed physical characterisation described here of the region surrounding this gene will be useful in future efforts to reveal the normal functions of this gene and how it causes oncogenesis.
# TABLE OF CONTENTS

Title

Abstract ............................................. (i)

Table of contents .................................... (ii)

Abbreviations ......................................... (xii)

Acknowledgements ..................................... (xv)

CHAPTER 1: INTRODUCTION ......................... 1

1.1 From cancer predisposition to malignancy: accumulation of multiple 
defects in tumour suppressor genes and oncogenes .......... 1
   1.1.1 The roles of oncogenes genes in malignancy ............. 2
   1.1.2 Ther oles of tumour suppressor genes in malignancy ........ 2
   1.1.3 The recessive model of cancer inheritance .............. 3
   1.1.4 The multistep nature of cancer ........................... 5

1.2 Cancer families and the genes responsible ............... 9
   1.2.1 The variety of family cancer genes ...................... 9

1.3 Methods of isolating cancer genes ....................... 13
   1.3.1 Functional methods of gene isolation ................... 13
   1.3.2 Subtractive hybridisation ............................... 14
   1.3.3 Positional cloning .................................... 14

1.4 Breast cancer predisposition .......................... 18
   1.4.1 Evidence for a genetic component ...................... 18
   1.4.2 Breast cancer susceptibility genes .................... 18
   1.4.3 The breast cancer susceptibility locus, \textit{BRCA1} ........ 20

1.5 Isolation of \textit{BRCA1} by positional cloning ........... 22
   1.5.1 Generation of a genetic map of the \textit{BRCA1} region. .... 22
   1.5.2 Hunting for cytogenetic abnormalities: clues to \textit{BRCA1} location. ........................................ 25
   1.5.3 Previously characterised candidate genes in the region. ........................................ 26
1.5.4 Construction of a physical map  .......................................................... 27
1.5.5 Genomic cloning of the BRCA1 region ................................................. 30
1.5.6 Gene isolation: identification of new candidates for BRCA1 .................. 32
1.5.7 Genes discovered in the BRCA1 region .............................................. 36
1.5.8 Mutation hunting .................................................................................. 40
1.5.9 The BRCA1 gene .................................................................................. 43

Aims of thesis .................................................................................................. 46

CHAPTER 2: MATERIALS AND METHODS .................................................... 47

2.1 DNA preparation ........................................................................................ 47
  2.1.1 Genomic DNA ..................................................................................... 47
  2.1.2 High molecular weight DNA for PFGE analysis .................................. 48
  2.1.3 Bacteriophage DNA ........................................................................... 48
  2.1.4 Cosmid and plasmid clone DNA .......................................................... 49
  2.1.5 YAC DNA .......................................................................................... 51

2.2 DNA restriction endonuclease digestion .................................................. 51
  2.2.1 Genomic DNA for conventional Southern analysis ............................... 51
  2.2.2 PFGE blocks ........................................................................................ 52
  2.2.3 Genomic clone and cDNA clone digestion .......................................... 52

2.3 Electrophoretic size separation of DNA ................................................... 52
  2.3.1 Conventional agarose gel electrophoresis .......................................... 52
  2.3.2 PFGE electrophoresis ......................................................................... 53

2.4 Preparation of radioactive probes .............................................................. 53
  2.4.1 DNA probes ......................................................................................... 53
  2.4.2 Cosmid riboprobes .............................................................................. 54

2.5 Southern analysis ....................................................................................... 55
  2.5.1 Alkali transfer ..................................................................................... 55
  2.5.2 Prehybridisation and hybridisation of filters ....................................... 56
  2.5.3 Washing of Southern filters .................................................................. 56

2.6 Northern analysis ....................................................................................... 57
  2.6.1 Prehybridisation and hybridisation of filters ....................................... 57
  2.6.2 Washing northern filters ...................................................................... 57
2.7 Subcloning ................................................................. 58
  2.7.1 Preparation of insert and vector DNA ............................ 58
  2.7.2 Preparation of chemically competent cells ....................... 58
  2.7.3 Ligation and plating of transformants ............................ 58
  2.7.4 Screening transformants ............................................. 59

2.8 cDNA and genomic clone isolation .................................... 59
  2.8.1 Screening of gridded libraries ..................................... 59
  2.8.2 Screening of random plated bacteriophage cDNA libraries ...... 60
  2.8.3 Plasmid rescue from λZAP phage clones ......................... 61

2.9 PCR analysis ................................................................ 62
  2.9.1 General PCR ............................................................ 62
  2.9.2 Subclone PCR analysis ............................................... 62

2.10 SSCP analysis ............................................................. 63
  2.10.1 Preparation of PCR products ...................................... 63
  2.10.2 Acrylamide gel electrophoresis .................................... 63

2.11 DNA sequencing .......................................................... 64
  2.11.1 Manual sequencing .................................................. 64
  2.11.2 Automated sequencing .............................................. 65

2.12 DNA sequence analysis .................................................. 66

2.13 Buffers and solutions ..................................................... 66

CHAPTER 3: PHYSICAL MAPPING OF THE BRCA1 REGION ............... 70

3.1 Construction of a physical map at the RARA locus .................... 72
  3.1.1 Isolation of linking clones from the RARA region ............. 72
  3.1.2 Localisation of RARA, TOP2 and linking clone 2H10 using the somatic cell hybrid mapping panel .......... 73
  3.1.3 Construction of a long range restriction map by PFGE analysis ............................................. 73
  3.1.4 Localisation of the insulin-like growth factor binding protein gene, IGFBP-4 .............................. 77
  3.1.5 Comparison to YAC data ............................................. 77

3.2 Construction of a physical map around EDH and D17S855 ............ 77
  3.2.1 Initial ordering of clones ........................................... 77
  3.2.2 PFGE analysis of the EDH/D17S855 region ....................... 79
3.3 Construction of a long range restriction map from
RNU2 to D17S858
.............................................................. 79

3.3.1 Comparison of PFGE data from IAl.3B and RNU2 genes and the construction of a restriction map ........................................... 81

3.3.2 Joining long range restriction maps at EDH/D17S855 and IAl.3B/RNU2 using an interstitial exon trap product .......................... 81

3.3.3 Comparison of the PFGE-generated physical map to PACs in the region ................................................................. 84

3.4 Physical mapping of PPY to GP2B
.............................................................. 84

3.4.1 Single copy DNA probes generated in the region .............................................................. 85

3.4.2 Results of PFGE Southern hybridisations .............................................................. 86

3.4.3 Construction of a long-range restriction map of the region .............................................................. 87

3.5 Summary of physical mapping analysis of the BRCA1 region ......................................................... 89

CHAPTER 4: INVESTIGATION OF THE FEATURES OF THE PHYSICAL MAP AROUND BRCA1 IDENTIFIED BY PFGE ANALYSIS .............................................................. 91

4.1 Analysis of a CpG island at D17S183.
.............................................................. 91

4.1.1 Identification of a CpG island cluster .............................................................. 92

4.1.2 Isolation of corresponding cDNA clones .............................................................. 92

4.1.3 Localisation of the human upstream binding factor (hUBF) gene .............................................................. 94

4.2 Length heterogeneity of the RNU2 gene cluster and identification of rare-cutter restriction fragment polymorphisms .............................................................. 96

4.2.1 Polymorphic hybridisation patterns around the RNU2 locus .............................................................. 96

4.2.2 Size variation of the intact U2 gene array at the RNU2 locus .............................................................. 99

4.3 Discussion of the features of the BRCA1 region identified by PFGE analysis .............................................................. 99
## CHAPTER 5: GENOMIC CLONING IN THE BRCA1 REGION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Conversion of YAC clones to cosmid contigs</td>
<td>100</td>
</tr>
<tr>
<td>5.1.1 Generation and ordering of a cosmid pool from YAC 12H4</td>
<td>101</td>
</tr>
<tr>
<td>5.1.2 Generation and ordering of a cosmid pool from YAC 26D6</td>
<td>103</td>
</tr>
<tr>
<td>5.2 Assessment of the gap between cosmid contigs</td>
<td>105</td>
</tr>
<tr>
<td>5.3 Construction of a 300 kb cosmid walk in the BRCA1 region</td>
<td>107</td>
</tr>
<tr>
<td>5.3.1 Generation of a cosmid walk from D17S858/D17S859 in the direction of the centromere</td>
<td>107</td>
</tr>
<tr>
<td>5.3.2 Generation of a cosmid walk from RNU2 in the direction of the telomere</td>
<td>110</td>
</tr>
<tr>
<td>5.4 Meeting of the cosmid walks between RNU2 and D17S858/D17S859</td>
<td>112</td>
</tr>
<tr>
<td>5.5 Confirmation of the cosmid walk integrity and physical mapping by FISH</td>
<td>112</td>
</tr>
</tbody>
</table>

## CHAPTER 6: LOCALISATION OF GENES BETWEEN 1A1.3B TO D17S78 AND CONSTRUCTION OF A DETAILED TRANSCRIPT MAP OF THE REGION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Identification of putative CpG islands in the region and the likely locations of genes in the cosmid contig.</td>
<td>115</td>
</tr>
<tr>
<td>6.2 Isolation of the VHR phosphatase gene, HDSPH</td>
<td>116</td>
</tr>
<tr>
<td>6.2.1 Isolation of cDNA clones</td>
<td>116</td>
</tr>
<tr>
<td>6.2.2 Complete sequence of HDSPH</td>
<td>120</td>
</tr>
<tr>
<td>6.3 Localisation of clusters of cDNA clones throughout the region</td>
<td>120</td>
</tr>
<tr>
<td>6.4 Identification of separate gene loci by northern analysis</td>
<td>123</td>
</tr>
<tr>
<td>6.5 Discussion of the transcriptional analysis and gene isolation work</td>
<td>129</td>
</tr>
</tbody>
</table>
CHAPTER 7: SEARCHING FOR DISRUPTIVE MUTATIONS IN THE
GERMLINE DNA OF AFFECTED FAMILY MEMBERS AND SPORADIC
TUMOURS .................................................131

7.1 Searching for gross mutations by PFGE analysis of genomic DNA
from affected family members ........................................132
7.2 Rapid surveying for mutations by conventional Southern analysis ........................................135
7.3 Analysis of the candidate gene HDSPH for causative mutations ........................................135
7.4 Discussion of the mutation analysis results ........................................140

CHAPTER 8: DISCUSSION .................................................142

REFERENCES .................................................149
FIGURES

Figure 1.1.2(i) The 'two hit mechanism of oncogenesis through tumour suppressor gene inactivation .................. 4
Figure 1.1.2(ii) Mechanisms of uncovering recessive mutations ................... 6
Figure 1.1.4 A genetic model for colorectal tumorigenesis ................... 8
Figure 1.3.3 The positional cloning approach .......................... 17
Figure 1.5.1(i) Genetic map of the BRCA1 region ..................... 23
Figure 1.5.1(ii) A typical breast ovarian cancer family linked to BRCA1 .................................. 24
Figure 1.5.6(i) Direct selection protocol .................................. 34
Figure 1.5.6(ii) Exon trapping .......................................... 37
Figure 1.5.8 Principle of SSCP analysis .............................. 42
Figure 3.1.1 Isolation of linking clones from the RARA gene region .................................................. 74
Figure 3.1.2 Localisation of 2H10 on the somatic cell hybrid mapping panel .................................. 74
Figure 3.1.3(i) Localisation of linking clone 2H10 .................................................. 75
Figure 3.1.3(ii) Construction of a regional restriction map distal to RARA .................................. 76
Figure 3.1.4 Pulsed field gel electrophoresis analysis of IGFBP-4 .......... 78
Figure 3.2.2 Construction of a long-range restriction map around EDH/D17S855 by PFGE analysis .............................. 80
Figure 3.3.1 Construction of a physical map from PFGE data using probes RNU2, 1A1.3B and 12E/Bd, a fragment from the D17S858/D17S859-positive cosmid, G05149 .................................. 82
Figure 3.3.2 Joining long-range restriction maps at EDH/D17S855 and 1A1.3B/RNU2 using an interstitial exon-trap product .......... 83
Figure 3.4.3 Long-range restriction map of the region PPY-GP2B ................. 88
Figure 3.5 Physical map of the BRCA1 region on chromosome 17 .................. 90
Figure 4.1.1(i) Genomic PFGE map of the D17S183 region ...................... 93
Figure 4.1.1(ii) Restriction digestion of cosmid SCG43 with NotI .................................. 93
Figure 4.1.2 Northern analysis of a cDNA clone isolated from a placenta cDNA library using two large NotI fragments from cosmid SCG43 as probes .................................. 93
Figure 4.1.3(i) Localisation of the human upstream binding factor gene: hybridisation to the chromosome 17 somatic cell hybrid mapping panel .................................. 95
Figure 4.1.3(ii) Localisation of the human upstream binding factor gene:
hybridisation to the D17S183 cosmid, SCG43 .................................. 95
Figure 4.2.2 PFGE Southern hybridisation results with the
RNU2 probe ................................................. 97
Figure 4.2.1(i) Map of the U2 RNA repeat unit with its flanking
junction fragments ........................................ 98
Figure 4.2.1(ii) Variation of the U2 RNA tandem array in different
cell lines .................................................... 98
Figure 5.1.1(i) Confirmation of the location of C1198 adjacent to the
RNU2 locus by PFGE analysis .................................. 104
Figure 5.1.1(ii) Ordering of the cosmid pool from YAC 12H4 ........ 104
Figure 5.1.2(i) Southern analysis of cosmids isolated with the
26D6 YAC ..................................................... 106
Figure 5.1.2(ii) Ordering of the cosmid pool from YAC 26D6 ........ 106
Figure 5.2 The gap between the cosmid contigs around
RNU2 and PPY .............................................. 107
Figure 5.3 Hybridisation of probes from each end of cosmid
B09127 to filters containing gridded cosmids from
the flow-sorted chromosome 17 cosmid library .................. 108
Figure 5.3.1 Construction of a cosmid walk from
D17S858/D17S859 to the centromere ................................ 109
Figure 5.3.2 Cosmid walk from RNU2 to the telomere .............. 111
Figure 5.4 Southern hybridisation results indicating overlap
between cosmid G124, from the distal end of the
RNU2-telomere walk, and cosmids C02179 and D12132
from the proximal end of the PPY-centromere walk .......... 113
Figure 5.5 Confirmation of the cosmid walk between RNU2 and
D17S78 by FISH analysis .................................... 114
Figure 6.1(i) CpG island screen of the cosmids from RNU2-
D17S858/859 by restriction enzyme analysis ..................... 117
Figure 6.1(ii) Transcriptional analysis of the cosmid contig
between RNU2 and D17S858/859 ................................ 118
Figure 6.2.1(i) Isolation of HDSPH cDNA clones from a gridded
selected cDNA library ...................................... 119
Figure 6.2.1(ii) cDNA clones corresponding to the HDSPH gene ........ 121
Figure 6.2.2 Nucleotide and deduced amino acid sequence of the dual-
specificity VH1-related VHR phosphatase cDNA ............... 122
Figure 6.3(i) Localisation of DNA probes corresponding to
elements of a human endogenous retrovirus ..................................124

Figure 6.3(ii),(iii) Localisation of ET6B and 8Eg2 gene loci by
Southern analysis of cosmids between RNU2 and
D17S858/859 .................................................................125

Figure 6.3(iv) Localisation of a novel G protein, HAL64 to the
cosmid contig .............................................................126

Figure 6.3(v) Localisation of cDNA clones corresponding to the
HDSPH gene ................................................................127

Figure 6.4 Northern analysis of cDNA clones from the RNU2-
D17S858/D17S859 region ..................................................128

Figure 7.1(i) Searching for cytogenetic rearrangements within the
BRCA1 region by PFGE analysis (i) 855 RF probe ..........133

Figure 7.1(ii) Searching for cytogenetic rearrangements within
the BRCA1 region by PFGE analysis (ii) PPy and
EDH probes .................................................................134

Figure 7.2(i) Mutational analysis by conventional Southern
analysis (i) HDSPH gene ..................................................136

Figure 7.2(ii) Mutational analysis of conventional Southern
analysis (ii) 1A1.3B and IGFBP-4 genes .........................137

Figure 7.3 Genomic SSCP analysis of the VH1-related VHR
phosphatase gene within the DNA of affected members
of breast/ovarian cancer families and early onset
breast tumours ............................................................139
TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.2.1</td>
<td>Genes responsible for hereditary cancer</td>
<td>10</td>
</tr>
<tr>
<td>Table 1.4.2</td>
<td>Breast cancer susceptibility loci</td>
<td>19</td>
</tr>
<tr>
<td>Table 1.5.5</td>
<td>Comparison of genomic cloning systems</td>
<td>31</td>
</tr>
<tr>
<td>Table 2.3.2</td>
<td>PFGE parameters</td>
<td>53</td>
</tr>
<tr>
<td>Table 3</td>
<td>Mapping results of probes in the BRCA1 region on a panel of chromosome 17 CMGTs and translocation hybrids.</td>
<td>71</td>
</tr>
<tr>
<td>Table 3.4.1</td>
<td>Details of single copy probes from the PPY to GP2B region used in PFGE Southern analysis</td>
<td>85</td>
</tr>
<tr>
<td>Table 3.4.2</td>
<td>Results of PFGE Southern hybridisation analysis using probes from the PPY to GP2B region.</td>
<td>86</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Results of YAC analysis in the 1A1.3B to D17S78 region</td>
<td>102</td>
</tr>
<tr>
<td>Table 7.3</td>
<td>PCR-SSCP conditions for analysis of HDSPH</td>
<td>138</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td>adenomatous polyposis coli</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
<td></td>
</tr>
<tr>
<td>AT</td>
<td>ataxia telangiectasia</td>
<td></td>
</tr>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
<td></td>
</tr>
<tr>
<td>BL</td>
<td>Burkitt's lymphoma</td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
<td></td>
</tr>
<tr>
<td>BRCA1</td>
<td>breast cancer 1 gene</td>
<td></td>
</tr>
<tr>
<td>BWS</td>
<td>Beckwith-Wiedemann Syndrome</td>
<td></td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
<td></td>
</tr>
<tr>
<td>CMGT</td>
<td>chromosome mediated gene transfectant</td>
<td></td>
</tr>
<tr>
<td>DCC</td>
<td>deleted in colorectal cancer</td>
<td></td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>dNTPs</td>
<td>deoxyribonucleoside triphosphates</td>
<td></td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
<td></td>
</tr>
<tr>
<td>EDH</td>
<td>oestradiol dehydrogenase</td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>disodium ethylenediaminetetraacetate</td>
<td></td>
</tr>
<tr>
<td>EPB3</td>
<td>erythrocyte surface protein band 3</td>
<td></td>
</tr>
<tr>
<td>ER</td>
<td>oestrogen receptor</td>
<td></td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescence in situ hybridisation</td>
<td></td>
</tr>
<tr>
<td>GAP</td>
<td>GTPase activating protein</td>
<td></td>
</tr>
<tr>
<td>GAS</td>
<td>gastrin</td>
<td></td>
</tr>
<tr>
<td>GP2B</td>
<td>glycoprotein 2B</td>
<td></td>
</tr>
<tr>
<td>HDSPH</td>
<td>human dual specificity phosphatase</td>
<td></td>
</tr>
<tr>
<td>HERV</td>
<td>human endogenous retrovirus</td>
<td></td>
</tr>
<tr>
<td>hUBF</td>
<td>human upstream binding factor</td>
<td></td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
<td></td>
</tr>
<tr>
<td>HNPCC</td>
<td>hereditary non-polyposis colorectal carcinoma</td>
<td></td>
</tr>
<tr>
<td>IGFBP-4</td>
<td>insulin-like growth factor binding protein</td>
<td></td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pair</td>
<td></td>
</tr>
<tr>
<td>KRT10</td>
<td>keratin 10</td>
<td></td>
</tr>
<tr>
<td>LB broth</td>
<td>Luria Bertani broth</td>
<td></td>
</tr>
<tr>
<td>LOH</td>
<td>loss of heterozygosity</td>
<td></td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
<td></td>
</tr>
<tr>
<td>MCC</td>
<td>mutated in colorectal cancer</td>
<td></td>
</tr>
<tr>
<td>MMTV</td>
<td>mouse mammary tumour virus</td>
<td></td>
</tr>
</tbody>
</table>
mRNA  messenger RNA
NCAM  neural cell adhesion molecule
NF1  neurofibromatosis type I
NOR-90  nucleolus organiser region autoantigen-90
PAC  P1 artificial chromosome
PBSA  phosphate buffered saline
PCR  polymerase chain reaction
PFGE  pulsed-field gel electrophoresis
PHB  prohibitin
PML  promyelocytic leukaemia gene
PPY  pancreatic polypeptide
RARA  retinoic acid receptor alpha
RB  retinoblastoma
RFLP  restriction fragment length polymorphism
RNA  ribonucleic acid
RNU2  U2 small nuclear RNA locus
rpm  revolutions per minute
rNTP  ribonucleoside triphosphates
RT-PCR  reverse transcription PCR
SDS  sodium dodecyl sulphate
SSCP  single-stranded conformation polymorphism
TCR  T-cell receptor
THRA1  thyroid hormone receptor
TOP2  topoisomerase 2
VHL  von Hippel Lindau disease
VNTR  variable number of tandem repeats
WT  Wilm's tumour
YAC  yeast artificial chromosome
Work in this thesis is included in the following publications:


Also by K. A. Jones:

ACKNOWLEDGMENTS

I have been very fortunate to work with many talented people who have provided support and invaluable advice during the course of my work at the ICRF. I especially thank Kathy Howe, Donny Black, Audrey Goddard, Julian Borrow, Hans Nicolai, Beatrice Griffiths, Julie Chambers, Melissa Brown, Chun-Fang Xu, David Grimwade and Stuart Whitelaw for their helpful suggestions and advice and for making my time at ICRF so enjoyable.

I am also enormously grateful to Ellen Solomon and Dallas Swallow who, as my PhD supervisors, have provided useful advice which has shaped the direction of my work.

Special to thanks to all my family and friends who have given so much support. Thank you, Bal for all the encouragement and patience from the start.
CHAPTER 1 INTRODUCTION

The work described in this thesis is concerned with the detailed characterisation of a chromosomal region containing a gene responsible for a large percentage of familial breast and ovarian cancer. The work began shortly after the publication of the initial chromosomal assignment of the gene and preliminary work defining the candidate gene region. It commenced with the physical mapping of the gene region (see chapter 3) and progressed to genomic cloning and contig construction (chapter 5) followed by transcriptional analysis of the region and gene isolation (chapter 6) and finally mutational analysis (chapter 7). The direction of the thesis throughout its entire duration was influenced by the findings of other groups in the field. Parallel genetic mapping studies successively narrowed down the region and indicated a small region in which to concentrate gene isolation work. The genomic cloning and transcriptional analysis described here concentrated on this small region and located a number of known and novel genes.

Following the recent publication of the susceptibility gene by another group, a consideration of the findings obtained by this thesis work in the light of the gene discovery has been possible. Discoveries made during the course of this thesis concerning the surrounding chromosomal region may help to understand the mechanisms concerning the susceptibility gene which go awry with the consequence of oncogenesis.

1.1 From cancer predisposition to malignancy: accumulation of multiple defects in tumour suppressor genes and oncogenes

Cell numbers are maintained by a balance between increase in cell number resulting from cell division and withdrawal from the cell cycle due to differentiation or cell death (reviewed in (Hoffman and Liebermann 1994)). The control mechanisms affecting these processes are complex, involving stimulatory and inhibitory signals encoded by proto-oncogenes and tumour suppressor genes respectively (reviewed in (Bishop 1991; Marshall 1991; Goddard and Solomon 1993)). The acquisition of tumorigenicity results from genetic changes within these control genes. They include dominant, gain-of-function mutations within proto-oncogenes resulting in abnormal, positive signals for cell proliferation and recessive loss-of-function mutations within the tumour suppressor genes which interfere with mechanisms restraining cell multiplication. Mutations within control genes may also have a "dominant-negative" effect, whereby an altered protein is produced which competes with its wild-type counterpart and prevents its activity.
1.1.1 The roles of oncogenes in malignancy

Neoplasia arising from the activation of proto-oncogenes, which are important in the maintenance of cell numbers in normal tissue, to oncogenes whose expression cause cell numbers to rise with the consequence of malignancy, was first suggested by observations of the transforming ability of retroviruses (reviewed in (Bishop 1983)). Since then, analyses of the consistent chromosomal rearrangements observed in leukaemias and solid tumours have identified proto-oncogenes at, or spanning, chromosomal breakpoints. The consequences of such chromosomal aberrations are either the activation of the proto-oncogene through its juxtaposition to T-cell receptor (TCR) or immunoglobulin gene; or the creation of a fusion gene encoding a chimaeric protein (reviewed in (Rabbitts 1994)). The activation of tumorigenic activities by the CMYC gene translocation in Burkitt's lymphoma (Taub et al. 1982), the expression of a bcr-abl chimeric gene product in patients with chronic myelogenous leukaemia (Kurzrock et al. 1988) and the activation of HRAS and KRAS oncogenes in other tumours (reviewed in (Bishop 1991)) are examples of the involvement of such transforming oncogenes in human cancer.

1.1.2 The roles of tumour suppressor genes in malignancy

The importance of loss of normal gene function in the expression of malignancy was first demonstrated by experiments showing the dominant suppression of the tumorigenic phenotype of a malignant cell upon fusion with a normal cell (Harris et al. 1969; Stanbridge 1976). Reversion back to malignancy was often observed coincidently with chromosomal loss (Klinger 1982). The hybrid suppression of tumorigenicity was explained by the contribution by the normal parental cell of 'suppressor' genes, absent or inactive in the malignant parent, which were able to regulate some aspects of the transformed phenotype.

A clear demonstration of the importance of tumour suppressor gene inactivation in humans was the suppression of tumorigenicity of HeLa cervical carcinoma cells upon fusion with normal human diploid fibroblasts (Anderson and Stanbridge 1993). The presence of multiple tumour suppressor loci was indicated by the observation that hybrids formed by the fusion of two malignant cell lines derived from tissues of different histological types did not exhibit a tumorigenic phenotype (Weissman and Stanbridge 1983; Goddard and Solomon 1993). This implied that the malignant parent cell lines possessed different genetic aberrations and therefore could complement each other. Multiple tumour suppressor loci was also suggested by correlations of tumour suppression with specific chromosomes (reviewed in (Goddard and Solomon 1993)).
1.1.3 The recessive model of cancer inheritance

The importance of loss of tumour suppressor gene function in tumours in general could be extended to the rarer inherited cancers following the ideas of Knudson in his study of the rare intraocular tumour of early childhood, retinoblastoma (RB) (Knudson 1971). The different ages of onset of the disease in the inherited and sporadic forms of the disease suggested the requirement for two rate-limiting mutational events for oncogenesis (Fig 1.1.3(i)) (reviewed in (Malkin and Friend 1992)). In individuals with the inherited form of the disease, predisposition is due to a heterozygous defect in the retinoblastoma tumour suppressor gene in every cell, as a result of a germline mutation. For tumour initiation, only one allele of the gene remains to be inactivated to achieve full deficiency for the gene. Conversely, sporadic RB requires inactivation of both alleles to occur in the same somatic cell. This 'two hit' model also fits the epidemiological data for other cancers observed in both familial and sporadic forms, including renal cell carcinoma, neurofibroma, breast cancer and colon cancer.

Evidence in support of the two hit model has been provided by comparisons of the constitutional genotypes of affected and unaffected parents to the constitutional and tumour genotypes of their affected children (Cavenee et al. 1985). They indicated that the chromosome carrying the functional retinoblastoma gene was lost from the tumours of two hereditary retinoblastoma cases. The remaining chromosome contained the defective allele which had been inherited from the affected parents. Hence the loss of a chromosome commonly observed in sporadic disease and observed here in inherited disease constitutes the second event in the two hit model and serves to remove the functional RBI gene from the cell with the consequence of oncogenesis.

However, two problems with this model are apparent. Firstly, inheritance of cancer predisposition follows a dominant pattern even though it is transmitted by recessive lesions. Secondly, incidences of non-hereditary cancer due to the same gene do occur despite the requirement for the seemingly unlikely event of two somatic mutations within the same gene occurring in the same cell. These dilemmas may be resolved if some suppressor gene mutations affected growth even in the presence of a wild-type allele by dominant-negative and dosage effects (discussed in (Fearon and Vogelstein 1990)) and if the likelihood of the second mutation event was extremely high. If cells heterozygous for defective tumour suppressor gene alleles had a slight growth advantage, it would enable a cell carrying a single gene defect to undergo clonal expansion to amass enough cells to allow a reasonable probability for the second inactivation to occur. Such clonal
HEREDITARY

Germline mutation 'FIRST HIT'

somatic mutation 'SECOND HIT'

multiple tumours
bilaterality
early onset

NON-HEREDITARY

somatic mutation 'FIRST HIT'

somatic mutation 'SECOND HIT'

single tumours
unilateral tumours
later onset

Fig 1.1.3(i) The 'two hit' mechanism of oncogenesis through tumour suppressor gene inactivation. Loss of function of these genes may result in disturbances in the control of cell proliferation and oncogenesis. In the case of the hereditary form, one inactivated allele of the tumour suppressor gene is present in every cell as a result of an inherited germline mutation. This represents the 'first hit' in the oncogenesis mechanism. A somatic mutation occurring at some stage in lifetime provides the 'second hit'. In non-hereditary cases both hits are somatic mutations, occurring in the same somatic cell. (Adapted from Malkin and Friend 1992)
expansion is not required in hereditary disease, due to the presence of a single defect in every somatic cell which explains the increased probability of disease and the early onset and multiplicity often seen in these cases.

The second mutational event in the two hit model may uncover the constitutional recessive mutation by a variety of mechanisms usually involving a loss of genetic material (Fig 1.1.3(ii)) (Cavenee et al. 1983). Such events can be detected by the loss of heterozygosity (LOH) for markers flanking the disease susceptibility locus. Regarded as the 'hallmark of tumour suppressor loci', LOH is identified when a polymorphic marker which detects two variants in constitutional cells only identifies a single variant in the tumour, indicating a reduction from heterozygosity to homozygosity or hemizygosity. Studies of LOH in tumours has confirmed earlier suggestions of multiple tumour suppressor loci (reviewed in (Solomon et al. 1991; Goddard and Solomon 1993; Yokota and Sugimura 1993)).

1.1.4 The multistep nature of cancer

The majority of human tumours pass through three main stages during their development which reflect the multistep nature of the carcinogenesis process: premalignant lesions, primary tumours and metastases (Yokota and Sugimura 1993). Exceptions include breast tumours which on the whole are not associated with premalignant lesions, and some tumours of differentiated tissue such as the brain which remain in situ and fail to metastasise.

When premalignant lesions do occur, they may take the form of dysplasia, hyperplasia, leukoplakia and adenoma, consisting of clonally expanding cells which have acquired selective growth advantage as a result of genetic alteration. Such cells may be less responsive to negative growth regulators and cell differentiation inducers. These cells may convert to malignant cells by additional genetic alterations to form primary tumours. Their constituent cells have a further reduced responsiveness to negative growth regulators and continue to expand by pushing surrounding tissue away. During expansion of the converted cells in the primary site, new clones with more malignant phenotypes may appear through further accumulation of genetic alterations in some of the converted cells. These new clones can be increasingly invasive and autonomous, becoming highly selected and metastatic. Evidence for the development of tumours through clonal expansion is seen in colorectal adenomas which appear to have a monoclonal composition as opposed to normal colonic epithelium which is polyclonal (Fearon et al. 1987).
**Fig 1.1.3(ii) Mechanisms of uncovering recessive mutations.**
The second somatic event in the two hit model can involve mitotic nondisjunction with loss of the wild-type chromosome resulting in hemizygosity at all loci on the chromosome; mitotic nondisjunction with duplication of the mutant chromosome resulting in homozygosity at all loci on the chromosome; or mitotic recombination, resulting in heterozygosity at some loci of the chromosome but homozygosity throughout the rest, including the disease susceptibility locus. Tumours not exhibiting LOH have probably suffered more localised lesions such as gene conversion, point mutations small deletions or insertions.
(Adapted from Cavenee et al 1983)
The genetic alterations underlying these changes occur in the many tumour suppressor genes and proto-oncogenes. The most common oncogene involved in solid tumours is HRAS, found in 10-15% (reviewed in (Bishop 1991)). The most frequently mutated tumour suppressor gene is TP53, mutated in roughly 50% of all tumours (reviewed in (Harris 1993)). Such mutations result in circumvention of the apoptotic machinery, promotion of cell division and proliferation, loss of cell differentiation pathways and disruption of cell-cell communication and interaction. The importance of multiple tumour suppressor gene mutations in such tumours is demonstrated by the large numbers of LOH sites within the same tumour at known and putative new tumour suppressor gene loci (reviewed (Goddard and Solomon 1993)). The number of sites of LOH and other chromosomal aberrations appear to increase in proportion to tumour progression, as demonstrated by molecular studies of colorectal carcinoma (Fig 1.1.4) (Fearon et al. 1987; Vogelstein et al. 1988; Fearon and Vogelstein 1990). This model for colorectal tumour development suggests that the accumulation rather than the order of genetic alterations is important in progression. Recent mutation analysis of the tumours which developed in hereditary non-polyposis colorectal cancer (HNPCC) patients agree with this model (Lazar et al. 1994). In these patients, inactivation of the hMSH2 gene is a source of a progressive accumulation of mutations in critical genes during colorectal tumorigenesis. Mutational analysis of two genes, APC and TP53 in the tumours of these patients demonstrated that whilst early non-malignant polyps contained single mutations in either APC or TP53 genes, malignant tumours from the same patients displayed multiple mutations within both these genes.

The number of genetic events associated with cancer appears to be a function of the kind of target tissue that is transformed (Knudson 1992). Embryonal tumours and some leukaemias and lymphomas require only a few events. For example, the development of retinoblastoma only requires abnormalities of both copies of the RBI gene to give rise to both hereditary and sporadic types of the disease(Friend et al. 1986). On the other hand carcinomas, which occur in epithelial tissues require in the region of eight events, producing functional mutations in four or five genes.

Tumours affecting bone, muscle and breast; tissues which undergo growth during adolescence or adulthood are intermediate in complexity. The following is a list of the common events which play a role in breast cancer progression. Multiple oncogenes and tumour suppressor gene mutations are involved in the process:
Fig1.1.4. A genetic model for colorectal tumorigenesis
(adapted from Vogelstein et al 1988)
• amplification of chromosome 11p including INT2, HST1 and PRAD1 loci in 15-20% of breast carcinomas (Lidereau et al. 1988; Theillet et al. 1989; Theillet et al. 1990; Solomon et al. 1991).

• amplification of ERBB2 (HER2/NEU), a gene related to epidermal growth factor receptor gene, occurring in more advanced stages of adeno-carcinoma of the breast and ovary (Liberman et al. 1985; Slamon et al. 1987; Slamon et al. 1989).

• loss of tumour suppressor activity of p53 (Hartwell 1992), maspin (Zou et al. 1994) and prohibitin (Sato et al. 1992).

• loss of tumour suppressor gene expression throughout the genome (Sato et al. 1990) and at multiple loci on chromosome 17 (Lindblom et al. 1993; Kirchweger et al. 1994).

1.2 Cancer families and the genes responsible

Although individuals belonging to families where heredity plays some role may only account for 5-10% of all people with cancer, the study of the of the genes responsible for their predisposition is important for several reasons. The same genes are often, if not always, involved in sporadic cancers and so the isolation of these susceptibility genes will identify and subsequently help elucidate genetic mechanisms underlying both inherited and sporadic cancer development. Clinically, such discoveries may allow premorbid predictive testing, screening of populations, genetic counselling and prevention, including removal of premalignant lesions. In the long term, a better understanding of the genetic mechanisms that have been disrupted may present ways for correction.

1.2.1 The variety of family cancer genes

Studying the functions of cloned cancer predisposing genes has identified some of the genetic mechanisms central to tumorigenesis which may also involve proteins encoded by cancer predisposing genes yet to be isolated. The genes responsible for the inherited predisposition to cancer so far isolated encode a variety of proteins with a diverse array of functions from components of the cytoskeleton such as NF2 (Rouleau et al. 1993; Trofatter et al. 1993) and probably APC (Groden et al. 1991) to transcription factors such as WT-1, TP53 and RB1 (reviewed in (Goddard and Solomon 1993)) (see Table 1.2.1).
<table>
<thead>
<tr>
<th>GENE (chromosome)</th>
<th>CANCER TYPES (arising from germline mutation)</th>
<th>PRODUCT LOCATION</th>
<th>POSSIBLE PRODUCT FUNCTION</th>
<th>HEREDITARY SYNDROME</th>
<th>REFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB1 (13p14.2)</td>
<td>retinoblastoma; small numbers of osteogenic sarcomas, soft tissue sarcomas, melanomas, brain tumours, epithelial tumours and tumours of the peripheral nervous system</td>
<td>NUCLEUS expression in all tissues</td>
<td>transcription factor; involved in regulation of cell cycle and transcription</td>
<td>Retinoblastoma</td>
<td>Friend et al 1986; Sopta et al 1992; Knudson 1993; Goddard and Solomon 1993;</td>
</tr>
<tr>
<td>NF1 (17q11.2)</td>
<td>neurofibromas, tumours of the central nervous system and neural crest origin</td>
<td>CYTOPLASM expression in all tissues</td>
<td>GTPase activator protein; inhibitor of ras-stimulated cellular mitosis</td>
<td>Neurofibromatosis type I</td>
<td>Wallace et al 1990; Cawthon et al 1990; Buchberg et al 1990; Xu et al 1990; Eng and Ponder 1993</td>
</tr>
<tr>
<td>NF2 (22q12)</td>
<td>schwannomas, meningiomas and other brain tumours</td>
<td>INNER MEMBRANE expression in most tissues</td>
<td>cytoskeleton-associated protein; communication between cell membrane and cytoskeleton</td>
<td>Neurofibromatosis type II</td>
<td>Trofatter et al 1993; Rouleau et al 1993; Knudson 1992; Eng and Ponder 1993</td>
</tr>
</tbody>
</table>

Table 1.2.1 Genes responsible for hereditary cancer
<table>
<thead>
<tr>
<th>Gene</th>
<th>Tumour Type</th>
<th>Expression</th>
<th>Known Similarities to Proteins</th>
<th>Disease/Condition</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC (5q21)</td>
<td>colon carcinoma</td>
<td>CYTOPLASM</td>
<td>unknown - strongest similarities to intermediate filament structural proteins such as the myosins and keratins</td>
<td>Familial adenomatous polyposis; Gardner's syndrome</td>
<td>Groden et al 1991; Kinzler et al 1991; Nishisho et al 1991; Goddard and Solomon 1993</td>
</tr>
<tr>
<td>RET (10q11.2)</td>
<td>medullary thyroid carcinoma</td>
<td>MEMBRANE</td>
<td>receptor tyrosine kinase</td>
<td>Multiple endocrine neoplasia type 2</td>
<td>Mulligan et al 1993; Schuffenecker et al 1994; Eng and Ponder 1993</td>
</tr>
<tr>
<td>WT-1 (11p13)</td>
<td>nephroblastoma</td>
<td>NUCLEUS</td>
<td>transcription factor; regulation of kidney and gonadal development</td>
<td>Wilm's tumour</td>
<td>Gessler et al 1990; Call et al 1990; Coppes et al 1993; Rauscher et al 1993</td>
</tr>
<tr>
<td>VHL (3p25-26)</td>
<td>renal cell carcinoma, haemangioblastomas of the central nervous system and retina, pheochromocytoma</td>
<td>MEMBRANE</td>
<td>unknown</td>
<td>von Hippel-Lindau disease</td>
<td>Latif et al 1993</td>
</tr>
</tbody>
</table>

Table 1.2.1 Genes responsible for hereditary cancer
The majority of the inherited cancer predisposing genes are of the tumour suppressor type. In normal tissue they function to restrain cell proliferation in a variety of ways. The products of the retinoblastoma gene, \textit{RB1} (Friend \textit{et al.} 1986) and the putative familial melanoma gene \textit{CDKN2} (Kamb \textit{et al.} 1994) prevent passage through G1 of the cell cycle by inhibiting the oncogenic activities of other cell cycle components. The \textit{RB1} product inhibits E2F and myc transcription factor activity (reviewed in (Goddard and Solomon 1993)) whereas the \textit{CDKN2} product blocks cyclin dependent kinase 4 action (Serrano \textit{et al.} 1993). The product of the \textit{p53} gene involved in the Li-Fraumeni syndrome appears to have a variety of effects, predominantly due to its ability to act as a transcription factor for a number of genes (reviewed in (Zambetti and Levine 1993)). It appears to halt the cell cycle at the G1-S checkpoint in response to DNA damage and facilitates DNA repair through the regulation of p21 and Gadd45 genes (Kastan \textit{et al.} 1992; Michieli \textit{et al.} 1994). In addition, \textit{p53} may be more directly involved in DNA damage recognition and repair through its interaction with the excision repair protein ERCC3 (Schaeffer \textit{et al.} 1993) and it can also trigger apoptosis (Yonish-Rouach \textit{et al.} 1991; Clarke \textit{et al.} 1993). Mutations in the \textit{p53} gene result in the failure of \textit{p53} to bind DNA and promote transcription (Bargonetti \textit{et al.} 1991; Kern \textit{et al.} 1991; Bargonetti \textit{et al.} 1992), whilst the effects of some oncogene products appear to be due to their ability to inhibit \textit{p53} action (Mietz \textit{et al.} 1992; Momand \textit{et al.} 1992; Yew and Berk 1992).

Because \textit{p53} has so many functions, mutations in it have devastating effects on the cellular machinery. This may explain why mutations in the \textit{p53} gene contribute to nearly 50% of all human cancers (Marx 1994).

The gene causing the cancer-prone syndrome ataxia telangiectasia (AT), is thought to function upstream of the \textit{p53} gene (Kastan 1993). Like the Li-Fraumeni syndrome, cancer predisposition in AT is due to the inherited genomic instability which results from a defect in the G1-S checkpoint (reviewed (Hartwell 1992)). Inherited genomic instability also appears to be a factor in hereditary non-polyposis colorectal cancer (HNPCC) (reviewed in (Jiricny 1994)). Four genes involved in DNA mismatch repair, \textit{hMSH2}, \textit{hMLH1}, \textit{hPMS1} and \textit{hPMS2} have been found to be mutated in the germline of affected members in some HNPCC families (Fishel \textit{et al.} 1993; Leach \textit{et al.} 1993; Bronner \textit{et al.} 1994; Nicolaides \textit{et al.} 1994; Papadopoulos \textit{et al.} 1994). Their failure to function results in cells with increased numbers of uncorrected random mutations.

Most of the inherited cancer predisposing genes so far identified are also involved in non-inherited forms of the disease and require two mutational events to trigger oncogenesis, in accordance with the Knudson two hit model. The recently identified gene which causes von Hippel-Lindau (VHL) disease, characterised by early onset and multifocal
clear cell renal cell carcinoma, and at least one of the DNA repair genes responsible for HNPCC, *MLH1*, can be added to this list of conventional tumor suppressor genes. Mutations within the VHL gene and allele loss at its locus on chromosome 3p25 appear to be critical events in non-familial renal cell carcinoma, compatible with the two hit model (Foster *et al.* 1994). Similarly, allele loss involving the wild type *MLH1* gene is a feature of HNPCC (Hemminki *et al.* 1994).

In other cancer genes, such as *TP53*, *AT* and possibly *hMSH2* however, a single defect within the genes is sufficient to trigger oncogenesis as a result of decreased dosage of their wild-type gene products in the cells (discussed (Hartwell 1992; Jiricny 1994)).

A single hit event in the triggering of oncogenesis is also applicable for the RET oncogene which gives rise to multiple endocrine neoplasia type 2A (MEN2A) (Mulligan *et al.* 1993) and the insulin-like growth factor-2 gene *IGF2*, implicated as the cause of Beckwith-Wiedemann syndrome (BWS) (Weksberg *et al.* 1993). In these cancer syndromes, a model of tumorigenesis distinct from Knudson's model is likely. Evidence that heterozygosity for the disease allele is sufficient for oncogenesis is the observation of hyperplasia in neuroendocrine organs in patients carrying MEN2A mutations and the lack of LOH at the MEN2A locus in the tumours (Eng and Ponder 1993; Mulligan *et al.* 1993). Similarly, in the BWS syndrome, which predisposes to Wilm's tumour, hepatoblastoma, adrenocortical carcinoma and rhabdomyosarcoma, non-tumorous tissue overgrowth, such as large birth weight and large tongue is a clinical feature of the heterozygous state.

### 1.3 Methods of isolating cancer genes

#### 1.3.1 Functional methods of gene isolation

A variety of techniques have been used to isolate cancer genes (Collins 1992). Probably the most straightforward technique relies on pre-existing knowledge about the basic defect resulting from the gene mutation. This may consist of, for example, partial amino acid sequence of the normal protein product. Alternatively, antibodies against the protein product could be used in the cloning strategy. In the isolation of the human non-polyposis colorectal cancer gene *hMSH2*, observations of instability of repeat elements throughout the genome indicated the involvement of a mismatch repair gene. A previously isolated yeast nuclear mismatch repair gene, *MSH2* was used to isolate homologous human cDNA clones, one of which was found to be from the culprit locus (Fishel *et al.* 1993; Jiricny 1994).
1.3.2 Subtractive hybridisation

In the absence of such biochemical aids to the identification of cancer genes, methods have been employed which exploit the differences between the genomes of tumour cells and their normal counterparts. The techniques result in the isolation of genes expressed in one cell population but not the other. Subtractive hybridisation compares the mRNA expression of genes from normal and tumour-derived cells which have been grown at similar growth rates, minimising mRNA level differences resulting from disparate growth conditions. Candidate tumour suppressor genes whose expression is lost in human mammary epithelial cells have been successfully cloned in this manner (Lee et al. 1991). This method has also successfully isolated a gene for a protease inhibitor known as maspin, whose loss of expression in human mammary epithelial tumour cells may be important in tumour progression (Zou et al. 1994). Refinements of this methodology have enabled identification of genes which show more subtle differences in expression between normal and tumour cells (Liang et al. 1992; Lisitsyn et al. 1993; Sager et al. 1993).

1.3.3 Positional cloning

Whilst subtraction hybridisation methodology can pinpoint general differences between normal and tumour cells, isolating genes responsible for specific defects in the absence of biochemical tools requires the use of positional cloning. This is essentially the isolation of a disease gene starting from knowledge of its genetic or physical location. In the simplest of cases, this might involve the isolation of genes disrupted by a reciprocal translocation. The underlying cause of oncogenesis can be revealed by identifying the genes involved in these disruptions. Gene disruptions resulting in sporadic haematological cancers have been identified following the discovery of such translocations. For example, the consistent t(15;17) translocation observed in the promyelocytes of acute promyelocytic leukaemia patients disrupts the retinoic acid receptor alpha (RARA) and PML genes resulting in the production of an oncogenic chimaeric fusion product (Borrow et al. 1990).

In inherited cancers, the discovery of chromosomal aberrations within the germline DNA has also aided the discovery of the underlying causes of oncogenesis. The identification of an interstitial deletion of chromosome 5q in the germline of an individual with colon cancer provided an early clue to the location of the familial adenomatous polyposis coli gene for example (Herrera et al. 1986). In the hunt for the neurofibromatosis gene,
translocations were identified in two patients with NF1 (Collins et al. 1989). In both patients, one breakpoint involved 17q11.2, later found to be the site of the NF1 gene.

Without such clues to the chromosomal location of the disease gene, a search of the whole genome becomes necessary. This begins with collection of pedigrees in which the gene responsible is segregating. Studying the inheritance of multiple polymorphic markers and determining positions of recombinations in these families can narrow down the region in which the disease susceptibility gene resides. Discoveries of meiotic recombinational events between the disease and various markers allows exclusion of portions of the candidate gene region, but the precision of this approach depends on the following factors.

- **Genetic heterogeneity.** A vital part of the genetic mapping stage of the positional cloning strategy is the collection of genetically homogeneous groups of families. This becomes difficult if disease pedigrees which are due to different genes cannot be distinguished from one another. Breast cancer families, which can be due to defects in many genes (Table 1.4.2) cannot be distinguished from one another on the basis of age at disease onset, for example (Margaritte et al. 1992). Similarly, the onset of Wilm's tumour can be due to any one of three loci (Goddard and Solomon 1993), whilst at least two loci at 9p and 1p have been found to cause familial malignant melanoma (Wainwright 1994). These problems may be minimised by concentrating the genetic analysis work on a small number of large families rather than on a collection of many small families which may include families which are due to different susceptibility genes.

- **Sporadic incidence.** This is a very large problem in the case of common diseases such as breast cancer and malignant melanoma. The majority of cases are sporadic and do not possess germline mutations within the cancer gene in question. The presence of such individuals within families under study can confuse the definition of the genetic region containing the disease gene.

- **Incomplete penetrance.** It is difficult to establish unequivocally the link between inheritance of a susceptibility allele and onset of a disease if some carriers escape the diseased phenotype, as for example in the case of the CDKN2 gene and familial malignant melanoma (Kamb et al. 1994).
• Number of informative meioses. This limits the amount of detail in the genetic map of the disease gene region. In rare syndromes, the paucity of affected individuals will make narrowing down the genetic region difficult. Subsequent physical mapping and cloning work may not be feasible if the genetic region defined by these few family members is too large. In such cases, it may be necessary to limit the search for candidate genes to regular checks of the human genome mapping database until a suitable candidate gene mapping to the region presents itself.

Linkage mapping has been revolutionised over the last few years with an explosion of readily available highly polymorphic DNA markers. They include restriction fragment length polymorphisms (RFLPs), which reveal DNA sequence differences by a loss or gain of single restriction enzyme recognition and cleavage sites, variable number tandem repeats (VNTRs), short repeat elements which show variation in copy number (Nakamura et al. 1987); and microsatellites, short di-, tri-, and tetranucleotide repeats which occur very frequently in mammalian DNA. These repeat elements are believed to have arisen as a consequence of DNA slippage and are highly polymorphic due to length heterogeneity (reviewed (Hearne et al. 1992)). They have become the preferred type of polymorphic marker because they can be analysed with the aid of the polymerase chain reaction (PCR). Primers can be designed from the regions flanking the microsatellite repeat and the genotype for these markers can be obtained by separation of the PCR products on acrylamide gels. Their main advantages are their high degree of polymorphism and the low quantity and quality of DNA required in the procedure. This has meant that archive material which may be paraffin-embedded and degraded, can be used in the analysis.

Coupling linkage analysis with other gene-mapping techniques, such as the search for disease-associated cytogenetic abnormalities and loss of constitutional DNA heterozygosity in tumours, can help narrow the search to a small region. The availability of any gross rearrangements in the gene, even if present in only a small subset of tumours, accelerates the process of gene identification. Most of the successful positional cloning projects so far have relied upon the discovery of relatively large rearrangements.

The process of cancer gene isolation from the initial genetic mapping stage usually requires a number of steps, summarised in Fig 1.3.3 (reviewed (Wicking and Williamson 1991; Collins 1992)). They involve detailed physical characterisation of the chromosomal region in which the gene is thought to reside, including details of the distances between the closest and most reliable genetic markers delimiting the region.
STAGE OF STRATEGY

1. LINKAGE MAPPING

RESOURCES

- genetically homogeneous groups of families

2. REFINING THE GENETIC MAP

RESOURCES

- large families with informative meioses, randomly distributed polymorphic markers throughout the region

3. PHYSICAL MAPPING

RESOURCES

- somatic cell hybrid mapping panel, pulsed-field electrophoresis mapping, fluorescence in situ hybridisation

4. GENOMIC CLONING

RESOURCES

- YAC, PAC, P1, cosmid libraries

5. GENE ISOLATION

RESOURCES

- cDNA libraries, zoo blots, CpG island clones, exon-trapping libraries, solution hybrid libraries, access to genbank, EMBL genome databases

6. MUTATION ANALYSIS

RESOURCES

- automated sequencing apparatus, sequencing analysis software, SSCP analysis, chemical mismatch cleavage

Fig 1.3.3 The positional cloning approach
This aids the saturation of the chromosomal region with genomic clones which can in turn be used as a resource in the isolation of genes residing in the region. The effort then reduces to a search through these transcripts for causative mutations in individuals with the disease.

In some lucky instances, human disease genes have been identified without the requirement for exhaustive cloning efforts. In such cases, the initial chromosomal assignment of the disease region has shown it to reside in the proximity of an obvious known candidate gene. In this way, the genes for rhodopsin, peripherin, cardiac myosin heavy chain and p53 were found to be responsible for two forms of retinitis pigmentosa, familial hypertrophic cardiomyopathy, and the Li-Fraumeni cancer syndrome respectively (Collins 1992).

1.4 Breast cancer predisposition and the genes responsible

1.4.1 Evidence for a genetic component

Several lines of evidence suggest genetic factors play a role in breast cancer risk. Breast cancer is an extremely common, predominantly sporadic disease which affects approximately one in eight American women during their lifetime (American Cancer Society, Cancer Facts and Figures 1994, Atlanta GA p13). Roughly 5% of breast cancer cases have a family history of the disease, a factor which appears to increase the risk of breast cancer development by two to threefold (Anderson 1974; Schwartz et al. 1985; Ottman et al. 1986; Lynch et al. 1988; Claus et al. 1990; Claus et al. 1991). Numerous epidemiological studies have investigated the transmission in such familial cases and have shown that at least a proportion of breast cancer cases can be explained by inherited mutations in one or more autosomal dominant genes (Newman et al. 1988; Claus et al. 1991). The contribution these genes make to total breast cancer cases is age-dependent; more that a third of the cases diagnosed before the age of 30 years are estimated to be due to inheritance of a susceptibility allele. This contribution is reduced to about 1% of cases diagnosed after the age of 80 years (Claus et al. 1990; Claus et al. 1991).

1.4.2 Breast cancer susceptibility genes

Despite the extensive heterogeneity of the disease (Narod and Amos 1990; Skolnick and Cannon-Albright 1992; Sobol et al. 1992) and the high incidence of sporadic disease, several breast cancer susceptibility loci have been identified (Table 1.4.2). Constitutional mutations of one gene, BRCA1, which maps to chromosome 17q21 (Hall et al. 1990)
<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosomal Region</th>
<th>percentage of familial breast cancer due to this gene</th>
<th>Types of cancer arising from germline mutations</th>
<th>References and notes</th>
<th>Gene product function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>17q21</td>
<td>&gt;80% of breast/ovarian cancer families 45% breast cancer only families</td>
<td>female breast, ovarian, prostate</td>
<td>Recently isolated (Miki et al 1994). LOH and early onset/bilaterality in inherited breast cancer suggestive of a tumour suppressor gene</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>BRCA2</td>
<td>13q12-13</td>
<td>45% breast cancer only families</td>
<td>female and rare cases of male breast cancer</td>
<td>Recently mapped (Wooster et al 1994). LOH also noted, but may be due to the nearby RB1 gene</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>AR</td>
<td>rare incidence male breast cancer</td>
<td>male breast cancer</td>
<td>Low androgen/excess oestrogen leads to breast cancer. Mutations in AR disrupt androgen-binding and DNA binding domains.</td>
<td></td>
<td>steroid nuclear receptor involved in male sex differentiation</td>
</tr>
<tr>
<td>ER</td>
<td>6q23</td>
<td>few % families with late onset disease (&gt;50yrs)</td>
<td>late onset breast cancer</td>
<td>1/11 families studied with late onset disease show linkage to oestrogen receptor locus</td>
<td>binds and transfers oestrogens to the nucleus</td>
</tr>
<tr>
<td>Cowden disease gene</td>
<td>rare cases of early onset breast cancer</td>
<td>thyroid tumours, breast cancer</td>
<td>Brownstein et al 1978; Starink et al 1986; reviewed in Anderson 1991</td>
<td></td>
<td>UNKNOWN</td>
</tr>
</tbody>
</table>

Table 1.4.2. Breast cancer susceptibility loci
predisposes heterozygous carriers to carcinomas of the breast and ovary and is thought to account for approximately 45% of families with significantly high breast cancer incidence and at least 80% of families with increased incidence of both early-onset breast cancer and ovarian cancer (Easton et al. 1993). A second locus \textit{BRCA2}, recently mapped to a six-centimorgan interval on chromosome 13q12-13 (Wooster et al. 1994) may account for a proportion of early onset breast cancer roughly equal to that resulting from \textit{BRCA1}. Previous studies have indicated \textit{BRCA1} mutations are not responsible for inherited male breast cancer (Stratton et al. 1994), whilst male breast cancer cases were present in a few families exhibiting linkage to \textit{BRCA2}. Initial observations also suggested \textit{BRCA1} played a more important role in ovarian cancer than \textit{BRCA2} (Wooster et al. 1994). However, more recent epidemiological data suggest the roles of \textit{BRCA1} and \textit{BRCA2} in ovarian cancer and male breast cancer are less clear cut (T. Bishop, pers comm).

Several breast cancer families have been identified which fail to show evidence of linkage to either \textit{BRCA1} or \textit{BRCA2} loci (Sobol 1994). These families may be due to inherited germline mutations in the other genes known to cause breast cancer predisposition or perhaps to defects in one or more unmapped gene or genes. Some of these breast cancer cases are members of Li-Fraumeni cancer families who have inherited defects within the \textit{TP53} gene on chromosome 17p (Malkin et al. 1990). Others have been associated with heterozygosity for the defective form of the gene predisposing to ataxia telangiectasia (AT) (Swift et al. 1991). In rare cases of premenopausal breast cancer an inherited defect in the Cowden disease gene is thought to be responsible. This disease is characterised by skin and oral cavity lesions that occur in association with benign and malignant thyroid tumours. More than 90% of the female carriers for this disease develop breast tumours and half of the carriers develop early onset breast cancer, often with bilateral disease (Brownstein et al. 1978; Starink et al. 1986); reviewed in (Anderson 1992)). Two breast cancer families were found which appeared to have a predisposition to post-menopausal disease which is coinherit with haplotypes from the oestrogen receptor (\textit{ER}) locus on chromosome 6 (Zuppan 1991). Rare instances of male only breast cancer which occurred in association with genitourinary abnormalities have been found to be due to a germline mutation in the androgen receptor gene (\textit{AR}) (Lobaccaro et al. 1993).

1.4.3. The breast cancer susceptibility locus, \textit{BRCA1}

The mapping of a susceptibility gene, \textit{BRCA1} to chromosome 17q12-24 (Hall et al. 1990) confirmed the predictions that one or more autosomal dominant susceptibility genes was
responsible for the occasional clustering of breast cancer in families (Newman et al. 1988; Claus et al. 1991). Inspection of the apparently linked families suggest that the penetrance conferred by \textit{BRCA1} is close to 100\% (Easton et al. 1993). Female carriers have a 60\% risk of breast cancer by the age of fifty and a 90\% lifetime risk (Ford et al. 1994).

The risk of ovarian cancer is also increased in carriers of defective \textit{BRCA1} alleles. Hints of a gene predisposing to both breast and ovarian cancer came from the observation that breast cancer was increased in ovarian cancer relatives and \textit{vice versa} (Schildkraut et al. 1988; Easton et al. 1993). This was confirmed by the discovery that nearly all families with both breast and ovarian cancer cases were linked to \textit{BRCA1} (Easton et al. 1993). In some families, possession of a defective \textit{BRCA1} allele confers an equal risk to breast and ovarian cancer, whilst in others, the risk of ovarian cancer is much lower than the risk of breast cancer. This suggests that either \textit{BRCA1} is responsible for both breast and ovarian cancer but allelic heterogeneity results in different phenotypes; or ovarian cancer is due to a very closely linked gene to \textit{BRCA1} (Ford et al. 1994).

In addition to conferring risks to both breast and ovarian cancer, defective \textit{BRCA1} alleles appear to increase the risk of colon cancer by four-fold in both male and female carriers. Male carriers of defective \textit{BRCA1} alleles are at a three-fold increased risk of prostate cancer (Ford et al. 1994).

Such observations have raised many questions concerning the normal functions of \textit{BRCA1} and how they are disrupted with the result of tumorigenesis. A tumour suppressor function is suggested by epidemiological studies of breast cancer which appear to be consistent with the Knudson model. Familial cases are increasingly represented among patients with premenopausal, or early onset, disease (Schwartz et al. 1985; Lynch et al. 1988; Claus et al. 1990). In addition, incidences of bilaterality appear to be more common amongst familial cases (Ottman et al. 1983; Bernstein et al. 1992). Observations of allele loss of the \textit{BRCA1} region in many sporadic and familial breast and ovarian cancers have supported theories of a tumour suppressor role for \textit{BRCA1} (Foulkes et al. 1991; Eccles et al. 1992; Futreal et al. 1992; Lindblom et al. 1993; Saito et al. 1993; Cropp et al. 1994; Lalle et al. 1994; Mori et al. 1994). Haplotype analyses of the breast and ovarian tumours from \textit{BRCA1}-linked families indicate the allele losses observed involve the wild type allele whilst the inherited defective allele is retained (Smith et al. 1992; Lalle et al. 1994). This again is consistent with Knudson's model; the allele loss represents the second event and serves to uncover the recessive \textit{BRCA1} mutation to trigger tumorigenesis.
Functional studies have also suggested the presence of a tumour suppressor gene on chromosome 17q. Transfection of the long arm of chromosome 17 into a breast cancer cell line MDA 231 resulted in suppression of tumorigenic phenotype (Negrini et al. 1992). Introduction of a normal chromosome 17 into wild-type p53 expressing MCF 7 cells resulted in growth arrest, suggesting the presence of a tumour suppressor gene other than TP53 on chromosome 17 (Casey et al. 1993).

Apart from these indications that BRCA1 is a tumour suppressor gene, no additional clues regarding its function or identity have been available. Therefore, in order to isolate and ultimately elucidate the functions of this important gene, it was necessary to employ a positional cloning strategy.

1.5 Isolation of BRCA1 by positional cloning

Since the BRCA1 gene was identified while the work described in this thesis was still in progress, this fifth section of the Introduction outlines the main approaches used and the steps that led to this discovery and refers to relevant contributions included in this thesis.

1.5.1 Generation of a genetic map of the BRCA1 region

The discovery in 1990 of linkage in breast cancer families to a polymorphic marker D17S74 (CMM86) (Hall et al. 1990) was followed by several other independent linkage studies. Several of their findings were inconclusive due to genetic heterogeneity and also, as it soon became clear, the marker D17S74 is actually a considerable distance from the gene (Fig 1.5.1(i)). However, linkage to this region of 17q was clearly confirmed in three large pedigrees with an inherited predisposition to both breast and ovarian cancers (Narod et al. 1991). Subsequent linkage studies concentrated on families with multiple cases of breast and ovarian cancer cases, such as the one depicted in Fig 1.5.1(ii) (Spurr et al. 1993) following the discovery that nearly all such families are linked to BRCA1 (Easton et al. 1993).
Fig. 1.5.1(i). Genetic map of the BRCA1 region.
The markers D17S776 and D17S78 were the closest published markers prior to the isolation of BRCA1 and narrowed the region of interest to 1.0-1.5cM, a distance of roughly 1-2 megabase pairs. The discovery of recombination events at D17S1321 and D17S1327, which are an estimated 600kb apart lead to the isolation of BRCA1. The polymorphic marker, D17S855 resides within the BRCA1 gene. Abbreviations for gene markers are THRA1, thyroid hormone receptor; RARA, retinoic acid receptor a; TOP2, topoisomerase 2; KRT10, keratin 10; IGFBP-4, insulin-like growth factor binding protein 4; GAS, gastrin; EDH, oestradiol dehydrogenase; 1A1.3B, a novel B-box protein; RNU2, a component of the small nuclear ribonucleoprotein complex; PPY, pancreatic polypeptide; EPB3, erythrocyte surface protein band 3; GP2B, glycoprotein 2B; PHB, prohibitin; NME1/NME2, polypeptides from a protein involved in reducing metastatic potential.
Fig. 1.5.1(ii). A typical breast/ovarian cancer family linked to BRCA1.
The mean age of onset of breast or ovarian cancer is 42 years. Three cases of bilateral breast cancer have occurred in the family. (from Spurr et al 1993)
Collective data from analyses of meiotic breakpoints in breast cancer families and high density genetic maps succeeded in narrowing down the region to 1.0-1.5 centimorgans (cM) between the markers D17S776 and D17S78 (Anderson et al. 1993; Bowcock et al. 1993; Chamberlain et al. 1993; Kelsell et al. 1993; Porter et al. 1993; Simard et al. 1993; Albertsen et al. 1994; Goldgar et al. 1994). The discovery of meiotic breakpoints between the markers D17S1321 and D171325, a distance of an estimated 600 kilobase pairs (kb), further reduced the region and led to the discovery of the BRCA1 gene (Miki et al. 1994).

The large numbers of polymorphic probes isolated from the BRCA1 region enabled searches for evidence of linkage disequilibrium; namely the occurrence in a population of two genetic markers together at a higher than expected frequency. Such a discovery might have lead to the identification of BRCA1 if one of these alleles was a BRCA1 susceptibility allele and the other a closely linked marker. The occurrence of linkage disequilibrium depends on the mutational history of the disease gene and flanking markers as well as the recombinational history of its surrounding chromosomal region. It is more likely to be evident if many affected individuals have the same BRCA1 mutation by descent. The failure to find evidence of linkage disequilibrium in breast cancer families is now known to be due to a large number of different susceptibility alleles at the BRCA1 locus (see section 1.5.8).

1.5.2 Hunting for cytogenetic abnormalities: clues to BRCA1 location.

Searches for gross chromosomal aberrations in attempts to provide clues to the location of BRCA1 have been largely unsuccessful. Studies of LOH, which have been crucial to the success of other positional cloning projects concerning tumour suppressor genes, failed to pinpoint the location of BRCA1. This is probably due to the presence of multiple tumour suppressor loci on chromosome 17 in addition to BRCA1 such as NF1, TP53, PHB a gene for a putative antiproliferative factor known as prohibitin (Sato et al. 1992) and NME1 and 2, genes encoding components of a protein whose loss may encourage metastasis (Steeg et al. 1988; Stahl et al. 1991; Kelsell et al. 1993). As a result, the regions of loss observed in breast and ovarian tumours have been very large and ill-defined (Cornelis et al. 1993; Lindblom et al. 1993; Tavassoli et al. 1993; Kirchweger et al. 1994).

Another method used to hunt for cytogenetic abnormalities utilises the technique of pulsed field gel electrophoresis. High molecular weight DNA is prepared from normal and tumour cell lines from familial breast and ovarian cancer cases and digested with a
number of restriction endonucleases, whose sites are found rarely in the genome. Electrophoresis of the digested DNA samples followed by Southern analysis using single copy probes from throughout the region was performed in order to identify any gross chromosomal rearrangements which may be associated with tumorigenesis (as shown in chapter 7). These rearrangements would be visible on the resulting autoradiographs as aberrant bands in comparison to normal control samples. This method identified the position of the Von Hippel-Lindau disease gene by identifying germline deletions in affected individuals (Richards et al. 1993) and localised the NF1 gene by identifying translocations in NF1 patients (Collins et al. 1989). Unfortunately, no such alterations were visible in the breast and ovarian cancer patients analysed in this thesis (see chapter 7, (Brown et al. 1994)).

As a result of the failure of these techniques to indicate precisely the location of BRCA1, progress towards its isolation had to rely entirely upon information provided by genetic studies.

1.5.3 Previously characterised candidate genes in the region

The criteria for deciding the candidacy of a gene for BRCA1 were multiple. As mentioned, previously characterised cancer genes have taken many forms (table 1.2.1). Hence perhaps not surprisingly, a number of biologically plausible candidate genes which had been previously characterised and mapped to the chromosome 17q region, presented themselves following the initial localisation of BRCA1. They included the retinoic receptor alpha gene RARA, a gene encoding a member of the steroid-thyroid hormone receptor superfamily. The importance of its product in cellular differentiation makes this a candidate for BRCA1, especially in the light of its involvement in a form of leukaemia (Borrow et al. 1990). Similarly THRAl, a gene encoding a thyroid hormone receptor was also an initial candidate due to its location proximal to RARA on chromosome 17q. Both genes were ruled out by recombination studies which suggested BRCA1 was located further distal (Bowcock et al. 1993; Simard et al. 1993). These same studies also ruled out the involvement of the insulin-like growth factor binding protein gene, IGFBP-4 (Tonin et al. 1993). Its product regulates of the activity of insulin-like growth factors which play an important role in normal breast development and are potent mitogens for breast cancer cells (Cullen et al. 1992).

Further telomeric on the chromosome the gene PHB, a putative tumour suppressor gene whose product may have antiproliferative effects, appeared a strong candidate. However, despite demonstrations of mutations in sporadic breast tumours (Sato et al. 1992), its
involvement in breast cancer predisposition was ruled out due to its location outside of the limits defined by genetic studies (Black et al. 1993; Bowcock et al. 1993; Chamberlain et al. 1993). These studies also eliminated two genes NME1 (Steeg et al. 1988) and NME2 (Stahl et al. 1991), which encode polypeptide chains of a heterodimeric protein thought to be involved in the reduction of metastatic potential of breast carcinomas.

Two candidate genes which encode the enzyme 17\(\beta\)-oestradiol dehydrogenase (EDH) remained even after the BRCA1 region had been narrowed to 1.0-1.5 cM. This enzyme catalyses the synthesis of the most important oestrogen, 17\(\beta\)-oestradiol from its weak precursor, oestrone. Considering the dominant role that oestrogens play in the development, growth and function of all tissues involved in reproduction and fertility in women, genes which can affect their abundance can be considered potential oncogenes. The EDHI gene encodes a truncated EDH-like protein whilst the adjacent EDHII gene encodes the fully functional enzyme (Luu The et al. 1989; Luu-The et al. 1990; Touitou et al. 1994). The production of two homologous EDH proteins with potential for dimerisation presented the possibility that the product of EDHI may be able to modulate the activity of the product of EDHII in a dominant-negative manner (Touitou et al. 1994). Mutations in either gene could therefore affect the abundance of EDH and as a consequence, the abundance of 17\(\beta\)-oestradiol in cells. Thus, the EDH genes were the subjects of intensive mutational searches (Kelsell et al. 1993; Simard et al. 1993; Mannermaa et al. 1994). However, no disruptive mutations were discovered within the germline or tumour DNA of affected breast cancer family members.

1.5.4 Construction of a physical map

As it became apparent that none of the initially tantalising candidate genes in the region were BRCA1, work moved on to characterise the chromosomal region further and to isolate other genes residing in the region with the hope of finding the culprit locus.

The basis for this work was provided by the segregation analyses of several large breast cancer kindreds (Easton et al. 1993). These analyses enabled the generation of high density maps (Anderson et al. 1993; Albertsen et al. 1994). This gave a probable order of the genetic markers and rough estimates of the genetic distances between them. Such genetic maps cannot provide accurate estimates of the physical distances separating the markers since recombination does not occur uniformly along each chromosome. Better definition of the candidate region could only be achieved through the use of physical mapping tools which could convert the genetic distances, measured in
Introduction

centimorgans into physical distances, measured in kilobase pairs. A detailed physical map could pinpoint exactly the location of markers and clones, identify features of the genomic region which could help orientate genomic cloning efforts and indicate the likely position of unidentified genes.

Physical mapping efforts began with the use of chromosome 17 somatic cell hybrid mapping panels. These consisted of a selection of human/rodent hybrid cell lines, some of which contained defined subchromosomal fragments of chromosome 17, whilst others contained well-characterised chromosome 17 translocations. Hybrid cell lines which contained chromosome 17 as the only human chromosome were also included in these panels (Xu et al. 1988; Abel et al. 1993; Black et al. 1993; O'Connell et al. 1994). The mapping panels were scored for the presence or absence of specific human chromosome 17q DNA markers by either Southern hybridisation or PCR. Markers could be ordered on the basis of their segregation in the hybrids by using a framework of anchor markers of known order, determined genetically on families (Fain et al. 1992; Solomon and Ledbetter 1992). The resolution achieved by this mapping method improved upon that which had been provided by linkage analysis and provided a useful framework for subsequent physical mapping and cloning work.

Fluorescence in situ hybridisation (FISH) also served as a bridge between genetic mapping and physical mapping in the \textit{BRCA1} region. A number of polymorphic markers and genes spanning several megabases throughout the \textit{BRCA1} region were ordered using metaphase and interphase chromosomes (Flejter et al. 1993). Adaptations of this technique also proved very useful in confirming clone contig development and the integrity of genomic clones carried out in this thesis (chapter 5) (Jones et al. 1994).

In order to further refine the physical map, long-range restriction mapping with the aid of pulsed field electrophoresis (PFGE) analysis was employed (shown in chapter 3). This technique allows DNA fragments as large as 10,000 kb to be separated (Schwartz et al. 1983; Schwartz and Cantor 1984). When combined with restriction enzymes which cut only rarely in the genome and Southern analysis using single copy fragments from the region of interest, this approach enables the construction of a map of spanning several megabases (Smith and Cantor 1987). Long range maps of the Duchenne muscular dystrophy region (van Ommen et al. 1986), the major histocompatibility complex (Lawrance et al. 1987) and the cystic fibrosis region (Drumm et al. 1988) have been constructed in this way.
In addition to identifying the distances between markers and their more precise localisation in the region, PFGE analysis can identify features such as restriction site polymorphisms and CpG islands (detailed in chapters 3 and 4). Genetically-determined polymorphisms are recognised by the occurrence of two bands of equal intensity in the completely digested genomic DNA upon hybridisation of a single probe. When the same probe is hybridised to a range of genomic DNA samples from different individuals, a characteristic polymorphic pattern of bands can be seen corresponding to the variety of different alleles. Restriction site polymorphisms can therefore be used as signposts in the physical map. Another useful feature recognised by PFGE analysis, the CpG island, is recognised by the occurrence of clustered sites for certain rare-cutting restriction endonucleases (Bird 1987; Larsen et al. 1992). These islands consist of a high density of non-methylated CpG and are situated around the transcription start sites of many genes.

In the creation of a long range restriction map by PFGE analysis, two main types of restriction endonucleases can be used: those which cut predominantly at CpG islands and indicate the location of gene transcription sites, and those which cut mostly at inter-island sites and can extend the map to adjacent areas (Larsen et al. 1992).

Restriction endonucleases which cut between CpG islands commonly result in partial digestion. This is due to methylation which occurs to varying degrees within inter-island sites and can prevent restriction (Smith et al. 1987). This feature can be useful in the construction of a long-range physical map as it can produce larger restriction fragments which may hybridise to more than one probe and therefore extend the physical map. However, the variation in methylation which occurs from cell line to cell line requires that PFGE mapping analysis must be carried out on one cell line only. Hence as a result of these differences direct comparisons between maps generated from different cell lines may reveal discrepancies.

Once several regional long range restriction maps were generated with this methodology, adjacent maps were joined together with the aid of linking clones (shown in chapter 3). These clones contain sequences that span rare-cutter sites in the genome (Poustka and Lehrach 1986) and therefore hybridise to two adjacent fragments on PFGE analysis. A NotI linking clone library from human chromosome 17q generated previously in this laboratory (Borrow et al. 1991) proved very useful in extending long-range restriction maps of this region (see chapter 3 ). In regions where linking clones were not available, maps could be joined together if two probes were seen to hybridise to the same partial restriction fragment.
1.5.5 Genomic cloning of the BRCA1 region

Several different types of genomic clone were utilised in the efforts aimed at cloning the entire BRCA1 region and using the clones to isolate genes (Table 1.5.5) (reviewed (Monaco and Larin 1994)). Clones isolated from the region provided additional probes which could be used to add detail to existing physical maps and to isolate neighbouring genomic clones. Analysis of the clones themselves helped to refine the physical map still further and indicated the exact locations of transcriptional activity.

Yeast artificial chromosomes (YACs) were the obvious first choice in the cloning efforts due to their ability to house several hundred kilobases of exogenous DNA ((Burke et al. 1987); reviewed (Schlessinger 1990)). They have been the major cloning vector in the genome projects of a variety of organisms, including a first-generation physical map of the entire human genome(Cohen et al. 1993). In human positional cloning projects such as the Huntington's disease gene on chromosome 4 (Bates et al. 1992) and the dystrophin gene on chromosome X (Coffey et al. 1992), YACs have proved a crucial part. In the isolation of BRCA1, several different YAC libraries, generated at the Imperial Cancer Research Fund (ICRF) (Larin et al. 1991), Centre d'Etude du Polymorphisme Humain (CEPH) (Albertsen et al. 1990), and the Washington School of Medicine (Brownstein et al. 1989), were screened to obtain YAC clones spanning the region. Screening methods included hybridisation with probes corresponding to known genes and anonymous markers and PCR-based screening with chromosome-specific sequence-tagged sites (STS) (Green and Olson 1990). YAC clone contigs obtained using these methods were extended using the techniques of Alu-PCR walking (Nelson et al. 1989; Breukel et al. 1990) and YAC end-clone walking (Traver et al. 1989; Nelson 1990). The results of these efforts were incomplete YAC contigs spanning the BRCA1 region, as detailed in chapter 5 (Albertsen et al. 1994; Jones et al. 1994; Brown et al. in press).

Although large portions of the BRCA1 region were cloned in YACs, problems of inter-chromosomal chimaerism, rearrangements and deletions made them an unreliable resource for physical mapping and prevented their use in the direct isolation of genes. These problems are partly due to repair mechanisms in S. cerevisiae which are initiated following strand damage (Albertsen et al. 1990) and an inherent instability of these types of clones. In addition, preparation of YAC DNA is a lengthy and cumbersome process as a result of its yeast host and low transformation efficiency. For these reasons, most groups involved in the BRCA1 project turned to alternative cloning systems.
<table>
<thead>
<tr>
<th>CLONE TYPE</th>
<th>CAPACITY</th>
<th>HOST</th>
<th>TRANSFECTION METHOD</th>
<th>VECTOR TYPE</th>
<th>DISADVANTAGES</th>
<th>ADVANTAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAC</td>
<td>100-1000 kb</td>
<td>yeast</td>
<td>spheroplast fusion</td>
<td>linear</td>
<td>chimaerisms, rearrangements, deletions, lengthy DNA preparation, poor yield</td>
<td>large capacity</td>
</tr>
<tr>
<td>cosmid</td>
<td>35-45 kb</td>
<td>E. coli</td>
<td></td>
<td>plasmid</td>
<td>small capacity</td>
<td>stability, high copy number and easy DNA preparation</td>
</tr>
<tr>
<td>PI</td>
<td>70-100 kb</td>
<td>E. coli</td>
<td>infection</td>
<td>plasmid</td>
<td>low copy number and DNA yield, elaborate packaging system</td>
<td>stability, selection for recombinants</td>
</tr>
<tr>
<td>BAC</td>
<td>~300 kb</td>
<td>E. coli</td>
<td>electroporation</td>
<td>plasmid</td>
<td>no selection for recombinants, low copy number and DNA recovery</td>
<td>stability, large capacity</td>
</tr>
<tr>
<td>PAC</td>
<td>130-150 kb</td>
<td>E. coli</td>
<td>electroporation</td>
<td>plasmid</td>
<td>low copy number and DNA yield</td>
<td>stability, selection for recombinants</td>
</tr>
</tbody>
</table>

Table 1.5.5. Comparison of genomic cloning systems
<table>
<thead>
<tr>
<th>Clone Type</th>
<th>Capacity</th>
<th>Host</th>
<th>Transfection Method</th>
<th>Vector Type</th>
<th>Disadvantages</th>
<th>Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAC</td>
<td>100-1000 kb</td>
<td>yeast</td>
<td>spheroplast fusion</td>
<td>linear</td>
<td>chimaerisms, rearrangements, deletions, lengthy DNA preparation, poor yield</td>
<td>large capacity</td>
</tr>
<tr>
<td>Cosmid</td>
<td>35-45 kb</td>
<td>E. coli</td>
<td>in vitro packaging</td>
<td>plasmid</td>
<td>small capacity</td>
<td>stability, high copy number and easy DNA preparation</td>
</tr>
<tr>
<td>PI</td>
<td>70-100 kb</td>
<td>E. coli</td>
<td>infection</td>
<td>plasmid</td>
<td>low copy number and DNA yield, elaborate packaging system</td>
<td>stability, selection for recombinants</td>
</tr>
<tr>
<td>BAC</td>
<td>~300 kb</td>
<td>E. coli</td>
<td>electroporation</td>
<td>plasmid</td>
<td>no selection for recombinants, low copy number and DNA recovery</td>
<td>stability, large capacity</td>
</tr>
<tr>
<td>PAC</td>
<td>130-150 kb</td>
<td>E. coli</td>
<td>electroporation</td>
<td>plasmid</td>
<td>low copy number and DNA yield, no selection for recombinants</td>
<td>stability, large capacity</td>
</tr>
</tbody>
</table>

Table 1.5.5. Comparison of genomic cloning systems (amended)
Cosmid clones, which have the advantages of quick and simple preparation and a vastly improved stability over that of YAC clones, provided the means for the rapid construction of reliable genomic clone contigs of the region. Previously isolated YAC clones were used to screen a flow-sorted chromosome 17 only cosmid library (Lehrach 1990) with the result that analysis of other genomic regions present in the chimaeric YAC clones was avoided and non-deleted copies of the DNA that is unstable in the YACs could be isolated (detailed in chapter 5) (Jones et al. 1994). The conversion of YACs into chromosome-specific cosmids has been a successful strategy in many other projects, including mapping of chromosome Xp22 (Wapenaar et al. 1994) and the Huntington's disease gene region at chromosome 4p16 (Baxendale et al. 1993; Zuo et al. 1993). The gaps between the cosmid contigs could be filled by cosmid walking, using probes generated from each end of the cosmid insert (see chapter 5). However, the comparatively small insert size (35-45 kb) of these clones limited their use in completing the genomic coverage of the \textit{BRCA1} region.

More recent efforts have used bacteriophage P1-based vectors to overcome the limitations of cosmids and YACs (Ioannou et al. 1994). Bacteriophage P1 vectors can accept inserts in the 70-100 kb range and clones have an increased stability over cosmid clones whilst being just as easy to prepare. A complete genomic clone contig spanning the minimal region containing \textit{BRCA1} has been constructed with the aid of these P1 clones (Neuhausen et al. 1994). The P1 clones have been improved by combining the P1 system with the F-factor based pBAC system which has the advantage of larger insert size, but has no positive selection and a low DNA recovery. The resultant PAC clones have an average insert size of 130-150 kb, can be easily prepared with a stability which is higher than any other cloning system to date. In our laboratory PAC clones were used to complete genomic coverage of part of the \textit{BRCA1} region (Brown et al. in press).

1.5.6 Gene isolation: identification of new candidates for \textit{BRCA1}

The genomic clone contigs described in chapter 5 were used as a resource to isolate genes specific to the \textit{BRCA1} region. Combined techniques of direct cDNA clone isolation, cross homology searches and structure-based techniques to select for functional elements involved in gene expression such as exons and promoters were employed. Whilst each of these techniques have their downfalls (discussed (Parrish and Nelson 1993)), it was hoped that a combination of all of them would result in the isolation of all of the genes in the region. The techniques used are described below and their results are detailed in chapter 6.


Introduction

Direct selection

This protocol involves hybridisation of a PCR-amplified cDNA library to biotinylated cosmid DNA. Streptavidin magnetic beads are then used to immobilise the target DNA for removal of nonspecific cDNAs, followed by elution and amplification of specific cDNAs (Fig. 1.5.6(i)) (Lovett et al. 1991; Korn et al. 1992; Lovett 1994).

This is a rapid technique, capable of detecting very rare transcripts, which has enabled the isolation of genes for a number of disorders (reviewed (Parrish and Nelson 1993)). It enables the production of cDNA libraries which have over 100,000 fold enrichment for the region under study and contain only low levels of contaminant clones. Unlike the other methods of cDNA isolation, it is insensitive to the presence of introns or cryptic splice sites and is not dependent upon the presence of CpG islands. It is also capable of detecting human genes that have diverged from those of other species, since cross-species homology is not required in the selection scheme. Since direct selection is dependent upon expression, its success is governed by the extent of representation in the cDNA library source used in the hybridisation selection. This can be vastly improved by using a pool of cDNA library sources.

The main disadvantages of direct selection are that there is no differentiation between genes and pseudogenes and cDNA clones with a high percentage homology to the genomic region in question, but which map elsewhere in the genome will be isolated with this method. Hence care must be taken to check that cDNA clones isolated actually map to the region in question.

Expression-independent methods to identify the location of transcripts

Methods of CpG island hunting, cross-species homology searches and exon-trapping identify fragments of DNA which contain transcribed sequences. Once identified, such fragments can be used as specific probes on random high density cDNA libraries in order to isolate the complete gene. Hence, although potential coding sequences are identified by these methods, expression must still be confirmed either through analysis at the RNA level or by the isolation of a cDNA clone. However, these stages are approached with a smaller fragment which is already suspected to contain potential exons.
BRCAl region cosmids

biotinylate cosmids by nick translation with biotin-16-dUTP

bind biotinylated cosmid fragments to streptavidin-coated magnetic beads and preanneal repetitive elements

cDNA library inserts

hybridise with cDNA libraries

Immobilise hybrids using a magnetic strip and remove non-specific cDNAs. Elute off the hybridised cDNAs, amplify and repeat the hybridisation step.

clone cDNAs, transform into *E. coli* host, pick colonies into microtitre plates and spot onto filters

Fig. 1.5.6(i). Direct selection protocol.
Cross-species homology searches

This technique has successfully isolated a number of genes, including those responsible for Duchenne muscular dystrophy (Monaco et al. 1986), retinoblastoma (Friend et al. 1986) and Wilm's tumour (Call et al. 1990). The basis of this technique is that coding sequences are much more highly conserved between species than non-coding ones. Candidate DNA fragments are hybridised to Southern filters, or 'zoo blots' containing genomic DNA from a variety of species to determine whether the candidate cross-hybridises across species, class, order or phylum boundaries. A clone which hybridises to other species is generally found to contain coding sequences.

This method requires no knowledge of expression and has no dependence upon having an appropriate cDNA library. However, for the best results only small fragments of DNA should be hybridised to the zoo blots to reduce background hybridisation from repetitive sequences. Therefore, screening a genomic region for the presence of genes using this procedure is a time-consuming process. Regions of transcriptional activity may be missed if they are embedded in DNA fragments which are highly repetitive. In addition, genes that have diverged significantly from other species will not be detected.

Selection for CpG islands

Certain rare-cutting restriction endonucleases can detect the locations of G-C rich sequences surrounding the transcription start sites of many vertebrate genes (Bird 1987; Linsay and Bird 1987). Using these enzymes to carry out restriction digestion analysis on the genomic clones spanning the region can identify DNA fragments containing putative CpG islands and hence the 5' regions of genes. This approach has been successful on numerous occasions, including the isolation of genes at the human major histocompatibility complex (Hanson et al. 1991) and the X chromosome (Yen et al. 1992). It has the advantage of not having a requirement for cDNA libraries and may detect less conserved genes than might be detected by zoo blotting. In addition, the discovery of CpG islands leads directly to the 5' region of the gene, often very difficult to obtain by other methods. The main disadvantage is that whilst the majority of mammalian genes have CpG islands, including all the RNA polymerase II housekeeping genes characterised to date and several tissue-specific genes, some tissue-specific genes do not possess such sequences. These genes therefore go undetected by this method.
**Exon-trapping**

This involves the cloning of small fragments of genomic DNA into a vector which contains splice donor and acceptor sites. Transfection of these constructs into monkey fibroblast COS cells results in the transcription and processing of the exons contained which are tagged with vector sequences for subsequent amplification and cloning (Fig. 1.5.6(ii)) (Buckler et al. 1991; Church et al. 1994). Artefacts are produced when splicing occurs around inserts without exons. Recent adaptations to the splicing vector used in this technique have decreased the artefacts commonly observed with this technique and have increased its efficiency (Church et al. 1994). Since this method allows the direct cloning of exons from genomic DNA with no requirement for expression, it has become very popular in ongoing positional cloning projects and has been central to the isolation of the Huntington's disease gene (Huntington's Collaborative Research Group, 1993), *NF2* (Trofatter et al. 1993) and *BRCA1* (Miki et al. 1994). However, it has variable reliability, produces numerous artefacts and requires a lengthy, involved procedure which can slow the gene hunting process.

### 1.5.7 Genes discovered in the *BRCA1* region

With the aid of direct selection, exon-trapping, zoo-blotting and CpG island identification, several new and interesting genes within the *BRCA1* region were isolated (as detailed in chapter 6) (Emi et al. 1993; Albertsen et al. 1994; Campbell et al. 1994; Futreal et al. 1994; Jones et al. 1994; Brown et al. in press). Initial assessment of the candidacy of these genes for *BRCA1* involved DNA and protein homology searches in order to find any similarities to existing genes which are known to play a role in tumour development. Further information on the newly isolated genes was obtained from analyses of their patterns of expression. Most of the previously isolated genes show a ubiquitous pattern of expression, reflecting their importance in general cell regulation. However, on rare occasions the disease susceptibility gene's expression pattern has reflected the site of tumour development, as in the case of the Wilm's tumour gene (reviewed (Rauscher 1993)). Therefore, initial assessments of this type were used to indicate genes with a high priority for further mutation analysis. Interesting genes isolated in the *BRCA1* region are detailed below:
Fig. 1.5.6(ii) Exon-trapping
Exons from two different genes, shown by striped and shaded boxes, can be isolated from total genomic DNA by cloning genomic DNA segments into an intron of the HIV-1 tat gene in vector pSPL3. Following transient transfection into monkey fibroblast cells, and in vitro transcription, the resulting RNA population is extracted and cDNA synthesised using vector specific primers. Restriction sites in the vector are included in the final PCR products facilitating cloning into an appropriate plasmid vector.
• **MDC**, a novel metalloprotease/disintegrin-like gene. Speculated to function in cell adhesion, this gene was investigated for any involvement in breast tumorigenesis (Emi *et al.* 1993). Although evidence for an involvement of this gene was discovered in two breast cancers out of a total of 650 sporadic breast cancers and 50 sporadic ovarian cancers studied, no mutations were found in familial breast or ovarian tumours. This gene was subsequently ruled out by genetic and physical mapping studies (Simard *et al.* 1993; Chandrasekharappa *et al.* 1994).

• **1A1.3B**, a novel gene encoding a protein with a putative B-box zinc finger coiled coil domain. This gene was isolated using a polyclonal antibody against an ovarian tumour antigen known as CA125 (Campbell *et al.* 1994). Considering this tumour antigen is a highly expressed serum marker in ovarian cancer cases, isolation of its putative gene which mapped to within the *BRCA1* region was of obvious interest. Although no causative mutations were found within the coding region of this gene in sporadic and familial breast and ovarian tumours (as shown in chapter 7), investigations of its promoter region and expression in normal and malignant tissue are still required. This gene is under renewed interest following the discovery of its location adjacent to *BRCA1* and the observation of putative shared bidirectional promoter elements between them (discussed in chapter 8) (Brown *et al.* 1994).

• **A diverged homeobox gene, MOXI** (Futreal *et al.* 1994). This gene is the human homologue of the mouse *Moxl* gene (Candia *et al.* 1992). It was considered a candidate because it resides within the *BRCA1* region and contains a homeobox domain with the potential to regulate growth and differentiation. The absence of mutations in the coding region of *MOXI* in *BRCA1* carriers suggested this was not *BRCA1* (Futreal *et al.* 1994).

• **The dual specificity VH1-related phosphatase gene, HDSPH.** A member of the protein tyrosine-phosphatases (PTPases), this gene has potential roles in cell signalling, cell growth and proliferation and oncogenic transformation (Ishibashi *et al.* 1992). A possible role in the reversal of the effects of protein tyrosine kinases implicated the PTPases as candidate tumour suppressor genes (Saito and Streuli 1991). Hence the discovery of this gene in the middle of the *BRCA1* region (Jones *et al.* 1994; Kamb *et al.* 1994) (detailed in chapter 6) led to its mutational analysis within sporadic and familial tumours. No causative mutations have been identified, but the possibility that decreased expression levels of *HDSPH* has a role in tumour progression has yet to be ruled out (see chapter 7).
• **HAL64, a novel ADP-ribosylation factor gene.** This gene appears to be related to the highly conserved family of GTP-binding proteins with homology to both ras oncogene superfamily and the heterotrimeric G protein α subunits (Kahn *et al.* 1993). It is most closely related to a processed chicken pseudogene, *CPSI* which has been implicated as an oncogene due to its similarity to the *C-HRAS* oncogene and possible involvement in early embryonic development (Alsip and Konkel 1986). As a result of these interesting homologies, *HAL64* was investigated as a potential *BRCA1* gene. Both SSCP and DNA sequence analysis failed to identify disruptive mutations in over 100 affected individuals from breast/ovarian families or over 100 sporadic breast tumours (detailed in chapter 6) (Black *et al.* in preparation).

• **Complex integration site for the human endogenous retrovirus, HERV-K.** Retrovirus related sequences exhibit a number of features giving them potential for involvement in carcinogenesis (Leib-Mosch *et al.* 1990). Proviral integration can cause activation of adjacent cellular protooncogenes by promoter insertion, as observed by induction of lymphomas by the viral promoter-induced activation of the cellular gene, *CMYC* (Hayward *et al.* 1981). In addition, stimulation of HERV-K expression in the breast carcinoma cell line, T47D by female steroid hormones has been observed. It has been suggested that this enhanced expression may be involved in the aetiology of certain human breast cancers similar to the MMTV-induction of tumours in mice (Ono *et al.* 1987). The discovery of HERV-K sequences stretching over ten kilobase pairs of genomic DNA within the *BRCA1* region was therefore of interest (detailed in chapter 6) (Jones *et al.* 1994). Although this integration site does not appear to be *BRCA1*, a possible role for it in tumour development has not been ruled out.

• **Other genes residing in the region.** Many other genes were found to reside within the 1.0-2.0 megabase pair region defined by genetic studies as the minimal region containing *BRCA1*. They include genes with homologies to the Ki antigen, a rat ribosomal protein, a dog GTP binding protein and a yeast transcriptional activator (Albertsen *et al.* 1994). Also, clones with homologies to the rho group of G proteins and the interferon-induced leucine zipper protein IFP-35 were discovered in the vicinity of *BRCA1* (Brown *et al.* in press). Many other cDNA clones were obtained which had no protein or nucleotide sequence homologies to anything in the databases. The role of these genes in breast and ovarian tumour development has not been completely assessed but, with the exception of the *BRCA1* gene, it seems unlikely that any of them are involved.
1.5.8 Mutation hunting

One of the criteria for \textit{BRCA1} was the presence of disruptive mutations within the gene in the germline and tumour DNA of affected chromosome 17q-linked breast cancer only and breast/ovarian cancer family members. Such disruptive mutations may be detectable at the DNA level through sequence analysis methods; the RNA level with the aid of northern analysis, RNase protection or RT-PCR; or the protein level through the protein truncation test (Roest \textit{et al.} 1993).

Mutations in previously isolated cancer genes have taken many forms. Small deletions and point mutations are frequent observations and result in a loss of protein expression or the production of a functionless truncated protein product, as in the cases of \textit{NFI} (Xu \textit{et al.} 1990), \textit{VHL} (Latif \textit{et al.} 1993) and \textit{hMSH2} mutations (Fishel \textit{et al.} 1993). Similar mutations result in a truncated protein product which can bind to wild-type protein and disrupt its activity in a 'dominant-negative' manner, as in the case of \textit{APC} and \textit{TP53} mutations (Su \textit{et al.} 1993; Dittmer \textit{et al.} 1994). Single base pair mutations can have devastating effect on the resulting protein, resulting in frameshift with a consequential premature termination of the protein product, or as in the case of \textit{MEN2} mutations in the \textit{RET} gene, a single amino acid change which results in an oncogenic gain of function in the protein (Schuffenecker \textit{et al.} 1994). Mutations at intron/exon boundaries may result in incorrect splicing which can be visible at the DNA level if sequence analysis includes the analysis of these sites or the RNA level since products of transcription will be altered. Mutations of this type have been found in the \textit{WT-1} gene (Bickmore \textit{et al.} 1992) and the \textit{hMSH2} gene (Fishel \textit{et al.} 1993). Alternatively, mutations at the promoter regions may abolish transcription all together.

Initial screening involved in the \textit{BRCA1} project Southern analysis of single copy probes on digested genomic DNA samples from \textit{BRCA1}-linked families, a rapid method which does not require detailed knowledge of the structure and sequence of the gene (as shown in chapter 7). This can detect large deletions and insertions by the presence of junction fragments or changes in band intensities in the resulting autoradiographs. Point mutations may also be detectable if they alter restriction sites. Unfortunately, such screening efforts failed to identify disruptive mutations within candidate \textit{BRCA1} genes and it became obvious that more rigorous detection methods were required to detect subtle changes in gene sequences.

The quickest and most popular method of mutation screening used in the \textit{BRCA1} project was single-stranded conformation polymorphism analysis (SSCP) (Orita \textit{et al.} 1989;
Hayashi 1992) (Fig 1.5.8). Small fragments of the gene are amplified by PCR, denatured and then electrophoretically separated through non-denaturing polyacrylamide gels. The two single-stranded DNA molecules from each denatured PCR product assume a three-dimensional conformation which is dependent on the primary sequence. In the event of a sequence difference between wild-type and mutant DNA, this may result in differential migration of the denatured strands. Products with altered migration patterns can then be analysed by DNA sequencing to determine the exact nature of the alteration. This method can be capable of detecting 70-95% of mutations in PCR products of 200 base pairs or less under optimal conditions (Michaud et al. 1992), although the sensitivity decreases with the size of the PCR product. For a complete analysis of a transcript, SSCP can be carried out on cDNA samples produced from affected family members. However, this may not be able to detect instances of mutations in genomic DNA which disrupt transcription. In addition, mutations affecting splice sites and untranslated parts of the gene may go undetected. Therefore for a thorough mutation search it is preferable to carry out SSCP analysis on genomic DNA samples, requiring detailed knowledge of the gene structure and the sequences surrounding intron/exon boundaries.

Ultimately, all putative mutations have to be sequenced in order to define the location and nature of the changes. Several protocols have been devised to make sequencing a more rapid, accurate and efficient method of mutation detection (reviewed (Grompe 1993), and with the advent of automation and new fluorescence detection technology, sequencing may well become the preferred technique.

These methods were used to search for sequence differences in both cDNA and genomic DNA samples from BRCA1 carriers using primers from many of the genes found to reside in the BRCA1 region. With the discovery of BRCA1, a variety of different techniques are now being explored to enable rapid screening of the many samples accumulated from affected family members and to give further insights into the nature and consequences of mutations within BRCA1.
Fig. 1.5.8 Principle of SSCP analysis.
Wild-type and mutant PCR products, shown at the top of the figure, are denatured by heating to 94°C and cooled rapidly on ice to allow the single DNA strands to adopt secondary conformation. The presence of a mutation, indicated here by a dot on the DNA strands, results in an altered single-stranded conformation leading to differential mobility in the gel.
1.5.9 The *BRCA1* gene

Using the techniques and approaches described above, the *BRCA1* gene was isolated and identified by a large group in October 1994. The novel gene was detailed in two papers which showed that it fulfilled the requirements for *BRCA1* as detailed below (Futreal *et al.* 1994; Miki *et al.* 1994).

- *BRCA1* mutations were found to cosegregate with *BRCA1* haplotypes
- mutations which could potentially disrupt the *BRCA1* protein function were found within the gene in carrier individuals from kindreds segregating 17q-linked susceptibility to breast and ovarian cancer
- population frequency analysis suggested these *BRCA1* alleles occurred only rarely in the general population.

Conceptual translation of a composite full-length sequence of the gene indicated it encodes a protein of 1863 amino acids with a zinc finger domain near the NH$_2$-terminus. The gene is composed of 23 coding exons distributed over roughly 100 kb of genomic DNA. Its full-length transcript is 7.8 kb and appears abundant in testis and thymus, but also present in other tissues including breast and ovary. Analysis of cDNA clones has revealed a complex pattern of alternate splicing at the 5' one third of the gene, the implications of which are as yet unknown.

Initial mutational evidence showed it was responsible for susceptibility to breast cancer in three of four families with strong linkage to 17q markers, all of which displayed early onset breast cancer and at least one ovarian cancer case (Miki *et al.* 1994). One of these mutations was an 11 base pair deletion in exon 2 which effectively removed the zinc finger domain and caused a premature termination of the protein at codon 36. Breast cancer susceptibility in a second kindred appeared to be due to a regulatory mutation in the gene. The cDNA from an affected individual from this family showed homozygosity at two polymorphic sites within the gene, whereas the genomic DNA from this same individual revealed heterozygosity at these loci. This indicated that the mRNA from the mutant allele of this individual was absent due to a mutation that affects RNA production or stability. The third mutation was a nonconservative amino acid change. In another family weakly linked to chromosome 17q, a single nucleotide insertion in coding exon 20 resulting in alteration of the protein 107 amino acids from the wild-type carboxy terminus was found to segregate with the disease. The accompanying paper focusing on primary
breast and ovarian carcinomas detected four additional germline alterations in this candidate gene. Three of these were missense alterations and one was a nonsense mutation (Futreal et al. 1994). Interestingly, all of the mutations found were disruptive which appeared to be consistent with the predictions that BRCA1 was a tumour suppressor gene.

Demonstrations of mutations within this gene in other families confirmed its identity as BRCA1 (Castilla et al. 1994; Friedman et al. 1994; Simard et al. 1994; Shattuck-Eidens et al. 1995; Black et al unpublished data; Xu et al unpublished data). Again, the majority of them appeared to lead to protein truncations and a loss of function of wild-type BRCA1 protein, consistent with the tumour suppressor theory. Possibilities of a dominant-negative disruption of wild-type BRCA1 protein function by these truncated protein products have yet to be investigated. However, these studies have demonstrated for certain that BRCA1 mutations occur in many different forms scattered throughout the gene, a fact which presents problems for the future development of accurate prognostic tests.

Part of the excitement about the identification of BRCA1 was due to the expectation that this gene is the initiator of breast and ovarian cancers not only in hereditary cases but also in the much more common sporadic forms of the disease, in accordance with the Knudson two-hit model. The frequent observations of LOH within sporadic breast and ovarian tumours, and the demonstration that this involved the wild-type allele supported such expectations (see section 1.3.3). Furthermore, the majority of the cancer susceptibility genes isolated so far also play a role in sporadic disease. Mutations within the cancer predisposing genes RBI, APC, TP53 and VHL have been shown to be early events in the development of sporadic retinoblastoma, colorectal tumours, breast tumours and renal cell carcinoma respectively for example. Hence, it was surprising that studies of primary sporadic breast and ovarian cancers have failed to identify any somatic mutations (Futreal et al. 1994) (Nicolai et al unpublished data). These studies, which have concentrated on early onset tumours showing allele loss on chromosome 17q, only identified a few mutations which on further analysis turned out to be germline and not somatic.

These findings suggest that although this novel gene appears to be BRCA1 and is the underlying cause of susceptibility to breast and ovarian cancer in a number of families, it does not seem to play a role in the great majority of breast cancers. Possibly, sporadic disease is a result of mutations in another gene on chromosome 17q. One candidate for such a gene is IA1.3B (Campbell et al. 1994) (see section 1.5.6). It appears to reside
immediately adjacent to \textit{BRCA1} with the 5' ends of the two genes lying head to head with bidirectional promoter elements between them (Brown \textit{et al.} 1994). Work is underway to investigate the regulation of these genes and to establish whether their expression is altered in breast and ovarian tumours through the activity of shared promoter and enhancer sequences.

As with the discovery of any novel gene, the identification of \textit{BRCA1} has presented more questions than answers. Work is now needed to determine the functions of the gene product and the implications of mutations. Wider issues regarding the importance of \textit{BRCA1} in familial breast and ovarian cancers and the identification of other predisposing genes must also be addressed. Tests which will sensitively and inexpensively detect mutations in \textit{BRCA1} will need to be devised. The possibility of other genes on chromosome 17q playing a major role in sporadic disease also needs to be investigated.
AIMS OF THIS THESIS

The work described in this thesis centred on the characterisation of the chromosome 17q21 region known to contain the familial breast/ovarian cancer gene, *BRCA1*. It began with constructing physical maps throughout the region and progressed through all the stages of a positional cloning strategy, complementing other efforts being made within the laboratory and elsewhere with the ultimate goal of isolating *BRCA1*.

When this project commenced in December 1991, genetic mapping efforts by a worldwide consortium including this laboratory had succeeded in narrowing the region containing *BRCA1* to a 4cM interval. The search for *BRCA1* required a detailed positional cloning strategy involving the construction of a physical map of the region and the compilation of a complete genomic clone contig. The genomic clones from the region would be used to isolate all of the genes within these limits which would in turn be scrutinised for their potential involvement in breast cancer.

An initial aim of this work was therefore to construct a detailed long-range physical map spanning the region to aid the building of genomic clone contigs. As ongoing genetic mapping efforts carried out by other groups in the consortium gradually narrowed the *BRCA1* region down to a 1.0-1.5 cM region, it became necessary to concentrate the physical analysis carried out in this project to this smaller region. The work aimed to obtain information regarding the precise physical characteristics of this region, including the location of both known and previously unidentified and unmapped genes and markers. It was hoped that a full understanding of the physical characteristics of the region surrounding *BRCA1* would aid future work concerning the mechanisms involved in breast tumorigenesis.

In order to achieve the thorough physical and transcriptional analysis of this smaller region, this project progressed to genomic cloning. This part of the work aimed to complete the genomic coverage in the distal half of the region spanning over 400 kb and to isolate and map new genes contained within the contig. This work would complement that of other members of the laboratory who were constructing genomic clone contigs over the proximal half of the *BRCA1* region.

At the final step in the positional cloning process and following the isolation and precise localisation of several genes, this work progressed to mutational analysis of a selection of genes with the aim of identifying *BRCA1*. 
CHAPTER 2 MATERIALS AND METHODS

The majority of protocols presented here have been previously described (Sambrook et al. 1989). Details of the buffers and solutions used throughout are given in section 2.14. All centrifugation was carried out in a Beckman J6B centrifuge using a JS4.2 rotor unless otherwise stated.

2.1 DNA preparation

2.1.1 Genomic DNA

High molecular weight genomic DNA for Southern analysis was prepared from cultured cell lines in the following manner. Cultured adherent cells were harvested by the addition of 5ml trypsin/versene solution (1:4 mixture) and gentle agitation for 15 minutes. Once the cells had become detached from the flask, 25ml cell culture medium plus serum was added and the cells were decanted into 50ml falcon tubes. The cells were centrifuged at 1.5K rpm for 5-10 minutes and the resultant cell pellets were washed twice with 25ml PBSA. Cells were finally resuspended in 5mls of lysis buffer for every 10^7 cells with gentle agitation to give a homogeneous opaque lysate. This was transferred to a 15ml snap-cap falcon tube and mixed with 0.5mls 2M sodium acetate. The lysate was extracted twice with a 25:24:1 phenol:chloroform:isoamylalcohol mix and once with a 24:1 chloroform:isoamylalcohol mix. During this extraction process, phases were clearly separated by centrifugation at 3K rpm for 15 mins. The aqueous phase was removed and an equal volume of propan-2-ol added. The total nucleic acid was allowed to precipitate at room temperature for 30 minutes and then centrifuged at 3K rpm at room temperature for 15 minutes. The pellet was washed in 80% ethanol/20% 0.01M Tris pH7.5 and resuspended in 5M lithium chloride. After 30 minutes at room temperature, the mixture was centrifuged as before and the supernatant containing the DNA was removed. The DNA was precipitated by the addition of one tenth volume of 3M sodium acetate pH5.2 and 2 volumes of cold absolute ethanol to the aqueous phase. The tube was gently inverted until the DNA became visible as a white stringy mass. This was pelleted by centrifugation for 30 minutes at 3K rpm, washed twice in 70% ethanol and after being allowed to dry, dissolved in TE pH7.5 by rolling the tube overnight at 40°C.
2.1.2 High molecular weight DNA for PFGE analysis

Cultured cell lines were harvested as described previously in section 2.1.1, washed twice in PBSA and resuspended to a concentration of $2 \times 10^7$ cells/ml. The cell suspension was warmed to 45°C and then added to an equal volume of 1.2% low melting point agarose (FMC Bioproducts, Rockland ME, USA) in PBSA also at 45°C. The cell mixture was cast into 0.1ml blocks using clean plexiglass block forms and after solidifying, dropped into 40mls proteinase K digestion buffer in a 50ml falcon tube (no more than 50 blocks per tube). After sealing with parafilm, the tubes were incubated in a 50-55°C water bath for 48 hours with occasional inversion. After lysis, the blocks were rinsed three times with TE and then incubated at room temperature with gentle rocking in two changes of TE with 1mM phenylmethylsulfonyl fluoride (PMSF), each for 1 hour. Blocks were rinsed several times with TE and stored at 4°C in 0.5M EDTA. All traces of EDTA were removed by three 20 minute washes in TE prior to digestion with restriction endonucleases.

2.1.3 Bacteriophage DNA

*Plating bacteriophage*

This method was used in the preparation of *NotI* linking clone DNA (sections 3.1,3.5) and in the preparation of bacteriophage cDNA clones (section 4.2 and section 6.5). The selected clones were picked from their respective microtitre wells from the gridded *NotI* linking clone library (Borrow *et al.* 1991) or from the cDNA library plates and mixed with 100µl of SM buffer. Serial dilutions of the phage stock at $10^{-2}$, $10^{-3}$, $10^{-4}$ and $10^{-6}$ were prepared, each giving a final volume of 100µl. The bacterial host cells were prepared by centrifuging a fresh overnight bacterial culture for 10 mins at 1.5K rpm and resuspending in half the original volume of LB broth supplemented with 10mM MgSO$_4$. To each 100µl aliquot of the phage dilutions, 100µl of prepared bacterial cells (XL1-Blue mrf)' (Stratagene, California USA) were added. The mixtures were shaken and incubated at room temperature for 20 minutes to allow the phage particles to adsorb. During this time, BBL top agarose was melted and cooled to 45°C in a water bath. Aliquots of 3.5ml of the top agarose were dispensed into 15ml falcon snap-cap tubes and kept at 45°C. Each phage/cell mixture was added to one of the
Materials and methods

tubes containing the agarose, gently shaken and poured immediately onto a ready-prepared BBL agar plate (85mm petri dish containing 30ml BBL agar) so that there was an even distribution of top agarose/phage/cells over the surface of the BBL agar. The lids were replaced on the plates and after 5 minutes, the plates were placed in an inverted position in a 37°C incubator overnight. Single plaques could be seen the following morning, usually on the plate on which the lowest dilution of phage had been plated.

Bacteriophage DNA preparation

Phage DNA was prepared by inoculating 50ml LB broth supplemented with 10mM MgSO₄ with an agar plug containing a single bacteriophage plaque. After overnight incubation in a rotatory shaker at 37°C, lysis was invariably apparent and 500µl of chloroform was added to ensure any remaining bacteria had also lysed. The lysate was centrifuged for 30 minutes at 2.8K rpm to remove the debris and the supernatant transferred to a fresh tube with 50µl of 10mg/ml RNAse and DNase. Bacterial nucleic acid was digested for 30 minutes at 37°C, followed by the addition of NaCl to 0.3M, EDTA to 10mM, Tris. HCl (pH7.6-8.0) to 100mM and 300µl of proteinase K (stock 10mg/ml) and the phage coats were allowed to digest for 30-60 minutes at 50°C. The DNA was precipitated by the addition of 0.6 volumes of propan-2-ol with incubation at room temperature for 20 minutes. After centrifugation at 2.8K rpm for 30 minutes, the DNA was dissolved in 200µl TE and extracted once with 25:24:1 phenol: chloroform: isoamylalcohol and again with a 24:1 chloroform: isoamylalcohol mixture. The bacteriophage DNA was precipitated by the addition of sodium acetate to 0.3M and 2 volumes of cold absolute ethanol. After a 70% ethanol wash, the DNA was dried and dissolved in 200µl TE.

2.1.4 Cosmid and plasmid clone DNA

All protocols employed the basic alkaline lysis method. For general restriction enzyme and Southern analysis, mini- and midi- preparation protocols were used. Where better quality DNA was required, as in the cases of sequencing, riboprobe preparation and FISH analysis, DNA preparation was performed using QIAGEN purification columns in accordance with manufacturer's instructions (QIAGEN Inc, USA).
Materials and methods

Mini-preparations

Small scale preparations were prepared by the following rapid method. Single colonies of transformed bacteria were inoculated into 1.5ml LB broth containing the appropriate antibiotic (50μg/ml ampicillin for plasmids based on pBluescript; 50μg/ml kanamycin for cosmids of lawrist 4 type) and incubated with vigorous shaking overnight at 37°C. The culture was then centrifuged at 13K rpm in a microfuge for 30 seconds. The supernatant was discarded leaving 50-100μl of liquid in the eppendorf tube and the pellet was resuspended by vortexing. Next, 300μl of TENS buffer was added and the mixture vortexed for 2-3 seconds. Following the addition of 150μl 3M sodium acetate, the mixture was vortexed again briefly and centrifuged for 5 minutes in a microfuge at 13K rpm. The supernatant was removed into a fresh tube and 0.9ml cold ethanol added to precipitate the DNA. After centrifuging again for 2 minutes, the DNA pellet was washed with 70% ethanol and air-dried. The pellet was resuspended in 30μl TE/RNase.

Midi preparations

Single colonies of transformed bacteria were inoculated into 50ml LB broth in a 500ml flask and incubated with shaking overnight at 37°C in the presence of the selective antibiotic. The culture was placed on ice for 15 minutes, transferred to a pre-cooled 50ml falcon tube and then centrifuged for 20 minutes at 2.8K rpm at 4°C. Next, 6mls of Solution I was added and the cells mixed up into an emulsion. Cells were lysed with 12mls Solution II and the mixture was gently inverted a few times and allowed to sit on ice for 5 minutes. The mixture was shaken vigorously following the addition of 8mls Solution III and placed on ice for a further 10 minutes. The white bacterial debris was pelleted by centrifugation at 2.8K rpm for 25 minutes at 4°C and the supernatant was transferred into a fresh falcon tube. Any remaining protein contamination was removed by two extractions with 24:1 chloroform: isoamylalcohol. Finally, DNA was precipitated by adding 15mls propan-2-ol to the extracted aqueous phase. After 15 minutes' incubation at room temperature, the DNA was pelleted by centrifugation at 2.8K rpm for 15 minutes at room temperature. The pellet was washed thoroughly in 30mls 70% ethanol and after air-drying, resuspended in 500μl TE/RNase.

Preparation of high quality cosmid and plasmid DNA

DNA from 50 ml overnight cultures was prepared using the QIAGEN-tip 100 purification kit according to the manufacturer's instructions (QIAGEN Inc, USA). The resulting DNA pellets were resuspended in 250μl TE.
2.1.4 YAC DNA

The YAC blocks prepared here were used for PFGE Southern and PCR analysis. The yeast containing the YAC were streaked out onto YPD agar and incubated at 30°C for 2-3 days. A single colony from the plate was used to inoculate 30mls YPD media and incubated with agitation overnight at 30°C. Following centrifugation at 1.5K rpm at room temperature, the yeast was resuspended in 15mls of 10mM TrisHCl/50mM EDTA. The yeast was pelleted again by centrifugation at 1.5K rpm, 15 minutes and resuspended in 15mls SCE. After another identical centrifugation step, the yeast was resuspended in 600μl of SCE+DTT and transferred to an eppendorf tube. Freshly made Novozyme (Novo Biolabs, see section 2.13) was added to 8mg/ml and the mixture was incubated at room temperature for 5-10 minutes. After this time, an equal volume of molten 1.2% LMP agarose in SCE was added and after thorough mixing, the yeast suspension was aliquotted into 100μl block formers on ice and allowed to solidify. The blocks were then transferred to a tube containing SCE+DTT (1 block/ml) and incubated for 1-2 hours at 37°C with occasional inversion of the tubes. After this, the SCE solution was replaced with yeast lysis buffer and the blocks were incubated at 37°C overnight. The following day, the blocks were washed twice in 50mM EDTA/10mMTris.HCl with agitation and stored at 4°C in the same solution.

2.2 DNA restriction endonuclease digestion

The enzymes used were purchased from New England Biolabs (Beverly MA, USA), Northumbria Biologicals Ltd (Northumberland, UK) (NBL) and Boehringer Mannheim (Lewes, UK) and in general restriction enzyme digestion was carried out according to the manufacturers' instructions. The number of enzyme units added and the length of digestion were determined assuming that 1 unit of enzyme digests 1μg of DNA in 1 hour, with a margin of error to ensure complete digestion. Where partial digestion was required, less enzyme was used and several digestions using different enzyme quantities had to be carried out in order to determine the enzyme concentration which would produce the desired effect. Where double digestions were required, both enzymes were added together if the buffers permitted, or sequentially if not.

2.2.1 Genomic DNA for conventional Southern analysis

In general, the high molecular weight DNA prepared using the protocol described was completely digested after a 4 hour incubation at the appropriate temperature with the required enzyme and buffer. To check that digestion was complete, a small aliquot of the
digest was electrophoresed through a mini-agarose gel containing ethidium bromide and visualised on an ultraviolet transilluminator. After complete digestion, the entire DNA was precipitated with a tenth volume 3M sodium acetate and 2 volumes of cold absolute ethanol, pelleted by centrifugation at 13K rpm for 15 minutes in a microcentrifuge and after washing with 70% ethanol, resuspended in an appropriate volume of TE.

2.2.2 PFGE blocks

Each block was washed twice in TE prior to digestion with restriction endonucleases. A block was transferred to an eppendorf tube and any residual TE removed before adding the digestion mix, as shown below.

<table>
<thead>
<tr>
<th>Typical digestion mix</th>
<th>Agarose block (80µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x buffer</td>
<td>(20µl)</td>
</tr>
<tr>
<td>BSA (10mg/ml)</td>
<td>(5µl)</td>
</tr>
<tr>
<td>dH2O</td>
<td>(90µl)</td>
</tr>
<tr>
<td>Enzyme</td>
<td>(15-20 units)</td>
</tr>
<tr>
<td></td>
<td>200µl</td>
</tr>
</tbody>
</table>

Incubation was carried out at the appropriate temperature for 4 hours and was stopped by the addition of 0.5M EDTA.

2.2.3 Genomic clone and cDNA clone digestion

In general for Southern analysis, 1-5µg of cloned DNA was digested in a 25µl volume with the desired enzyme and buffer and at the appropriate temperature for 30 minutes - 2 hours. The digest was immediately electrophoresed through an agarose gel after the addition of loading buffer.

2.3 Electrophoretic size separation of DNA

2.3.1 Conventional agarose gel electrophoresis

DNA fragments were size-separated in electrophoresis tanks (BioRad DNA sub cells) (BioRad Laboratories Inc, California USA) through agarose gels containing ethidium bromide at 1µg/ml with TAE buffer. The concentration of the agarose gel was varied according to the average fragment size of the DNA in order to achieve optimal separation: genomic DNA restriction digests were separated through 0.8% gels, PCR
fragments through 1.0-1.5% gels. Size markers used included HindIII digested λ DNA (NBL), PstI digested λ DNA (NBL) or 100bp ladder (Pharmacia Biotech, USA). In general, genomic DNA which required long separation in order to achieve good resolution was electrophoresed overnight at 25V using a 15x20cm agarose gel. Cloned DNA was usually run much faster at around 100V for a few hours. Electrophoresed DNA was visualised with UV light and photographed with Polaroid 667 film.

2.3.2 PFGE electrophoresis

All gels were made up in 0.5x TBE electrophoresis buffer with 1% agarose. Blocks were placed on a comb which was gently lowered into the casting tray. The molten agarose, cooled to 45°C was slowly poured into the tray and allowed to surround the blocks. After the gel had solidified, the comb was gently removed and the holes filled with molten agarose. Gels were run at a circulating temperature of 140°C using the BioRad CHEFII system according to the manufacturer. The parameters used to obtain the desired separation are detailed in Table 2.3.2 and were obtained from the BioRad CHEFMAPPER program. Size markers for the gels were *Saccharomyces cerevisiae* (Sigman) or 50kb lambda concatamers (BioRad).

<table>
<thead>
<tr>
<th>Min size (kb)</th>
<th>Max size (kb)</th>
<th>Run Time (hrs:mins)</th>
<th>Initial switch time (secs)</th>
<th>Final switch time (secs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>400</td>
<td>16:00</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>0</td>
<td>800</td>
<td>22:00</td>
<td>20</td>
<td>84</td>
</tr>
<tr>
<td>0</td>
<td>1000</td>
<td>22:00</td>
<td>20</td>
<td>150</td>
</tr>
<tr>
<td>50</td>
<td>2000</td>
<td>27:00</td>
<td>13</td>
<td>150</td>
</tr>
<tr>
<td>40</td>
<td>705</td>
<td>16:00</td>
<td>2.3</td>
<td>50.3</td>
</tr>
</tbody>
</table>

Table 2.3.2 PFGE parameters

2.4 Preparation of radioactive probes

Probes for Southern and northern analysis were labelled in the same way. The majority were also pre-annealed with human DNA in order to compete out repetitive elements.

2.4.1 DNA probes

Labelling was carried out by the random priming method using [α-32P]dCTP (Feinberg and Vogelstein 1984) Most of the DNA fragments were cut from 1% low melting point gels following electrophoresis. A volume of distilled water equal to that of the gel slice
was added and the slice melted in a 65°C water bath. For a single label reaction, 20-50ng of DNA was aliquotted into an eppendorf tube and made up to 34µl with water. The DNA aliquot was denatured by boiling for 5 minutes and after cooling at room temperature for 1-2 minutes, 2µl of 10mg/ml bovine serum albumin (BSA) was added together with 10µl of oligolabelling buffer, 3µl (30µCi) of [α-32P]dCTP and 1 unit of Klenow polymerase. The reaction tube was placed inside a lead pot and allowed to incubate in a 37°C water bath for 1 hour.

Unincorporated nucleotides were removed on a Sephadex G50 column and the radioactivity of the purified product was measured. In general, probes of 5-6x10^6cpm/ml hybridisation solution were used for genomic Southern and northern analysis, whilst a much lower degree of radioactivity was required for cloned DNA fragment probes which were being hybridised to Southern filters containing cloned DNA. Repetitive elements within the probes were removed by the addition of 250µg of sheared human placental DNA (Sigma, Dorset UK), or human COT-1 DNA (GIBCO Life Technologies Ltd, Paisley UK). The probe was then denatured by boiling and preannealled by incubation at 65°C for 2 hours before adding to the hybridisation solution. Probes which did not require competition were denatured by the addition of a quarter volume denaturing solution and incubation at room temperature for 5 minutes.

2.4.2 Cosmid riboprobes

Runoff RNA transcript probes were made from each end of the cosmid insert in an in vitro transcription reaction using the dual opposed SP6 and T7 promoters flanking the cloning site in the cosmid vector Lawrist 4 (D. Nizetic, unpublished data). Cosmid DNA (1µg) was digested to completion using Rsal. Afterwards, the digest was extracted with 24:1 chloroform: isooamylalcohol and precipitated with a 0.6 volume of propan-2-ol. After centrifugation for 15 minutes at 13K rpm in a microcentrifuge, the pelleted DNA was washed in 1ml 70% ethanol, centrifuged and air-dried before redissolving in 20µl DEPC-treated distilled water. The DNA was used as a template in an in vitro transcription reaction using the riboprobe II core system kit (Promega Ltd, Southampton, UK).
Materials and methods

Riboprobe synthesis was set up by mixing the following in an eppendorf tube:

5μl of 5x in vitro transcription reaction buffer (supplied in kit)
1μl RNAsin (1 unit) (Pharmacia Biotech, USA)
1μl 1M DTT
1μl 10mM ribonucleoside triphosphates rGTP, rATP, rCTP (each)
2.5μl α-32P-rUTP (20mCi/ml) (Amersham Life Science, Amersham, UK)

To this mixture, 6μl of the digested DNA template, 20 units of T7 or SP6 RNA polymerase and DEPC-treated H2O to a final volume of 25μl was added. The mixture was incubated at 37°C for 1-2 hours. Finally, 10 units of RNase free DNase was added to remove the DNA strand leaving a single-stranded RNA probe. After incubation for 15 minutes at 37°C, the reaction was terminated by heating to 65°C for a further 15 minutes. The probe was then added to the hybridisation solution and incubated with the cosmid library filters overnight at 65°C. Hybridisation and washing conditions were the same as those for Southerns.

2.5 Southern analysis

2.5.1 Alkali transfer

To enable efficient transfer of DNA from the agarose gel to membrane, the DNA was acid-nicked. This step was performed in all cases with the exception of the transfer of DNA cloned into plasmid vectors which does not require acid nicking due to its small size. Following electrophoresis, the agarose gel was placed in 0.25M HCl with gentle rocking for 10-15 minutes, until the running buffer dyes in the gel had changed colour. The gel was then rinsed briefly in distilled water and neutralised with 0.4M NaOH until the dyes in the gel had changed back to their original colour. The DNA fragments in the gel were then transferred to a charged nylon membrane, Hybond N+ (Amersham) by the alkali blotting procedure. This involved filling a tray with alkali transfer buffer (0.4M NaOH for DNA) A platform covered with a wick made from three sheets of Whatman 3MM filter paper saturated with alkali transfer buffer was placed in the tray and the agarose gel was then placed on the wick, carefully avoiding trapping air bubbles beneath it. A sheet of Hybond N+ the same size as the gel was placed on top of the gel, again avoiding any air bubbles and three sheets of 3MM paper wetted with the transfer buffer were placed on top of the membrane. A stack of absorbant paper towels was then placed on top of the 3MM paper and a weight was placed on top of the entire stack. In this way, the DNA was simultaneously denatured and drawn up through the gel to the membrane.
by the capillary action of the transfer buffer. The gel was surrounded by cling film to prevent the transfer buffer being absorbed directly into the paper towels above.

The transfer was allowed to proceed overnight in the case of genomic Southerns or a few hours in the case of the transfer of cloned DNA fragments. The apparatus was then dismantled and the membrane rinsed twice in 2xSSC for 20 minutes each in order to remove all traces of alkali. The membranes were then stored moist at 4°C until required.

2.5.2 Prehybridisation and hybridisation of filters

Prehybridisation solution was made up according to Hybond N+ manufacturer's instructions (Amersham). Prehybridisation was carried out in hybridisation bottles (Hybaid, UK) which were rotated in a 65°C oven (Hybaid, UK). Each bottle was filled with a maximum of 4 membrane filters interspersed with nylon gauze and carefully rolled so that there were no air bubbles between the layers and the bottle. Around 25ml of prehybridisation solution was added and the bottles were prehybridised for 2-12 hours to allow complete prehybridisation over the entire surface of the membranes.

In the case of pulsed-field gel Southern membranes, a further prehybridisation step using prehybridisation solution and 100μl denatured sonicated human placental DNA was performed for 2 hours in order to compete out repetitive sequences in the genomic DNA within the membranes.

Following prehybridisation, the solution in the bottles was replaced with hybridisation solution, consisting of prehybridisation solution with the addition of denatured radioactive probe. Hybridisation was carried out overnight at 65°C in the rotating oven. Shorter prehybridisation and hybridisation steps were performed with membranes containing digested cloned DNA.

2.5.3 Washing of Southern filters

Following hybridisation, the solution inside the bottles was replaced with 2xSSC and the bottles were rotated at 65°C for a further 5-10 minutes. The solution was discarded and the filters with their gauzes removed and placed in a plastic sandwich box containing 500ml-1L of 1x SSC, 0.1% SDS. The filters were rinsed with gentle rocking in a 65°C water bath for 10-15 minutes. The solution inside the sandwich box was replaced with fresh 1xSSC, 0.1%SDS and the filters were rinsed for a further 15 minutes in the same way. The washing stringency was increased by replacing the solution with 0.2x SSC,
0.1% SDS and rinsing at 65°C with gentle agitation for a further 15-30 minutes. A final 10 minute wash with 500mls of 0.1x SSC, 0.1% SDS was carried out if the background radiation on the filters was thought too high.

At the end of the washing the filters were removed, immersed briefly in 2x SSC to remove most of the SDS, and wrapped in saran wrap before exposure to X-ray film at -70°C.

2.6 Northern analysis

Northern analysis using MTNI and MTNII multiple tissue northerns (Clontech, California, USA) was carried out according to manufacturer's instructions.

2.6.1 Prehybridisation and hybridisation of filters

The membranes were rolled up with a protective gauze and placed inside a small hybridisation bottle with 10ml prehybrisation/hybridisation solution. Prehybridisation was carried out by rotating the bottle in a hybridisation oven at 42°C for 3-6 hours. After this time, the solution was replaced with fresh solution containing the radiolabelled probe at a concentration of 5-6x10⁶ cpm/ml hybridisation solution. Hybridisation was carried out at 42°C for 18 hours in the hybridisation oven.

2.6.2 Washing northern filters

Northern filters were removed from the hybridisation bottles and rinsed several times at room temperature in 2xSSC, 0.05% SDS for 30-40 minutes. Subsequently, the wash solution was replaced with 0.1xSSC, 0.1% SDS and washing was continued for a further 40 minutes at 50°C with gentle rocking. Finally, the filters were removed and shaken to remove excess wash solution before wrapping in Saran wrap. Exposure to X-ray film was carried out at -70°C using two intensifying screens for a minimum of 3 days.

2.7 Subcloning

DNA fragments from cosmids, bacteriophage and plasmids were subcloned into the multiple cloning site of the plasmid vector pBluescript (Stratagene). Subcloning was generally performed using chemically competent cells (section 2.7.2).
2.7.1 Preparation of insert and vector DNA

Vector DNA (2µg) was digested in a 20µl volume with the desired restriction enzymes then treated with calf-intestinal alkaline phosphatase (CIP) for a further 30 minutes at 37°C with the restriction enzyme buffer. The enzymes were inactivated by addition of trinitriloacetic acid to 10mM and EDTA to 15mM and incubation at 68°C for 15 minutes. The digest was extracted once with phenol: chloroform: isoamylalcohol and once with chloroform: isoamylalcohol, precipitated, dried and redissolved in 30µl of TE. For subcloning of entire cosmids or phage, insert fragments were prepared by digestion of 2µg of the cosmid or phage DNA, extraction as before followed by precipitation and dissolving in TE. When a specific cosmid or phage DNA fragment was required for subcloning, the DNA was digested, electrophoresed through a low melting point agarose gel and the desired DNA fragment excised.

2.7.2 Preparation of chemically competent cells

A single colony of XL1-Blue E. coli host cells (Stratagene) was used to inoculate 20mls of LB broth with tetracyclin (25µg/ml) and incubated overnight at 37°C. A 0.5ml aliquot of this overnight culture was used to seed 100mls LB broth which was incubated until the cells reached early log phase giving an OD600nm reading of 0.3-0.6; around 5x10⁷ cells/ml. When this had been achieved, the cells were pelleted by centrifugation at 1.5K rpm for 10 minutes at 4°C and the supernatant removed and discarded. The cells were resuspended in 10mls transformation and storage buffer and kept at 4°C or quickly frozen in a dry ice/ethanol bath and stored at -70°C for longer periods.

2.7.3 Ligation and plating of transformants

Insert and vector were ligated in a 3:1 molar ratio in a 20µl volume with T4 ligase (Northumbria Biologicals Ltd) and 4µl of 5x ligation buffer (Northumbria Biologicals Ltd). Control ligations were also set up as below:

1. Digested vector DNA, ligase, ligase buffer only. This demonstrated the efficiency of the CIP, the completeness of vector digestion and the presence of contaminating DNA.

2. Digested insert DNA, ligase, ligase buffer only. This control showed the presence of contaminating vector DNA in the insert.
3. Intact pBluescript DNA only. This demonstrated the transformation efficiency of the cells (which were found to be up to $2 \times 10^8$ transformants/µg of DNA with this method).

The ligation proceeded at room temperature for a minimum of 4 hours. During this time, fresh LB agar plates supplemented with ampicillin (20 µg/ml) were spread with 40 µl of X-gal (20 µg/ml stock in dimethylformamide) and 40 µl IPTG (100 mM stock) for subsequent plating. Following ligation, a 10 µl aliquot of the ligation mixture was mixed with a 150 µl of the prepared plating cells in a cold polypropylene tube. The mixture was incubated on ice for 15 minutes, and subsequently plated onto the ready-prepared agar plates. The plates were incubated at 37°C overnight and after a 1-2 hour refrigeration, white transformed colonies were visible against a background of blue colonies which did not possess inserts. A control plate on which only cells had been plated revealed whether contamination of the plating cells with ligation product had taken place.

2.7.4 Screening transformants

Transformants were screened for the desired clone by bacterial colony hybridisation (Grunstein and Hogness 1975). After overnight incubation and refrigeration, Hybond N filter circles (Amersham) were placed on top of the colonies. Each filter and agar was marked with a sterile needle to ensure correct orientation of the colonies. After a minute, the filters were peeled off, labelled with pencil and placed colony side up on 3MM Whatman paper soaked with 0.5M NaOH, 1.5M NaCl for 7 minutes. The filters were neutralised by placing onto 3MM paper soaked with 0.5M Tris pH 7.5, 1.5M NaCl for 5 minutes and then placed onto another fresh 3MM paper containing the same solution for a further 3 minutes. The filters were finally rinsed by placing onto a 3MM whatman paper soaked in 2xSSC for 3 minutes and air-dried before baking at 80°C under vacuum for 2 hours.

Filters were hybridised with the probe of interest using the same solutions and conditions as for Southernns. The reference markers on the filters were marked on the resulting autoradiographs which enabled positively hybridising colonies to be identified.

2.8 cDNA and genomic clone isolation

2.8.1 Screening of gridded libraries

Cosmid, YAC and direct selection cDNA libraries existed as ordered arrays of clones on charged nylon filters. Hybridisation was carried out with the probe of interest using the
same conditions and solutions as for Southern hybridisation. The resulting autoradiographs were superimposed on top of their corresponding filters and the reference and grid markings transferred from the filter to the autoradiograph. This enabled the grid references for the positively-hybridising clones to be calculated. In the case of the direct selection cDNA libraries which were generated in the laboratory, positive clones could be directly picked from the respective microtitre wells in the library and streaked onto LB agar plates for single colonies. For cosmid and YAC libraries which were generated at the ICRF in the genome analysis laboratory, the locations of the desired clones were calculated from their grid references by the reference library database (RLDB). The cosmid and YAC clones were picked and streaked out onto LB/kanamycin and YPD agar plates respectively.

2.8.2 Screening of random plated bacteriophage clone cDNA libraries

Two placenta cDNA libraries were used in this project. One library, constructed in the λgt11 vector was used to isolate genes from the GP2B/EPB3 loci (see section 4.2), whilst a newer library, cloned into a λZAP phage vector (Stratagene) was used to isolate cDNA clones from the HDSPH locus (section 6.5). Both libraries were plated onto 23cm x 23cm square megaplates to a density of $10^5$ plaque-forming units (PFU)/plate. A total of 8 such plates were set up and duplicate lifts of each plate were taken onto large Hybond N+ squares (Amersham). Clear registration marks were made in the agar and filter for each plate, as described in section 2.7.4. In screening the library, the duplicate lifts from each plate were hybridised to the probe of interest using the same solutions and conditions as for Southern hybridisations (see 2.5.2 and 2.5.3). The registration marks from the filters were transferred to their corresponding autoradiograph and regions containing possible positively hybridising plaques identified. Genuine positive clones were identified if they were visible at the same place on each autoradiograph of the duplicate filters. When this occurred, the autoradiograph from the first lift was used to pick the region in the plate containing the putative positive clone. The autoradiograph was placed on a light box and the corresponding library plate was place on top of it, making sure to align the registration marks. A plug from the centre of the positive spot was picked from the plate using a sterile pasteur along with several more plugs from around the edges of the positive. The plugs were dropped into a glass bijoux containing 1ml SM and 2-3 drops of chloroform and the bijoux was then sealed with parafilm. After brief vortexing, the bijoux was placed on a roller at room temperature for 1-4 hours to allow the phage to elute into the SM buffer.
The second round of phage library screening was then carried out by making serial
dilutions of the phage stock and plating onto 9cm round LB plates as previously
described (2.1.3). Screening the phage lifts was performed in the same way as described
for bacterial colony screening (section 2.7.4). The process was repeated until a single
hybridising plaque could be picked.

2.8.3 Plasmid rescue from λZAP phage clones

The placenta cDNA library used to isolate clones from the *HDSPH* locus (section 6.5.1)
was constructed using the Uni-ZAP XR vector (Stratagene) which was designed to allow
*in vivo* excision and recircularisation of any cloned insert contained within the vector to
form a phagemid containing the cloned insert. This useful aspect of the vector removed
the requirement for subcloning the phage clone insert into a plasmid vector.

Once the final single positive plaque had been picked and dropped into a bijoux
containing SM+ chloroform, vortexed and incubated for 1-2 hours, the following
infections were set up in 15ml snap-cap falcon tubes using XL1-Blue MRF' cells
(Stratagene).

1. 20μl phage stock + 200μl of OD600nm=1.0 XL1-Blue cells.

2. 100μl phage stock + 200μl of OD600nm=1.0 XL1-Blue cells.

3. 200μl of OD600nm=1.0 XL1-Blue cells with no recombinant phage stock (control).

To each infection 1μl of R408 helper phage (>1x10⁶ pfu/ml) (Stratagene) was added and
the tubes incubated at 37°C for 15 minutes. After this time, 5ml of LB broth was added
and the tubes were incubated for 3 hours at 37°C with shaking. Following this
incubation, the tubes were placed into a 70°C water bath and incubated for a further 20
minutes before centrifuging for 5 minutes at 2.5K rpm. The supernatants were decanted
into sterile tubes and 100μl aliquots were combined with 100μl of OD600nm=1.0 XL1-
Blue cells. These mixtures were incubated at 37°C for a further 15 minutes and finally
plated onto LB/ampicillin agar plates in two dilutions:

1. 2μl of phagemid/XL1-B supernatant +100μl LB broth

2. 50μl of phagemid/XL1-B supernatant +100μl LB broth
Materials and methods

Colonies appeared on the plates containing the pBluescript double stranded phagemid with the cloned DNA insert. By plating effectively 4 dilutions from each original phage stock, plates containing single colonies were obtained. The bacteria infected with helper phage alone did not grow because they did not contain ampicillin resistance genes.

2.9 PCR analysis

2.9.1 General PCR

PCR was performed using an MJ thermocycler with hot bonnet or a Hybaid thermocycler. Cycling conditions were generally 1 minute denaturation at 94°C, 1 minute annealing at 55°C, 1 minute extension at 72°C for 35 cycles, followed by 5 minutes extension at 72°C for 1 cycle. For PCR using genomic DNA, a single 5 minute denaturation step was added before the 35 cycles to ensure complete separation of the DNA strands. The annealing temperature was adjusted to 5°C below the average Tm of the primers, which was calculated using the equation, Tm = 4 (G+C) +2 (T+A). The extension times were lengthened or shortened according to the expected size of the PCR product. A PCR was set up by mixing the following components.

(per reaction)
5μl 10x Taq buffer with MgCl₂ (Promega)
1unit Taq DNA polymerase (Promega)
2.5μl dNTP mix: dATP 2mM, dTTP 2mM, dCTP 2mM, dGTP 2mM
1μl each primer (from 50μg/ml stock)
stereile dH₂O to 50μl
2-3 drops PCR oil (for Hybaid thermocycler only)

Once these components had been mixed, 50-100ng of DNA template was added and the tube placed into the thermocycler for the required program. For multiple PCR reactions, the above quantities were scaled up accordingly and added to tubes containing the readily aliquoted DNA samples. Care was taken to avoid contamination of the component solutions with DNA. In some instances, this involved setting up the reaction in a tissue culture hood.

2.9.2 Subclone PCR analysis

This method enabled the analysis of colonies to see whether they contained the desired insert without having to grow overnight cultures and prepare DNA.
A single colony was picked with a sterile small Gilson tip which was touched onto an LB/ampicillin plate marked with a grid and subsequently placed into a tube containing 50µl of sterile distilled water. The tip was flicked to remove some cells into the water and the tube boiled for 5 minutes. Cell debris was pelleted by centrifuging the tube for 5 minutes at 13K rpm in a microcentrifuge and the supernatant was transferred to a clean tube. A 1µl aliquot of the supernatant was then used in a 10µl PCR reaction. The entire reaction mixture was then run out on an eithidium-stained agarose gel and visualised under UV light.

2.10 SSCP analysis

2.10.1 Preparation of PCR products

This was carried out on over 100 DNA samples at a time. Hence, an SSCP reaction mix was made and 50µl aliquots added to the PCR tubes containing the DNA samples. The reactions were carried out using the MJ thermocycler which does not require the use of PCR oil and can hold up to 96 samples.

SSCP mix (for 100 samples)

500µl Taq buffer with MgCl₂ (Promega)
30µl Taq DNA polymerase (Promega)
20µl dNTP mix (as for PCR, see 2.9)
20µl each primer (from 500µg/ml stock)
3900ml dH₂O
10µl α³²P dCTP (10mCi/ml) (Amersham)

A PCR reaction was carried out using the conditions outlined in 2.9 with the 5 minute denaturation step at the beginning of the program. After the reaction had completed, 1µl of each sample was aliquotted into sequencing microtitre wells containing 4µl of stop buffer. The microtitre dish was then placed on a heating block at 94°C for 10 minutes to denature and then quickly cooled on ice prior to loading onto the gel.

2.10.2 Acrylamide gel electrophoresis

The best resolution of SSCP products was achieved using MDE heteroduplex detection gels (J.T.Baker Inc, Phillipsburg NJ, USA). Glass sequencing plates of 20cmx40cm dimensions were used with 1.0mm spacers and well-forming combs. The gels were
Materials and methods

Poured using the conditions and solutions suggested by the manufacturer. Gels were run using a model S2 BRL sequencing apparatus at room temperature at 300V for 18 hours to give good resolution of 200-400bp PCR products.

Following electrophoresis, the apparatus was dismantled and the gel plates separated. The gel was carefully blotted onto dry 3MM Whatman paper, covered with Saran wrap and dried in a vacuum gel dryer for 20 minutes at 80°C. The dried gel was exposed to X-ray film for 5-12 hours at room temperature.

2.11 DNA sequencing

Sequencing reactions were performed on double-stranded templates using vector-derived or insert-specific primers using the dideoxy chain-terminator procedure (Sanger et al. 1977). The RNA-free DNA templates were prepared using the QIAGEN purification kits (see section 2.1.3). Both manual sequencing using the Sequenase Kit (United States Biochemicals, Cleveland, USA) and automated sequencing using PRISM ready reaction dyedideoxy terminator cycle sequencing kit (Applied Biosystems, Warrington, UK) were performed.

2.11.1 Manual sequencing

Alkaline denaturation

A 3ug aliquot of the purified plasmid DNA in a 50μl volume was denatured by adding a 0.1 volume of 2M NaOH, 2mM EDTA and incubating at room temperature for 10 minutes. The reaction was neutralised with 0.1 volumes of 3M sodium acetate (pH4.5-5.5) and precipitated with 2 volumes of cold ethanol. After incubation on dry ice for 10 minutes, the DNA was centrifuged for 15 minutes at 13K rpm in a microcentrifuge. The DNA pellet was washed with 70% ethanol, respun, air-dried and dissolved in 7μl H2O.

Annealing

Reaction buffer (2μl) (provided in Sequenase kit) and 5ng of primer were added to the denatured DNA. The mixture was placed in a glass beaker containing water at 65°C for 2 minutes and allowed to cool slowly over 15-30 minutes to below 35°C. After slow cooling the sample was placed on ice or refrigerated until required for the next stage.
Materials and methods

Labelling

The components of the labelling reaction with the exception of the DNA were mixed together and kept on ice.

Labelling reaction mix (1 reaction)

1μl DTT (0.1M)
2μl 1x labelling mix (supplied in kit as 5x and therefore diluted in distilled water to a working concentration)
0.5μl α³⁵S dCTP
2μl Sequenase polymerase (diluted 1/8 from stock solution)

The DNA was added and incubated at room temperature for 5 minutes.

Termination

After the 5 minutes incubation, a 3.5μl aliquot of the labelling reaction was added to each of 4 pre-warmed eppendorf tubes containing 2.5μl of one of the ddNTP termination mixes. After 5 minutes incubation at 37°C, the 4 reactions were terminated by the addition of 4μl stop solution. The samples were stored on ice at 4°C or at -20°C for longer periods prior to use. Shortly before loading onto the acrylamide gel, samples were denatured at 90°C for 2 minutes.

Polyacrylamide gels (6%) were cast between glass plates using the Sequagel (National Diagnostics) concentrate, buffer and diluent. The larger of the plates was treated with Sigmacote. The catalysts, 150μl of 20% ammonium persulphate and 60μl of TEMED (BioRad) were added to the acrylamide immediately before the gel was poured. The apparatus (Model S2, BRL) was pre-run in 1x TBE buffer at 85W until the temperature of the gel had reached 50-55°C. After the samples had been loaded, the gel was run at 75W for 4-8 hours depending on the extent of sequencing required.

2.11.2 Automated sequencing

Preparation of the DNA for sequence analysis was performed exactly according to the manufacturer's protocol (Applied Biosystems). After a cycle sequencing reaction using vector or insert primer, excess dyedexoxy terminators were removed by phenol/chloroform extraction. The samples were denatured by incubation at 90°C for 5 minutes prior to
electrophoresis and data collection using an ABI Model 373A automated sequencer and software (Applied Biosystems). Apparatus assembly and all other preparations were carried out according to the instruction manuals provided by the manufacturer (Applied Biosystems).

2.12 DNA sequence analysis

The sequencing data from the autoradiographs generated by manual sequencing was entered into computer files with the aid of a sequencing digitizer (DNASTar, London UK). Sequencing files generated through automated sequencing were edited and exported into a usable format with the aid of the SeqEd program (Applied Biosystems). Assembly of cDNA sequence contigs was performed using the Intelligenetics suite of programs under VMS on a VAX computer and subsequently using the GCG suite of programs running under UNIX OSF/1 on Digital alpha AXP computers. Sequence homology searches were carried out using TFASTA and FASTA from the GCG database searching program package (Lipman and Pearson 1985) using EMBL and GENbank databases.

2.13 Buffers and solutions

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBL agar Biological</td>
<td>10g trypticase peptone (Baltimore Laboratories 11921)</td>
</tr>
<tr>
<td></td>
<td>5g NaCl</td>
</tr>
<tr>
<td></td>
<td>10g agar (pH 7.2)</td>
</tr>
<tr>
<td>BBL top agarose</td>
<td>as above, but substitute the agar with 6.5g agarose per litre. Supplement</td>
</tr>
<tr>
<td></td>
<td>with 2g of MgSO_4\cdot 7H_2O.</td>
</tr>
<tr>
<td>Denaturing solution</td>
<td>25mM NaOH</td>
</tr>
<tr>
<td></td>
<td>1mM EDTA</td>
</tr>
<tr>
<td>Denhardt's solution (50x)</td>
<td>5g Ficoll</td>
</tr>
<tr>
<td></td>
<td>5g polyvinylpyrrolidine</td>
</tr>
<tr>
<td></td>
<td>5g bovine serum albumin</td>
</tr>
<tr>
<td></td>
<td>500ml dH_2O</td>
</tr>
</tbody>
</table>
Materials and methods

**LB broth (1L)**
- 10g bacto tryptone
- 5g yeast extract
- 10g NaCl (pH7.2)

**Loading buffer**
- 0.25% bromophenol blue
- 0.25% xylene cyanol
- 30% glycerol in dH2O

**Lysis buffer (section 2.1)**
- 4M guanidine thiocyanate
- 2% β-mercaptoethanol
- 0.5% sodium-N-lauroyl sarcosine
- 0.1M Tris pH7.2

**Oligolabelling buffer (OLB) (5 X)**
- 0.2M Hepes, pH6.6
- 100nM/ml of each of dGTP, dATP, dTTP
- 0.25M Tris, pH8.0
- 25mM MgCl2
- 0.35% 2-mercaptoethanol
- 25 units/ml random hexamers (Pharmacia)
  (store at -20°C until required)

**Novozyme (50x)**
- 80mg novozyme (Novo Biolabs)
- 40ml glycerol
- 160ml SCE
  (store at 4°C for 1 week only)

**PBSA (1L)**
- 10g NaCl
- 0.25g KCl
- 0.25g KH2PO4
- 1.43g Na2HPO4
- 1g MgCl2

**Prehybridisation solution (Southerns)**
- 5x SSPE
- 5x Denhardt's solution
- 0.5% SDS
  (solution is stored at 4°C and prewarmed to 65°C before use)
Materials and methods

Prehybridisation/hybridisation solution (Clontech northerns)

- 5x SSPE
- 10x Denhardt's solution
- 100μg/ml denatured sheared salmon sperm DNA
- 50% freshly deionised formamide
- 2% SDS
- (store at 4°C for a max of 7 days, prewarm to 50°C before use)

Proteinase K digestion buffer

- 500mM EDTA pH8.0 (Sigma)
- 1% sodium-N-lauroyl sarcosine
- 2mg/ml proteinase K (BDH)

SCE

- 1M sorbitol
- 100mM sodium citrate
- 10mMEDTA
- autoclave
- (+/- 10mM DTT)

SM buffer (1L)

- 5.8g NaCl
- 2.0g MgSO₄·7H₂O
- 50ml 1M Tris.pH7.5
- 10ml 2% gelatin (0.2g in 10ml H₂O, boiled to dissolve

Solution I

- 50mM glucose
- 25mMTris.HCl (pH8.0)
- 10mMEDTA (pH8.0)

Solution II

- 0.2M NaOH
- 1% SDS
- (make fresh)

Solution III

- 5M potassium acetate (pH4.8)
- 60mls potassium acetate
- 11.5mls glacial acetic acid
- 28.5mls H₂O
- (store at 4°C)
Materials and methods

20xSSC (1L)  
175.3g NaCl  
88.2g sodium citrate  
(pH7.0)

Stop buffer  
95% formamide  
20mM EDTA  
0.05% bromophenol blue  
0.05% xylene cyanol

TAE electrophoresis buffer (1x)  
40mM Tris-acetate  
1mM EDTA

TBE electrophoresis buffer (0.5)  
45mM Tris-borate  
1mMEDTA

TE  
10mM Tris.HCl pH7.6  
1mM EDTA

TENS buffer (100mls)  
1ml 100x TE  
1ml 10M NaOH  
2.5ml 20% SDS  
(final concentration, 0.1M NaOH;0.5%SDS)

Transformation and storage buffer  
10% PEG (3350)  
5% DMSO  
20mM Mg (10mM MgSO4 + 10mMgCl2)  
made up in LB broth  
(store at -20°C until needed)

Yeast lysis buffer (200mls)  
1% lithium dodecyl sulphate  
100mM EDTA  
10mM Tris. HCl  
(filter sterilise)

YPD (1L)  
10g yeast extract  
20g peptone  
(for plates)

(20g agar  
autoclave, then add glucose to 2%)
CHAPTER 3 PHYSICAL MAPPING OF THE BRCA1 REGION

Initial genetic mapping of BRCA1 suggested the locus resided within a 4cM region flanked by polymorphic markers at 
THRA1 and D17S579 (Easton et al. 1993; Goldgar et al. 1993). The first aim of this thesis was to construct a physical map of this region to convert the genetic distances measured in centimorgans into physical distances measured in kilobase pairs and identify useful signposts such as polymorphic DNA stretches, CpG islands and the exact location of genes and markers. This information would help to orientate genomic clone walking efforts and serve as a useful reference so that as each new clone was isolated its location in the BRCA1 region could be verified.

The basis of the physical map was provided by the initial assessment of clones and genes on the chromosome 17 somatic cell hybrid mapping panel developed in the laboratory (Table 3). The new clones and DNA markers developed during the course of this work were ordered with respect to previously mapped genes and polymorphic markers by Southern hybridisation of the relevant repeat free fragments to the somatic cell hybrid mapping panel, as shown in Fig. 3.1.2.

Once the approximate order of the markers along the chromosome was established in this way, regional restriction maps which encompassed several neighbouring markers were developed throughout the region with the aid of pulsed field gel electrophoresis (PFGE). Southern filters were generated from pulsed-field gels containing high molecular weight genomic DNA from an immortalised B cell line digested with a number of rare cutter restriction endonucleases in single and double combinations. These filters were sequentially hybridised with single copy probes throughout the BRCA1 region and restriction maps constructed from the resulting autoradiographs.

At the most proximal end of the BRCA1 region, a map was developed around the RARA locus including the location of TOP2 and IGFBP-4 (see section 3.1). This work is to be included in a paper concerning the IGFBP-4 gene and its mutational analysis in breast tumour DNA (Nicolai et al in preparation). Subsequent genetic studies by other groups identified sites of recombination which indicated BRCA1 was situated below D17S776 and D17S857 in affected family members which effectively ruled out this region as the probable site of the BRCA1 gene (see Fig. 1.5.1(i)) (Goldgar et al. 1993; Kelsell et al. 1993).
Table 3. Mapping results of probes in the \textit{BRCA1} region on a panel of chromosome 17 CMGTs and translocation hybrids.  
All probes and markers included here were analysed by Southern hybridisation or PCR across the mapping panel. The presence or absence of each probe was noted and using the data of the chromosomal fragments present in each hybrid, the position of each probe along the chromosome could be estimated. +, probe present in hybrid; -, probe absent in hybrid.
Further distal along the chromosome, another long-range physical map including $EDH$, D17S855, exon-trap products and anonymous DNA fragments was developed (see section 3.2). This work is included in a paper concerning the physical map and transcriptional analysis of the region between $EDH$ and $1A1.3B$ which includes $BRCA1$ (Brown et al. in press). A map including $1A1.3B$, $RNU2$ and markers D17S858 and D17S859 was constructed distal to this (see section 3.3). This work is published in a paper detailing the physical mapping and transcriptional analysis of the $BRCA1$ region between $1A1.3B$ and D17S78 (Jones et al. 1994). The accumulating family data with notable absence of recombination events within this region gave strong indication that this was the likely location of $BRCA1$. Hence subsequent genomic cloning, gene isolation and analysis, carried out as part of this thesis work (chapters 5, 6 and 7) and also by other members of the laboratory concentrated on this region.

Two further physical maps around $PPY$ and $EPB3$ were also constructed (see section 3.4) which will be included in a paper concerning the physical mapping of this region (Jones et al. in preparation). With the isolation of clones between these regions and through observations of shared partial restriction fragments, physical maps around $EDH$, $RNU2$, $PPY$ and $EPB3$ were joined together to create a detailed long-range restriction map spanning over 3 megabase pairs (see Fig. 3.5). The region distal to PPY/D17S78 was excluded from further studies following the publication of convincing recombinations which placed the $BRCA1$ gene proximal to D17S78 (Simard et al. 1993).

The physical maps were confirmed by analyses of genomic clones in the region and comparisons to the data from ongoing genetic mapping efforts.

### 3.1 Construction of a physical map at the $RARA$ locus

Initial mapping with the somatic cell hybrid mapping panel suggested the proximity of $RARA$ and $TOP2$. A long-range restriction map around the $RARA$ locus was constructed using probes generated during this thesis work which included PCR generated probes from the coding regions of the $RARA$ and $TOP2$ genes and a linking clone which was isolated from a previously constructed chromosome 17 specific $NolI$ linking clone library (Borrow et al. 1991).

#### 3.1.1 Isolation of linking clones from the $RARA$ region

Two linking clones, 2H10 and 2A3 from the chromosome 17 $NolI$ linking library hybridised to radioactively-labelled YAC $Alu$-PCR products from a $RARA$ - and $TOP2$-
positive YAC, 199C3 (Washington University library, (Burke et al. 1987; Brownstein et al. 1989) (Fig.3.1.1). When a fragment from one of the linking clones, 2H10 was hybridised back to the linking clone library, both 2H10 and 2A3 hybridised, indicating that both linking clones were from the same NotI site. Hence only one clone, 2H10 was used in further analyses.

3.1.2 Localisation of RARA, TOP2 and linking clone 2H10 using the somatic cell hybrid mapping panel

The location of RARA and TOP2 distal to THRAl on chromosome 17q was suggested by PCR analysis on the somatic cell hybrid mapping panel (Table 3). When a repeat free fragment from linking clone 2H10 was hybridised to the same panel, hybridisation results suggested it came from the same region (Fig. 3.1.2). It hybridised to the chromosome 17-only hybrid PCTBA1.8, the chromosome 17q-only hybrid TRID62, the hybrid containing the t(1;17) NF1 translocation, the hybrid containing the 15q+ t(15;17) chromosome from an acute promyelocytic leukaemia patient and a chromosome mediated gene transfer (CMGT) hybrid, KLT8. The same linking clone failed to hybridise to a hybrid containing the derivative 17q- t(15;17) chromosome from an acute promyelocytic leukaemia patient, the CMGT hybrids PLT6B and PLT8 and the control mouse cell line, EL4. This placed it distal to the APL breakpoint but close to both RARA and TOP2.

3.1.3 Construction of a long range restriction map by PFGE analysis

A detailed PFGE map at RARA using a linking clone 4A3 and cosmids 121 and 124 had previously been constructed (Borrow et al. 1990). It was possible to extend this map towards the telomere using probes from both sides (A and B) of the NotI site from the linking clone 2H10.

The precise location of linking clone 2H10 on this previously constructed map was clearly deducible from the results of Southern hybridisation of 2H10/A to filters containing digested DNA samples from cosmids 121 and 124 (Fig.3.1.3(i)). The newly generated 2H10/A probe hybridised to the most distally extending RARA cosmid, 121 and the restriction fragment sizes of the hybridising bands indicated the location of 2H10 just distal to the previously isolated 4A3 linking clone. Pulsed field gel electrophoresis analysis carried out at the beginning of this thesis indicated probes 2H10/A and 4A3/B hybridised to the same NotI and MluI fragments, indicating the location of linking clone 2H10 at the next CpG island distal to 4A3 (Fig. 3.1.3(ii)). Probe 2H10/B from the other side of the NotI site in linking clone 2H10, hybridised to a larger 980kb NotI fragment. A
Fig. 3.1.1 Isolation of linking clones from the RARA gene region. YAC199C3 (a gift from F. Collins, Washington) was used in an Alu-PCR reaction using primers TC-65 and 517 (Nelson et al 1989). The radioactively-labelled products hybridised to two NotI linking clones, 2H10 and 2A3, as shown by the arrows.

Fig. 3.1.2 Localisation of 2H10 on the somatic cell hybrid mapping panel. A repeat free fragment from one side of the NotI site in 2H10 was hybridised to a Southern filter containing digested DNA from a selection of hybrid cell lines from the mapping panel. By scoring the presence or absence of the probe in the hybrids the chromosomal position of the probe could be calculated (see table 3).
Fig. 3.1.3(i) **Localisation of linking clone 2H10.** Probe 2H10/A, from linking clone 2H10 was hybridised to a Southern filter containing digested DNA from two RARA cosmid, 121 and 124 (J. Borrow PhD thesis 1992). The 2H10/A probe hybridises to the end of cosmid 121, telomeric to RARA.
**PFGE Southern hybridisation results**

<table>
<thead>
<tr>
<th>Probe</th>
<th>NotI Fragment sizes (Kb)</th>
<th>MluI Fragment sizes (Kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4A3/A</td>
<td>(190), 170</td>
<td>(260), 100</td>
</tr>
<tr>
<td>4A3/B</td>
<td>(190), 20</td>
<td>(260), 100, (540)</td>
</tr>
<tr>
<td>2H10/A</td>
<td>(190), 20</td>
<td>(260), 100, (540), 220</td>
</tr>
<tr>
<td>2H10/B</td>
<td>(1.2MB), 980</td>
<td>220, (540), (970)</td>
</tr>
<tr>
<td>IGFBP-4</td>
<td>(1.2MB), 980</td>
<td>220, (540), (970)</td>
</tr>
<tr>
<td>TOP2</td>
<td>(1.2MB), 980</td>
<td>220, (540), (970)</td>
</tr>
</tbody>
</table>

( ) = partial fragment produced by methylation of a restriction site

---

**Fig.3.1.3(ii) Construction of a regional restriction map distal to RARA.** Hybridisation data from PFGE Southern analysis (see table) was used to construct a long-range restriction map incorporating probes from linking clones 4A3 and 2H10, TOP2 and IGFBP-4 genes. The location of the RARA gene, discovered from work carried out previously in the laboratory (Borrow et al 1991) is also indicated.
probe from the TOP2 gene hybridised to the same NotI and MluI fragments as 2H10/B, placing it distal to RARA.

3.1.4 Localisation of the insulin-like growth factor binding protein gene, IGFBP-4

Following the localisation of the IGFBP-4 gene to chromosome 17q21.1 by genetic mapping (Bajalica et al 1992), information on the precise physical location of this gene was required. Analysis of the chromosome 17q somatic cell hybrid mapping panel suggested it was located close to RARA and TOP2 (Table 3) (Black et al. 1993).

Pulsed field gel electrophoresis analysis revealed the IGFBP-4 gene hybridised to the same NotI and MluI fragments as 2H10/B and TOP2 (see Fig. 3.1.3(ii)). Further hybridisation analysis of PFGE blots of BssHII digested DNA with probes 2H10/B, TOP2 and IGFBP-4 showed that 2H10/B and TOP2 reside on the same BssHII fragment at 150kb, whilst IGFBP-4 resides on a 800kb BssHII fragment (Fig. 3.1.4). This places IGFBP-4 distal to TOP2.

3.1.5 Comparison to YAC data

Work by other members of the laboratory involved in the analysis of YACs isolated in the region appeared to confirm the PFGE map. The 199C3 YAC used to isolate 2H10 appeared positive for TOP2 and RARA, but was negative for IGFBP-4 indicating the order \( \text{CEN-RARA-TOP2-IGFBP-4-TEL} \) (M. Bonjardim data not shown).

3.2 Construction of a physical map around EDH and D17S855

3.2.1 Initial ordering of clones

Initial analysis of this region carried out previously in the laboratory using the somatic cell hybrid mapping panel was inconclusive due to the absence of hybrids harbouring breakpoints in this region of the chromosome. It was further complicated by the discovery of a small hole in hybrids KLT12 and KLT8 which had not been accounted for and which lead to the wrong assignment of the EDH gene to below D17S579 (Black et al. 1993). Other mapping efforts suggested the ordering of markers in this region as CEN-RARA-TOP2-D17S857-GAS-EDH-D17S855-PPY-D17S78-D17S579 (Abel et al. 1993; Flejter et al. 1993; Kelsell et al. 1993; Simard et al. 1993), (see Fig. 1.5.1(i)). Hence it became necessary to carry out a more detailed physical analysis to determine the exact location of EDH and its neighbouring markers and to further characterise this region.
Fig. 3.1.4 Pulsed field gel electrophoresis analysis of IGFBP-4. The autoradiographs in panels A, B and C indicate that the probes 2H10/B, TOP2 and IGFBP-4 all hybridised to the same large 980kb NotI fragment and to the same 300kb MluI fragment. Probes 2H10/B and TOP2 hybridised to a 150kb BssHII fragment, whilst IGFBP-4 hybridised to a much larger 800kb BssHII fragment, placing it distal to 2H10/B and TOP2 in the position shown.

N, NotI; M, MluI; Bs, BssHII
3.2.2 PFGE analysis of the EDH/D17S855 region

Initial PFGE analysis using a PCR-generated probe from the EDH I gene and a repeat free fragment from the D17S855 locus, 855RF indicated that they both resided on the same 550kb NotI fragment. This enabled a restriction map incorporating these probes to be generated (Fig. 3.2.2, maps A and B). Ongoing genomic clone efforts carried out by other members of the laboratory generated new DNA probes in the region which were also analysed by PFGE and helped to add detail to the resulting long-range restriction map (Fig. 3.2.2 maps C and D). The probes EDH, 855RF and a repeat free fragment from a PAC from the region, RF18 (probe C) all hybridised to the same 550kb NotI fragment. In addition, RF18 mapped to the same NruI and NruI/NotI fragments as 855RF, but to different EagI and BssHII fragments (autoradiograph not shown). An exon-trap product, ET-A37 hybridised to the same 480kb NruI fragment as RF18 (and therefore 855RF), as indicated by the arrowed bands in autoradiographs C and D, but it hybridised to a larger NotI fragment of 750kb and to different MluI, EagI and BssHII fragments. Taken together, these results suggest that EDH and RF855 reside on the same 550kb NotI fragment and that the ET-A37 probe resides on the next distal NotI fragment. Another exon-trap product, ET-A38 revealed a hybridisation pattern identical to that of RF855 (autoradiograph not shown).

The long-range restriction maps generated at each locus were superimposed at the regions of shared restriction fragments resulting in a detailed long-range restriction map depicted in A+B+C+D, Fig 3.2.2. Following the suggestion by Albertsen and colleagues that the Ki antigen gene was located in this region, PFGE analysis was carried out using a PCR-generated probe from this gene and confirmed its location between EDH and D17S855 (Fig. 3.2.2 map A+B+C+D).

3.3 Construction of a long range restriction map from RNU2 to D17S858

Genetic mapping studies and FISH analysis suggested RNU2 and two markers, D17S858 and D17S859 were situated distal to EDH and proximal to PPY/D17S78 (refer to Fig. 1.5.1.(i)) (Anderson et al. 1993; Flejter et al. 1993). Analysis by PFGE in this thesis provided further information regarding the characteristics of this region by pinpointing the exact location of RNU2 and indicating the locations of markers D17S858/D17S859 and the 1A1.3B gene (see Fig. 3.3.1). The discovery that both 1A1.3B and ET-A37 (see section 3.2.2) resided on the same 750kb NotI and 220kb EagI fragments enabled the restriction maps around EDH/D17S855 and RNU2/D17S858 to be joined together to give a detailed long-range restriction map covering over 2.5 megabase pairs (see Fig. 3.5).
Fig. 3.2.2 Construction of a long-range restriction map around EDH/D17S855 by PFGE analysis. Autoradiographs A and B resulted from the sequential hybridization of one filter whilst C and D were produced by sequential hybridization of a second filter. Probe A was a PCR product from the 3' untranslated region of EDH, probe B was a repeat-free fragment from a cosmid containing marker D17S855 (855RF), probe C was a repeat-free fragment from PAC 22157 (RF18) and probe D was ET-A37. Size markers corresponding to each of the two PFGE filters used are indicated to the left of each pair of autoradiographs. Individual restriction maps A, B, C and D were constructed using the data provided by each of the corresponding autoradiographs. The maps were superimposed at the regions of the shared restriction fragments (A+B+C+D). Subsequent PFGE Southern analysis indicated the location of the KiAg gene (Albertsen et al 1994) and ET A38. N, Nol; M, MluI; Bs, BssHII; Ea, Eal; Nr, NruI.
Physical mapping

Analysis of PAC clones in the region by other members of the laboratory confirmed the ordering provided by this PFGE analysis. In addition, construction of a cosmid contig from \textit{RNU2} to \textit{D17S78} as part of this thesis (see chapter 5) confirmed the order of markers and genes and the distances between them suggested by the PFGE analysis.

3.3.1 Comparison of PFGE data from \textit{IA1.3B} and \textit{RNU2} genes and the construction of a restriction map

PFGE analysis revealed that \textit{IA1.3B}, \textit{RNU2} and 12E/Bd, a fragment from a D17S858/D17S859-positive cosmid, G05149 reside on the same 750kb \textit{NotI} fragment and that \textit{IA1.3B} and \textit{RNU2} share the same 380kb \textit{MluI} fragment but different \textit{NruI} fragments (Fig. 3.3.1). The \textit{NruI} fragment which contains \textit{RNU2} also hybridises to 12EB/d, giving the order \textit{IA1.3B}-\textit{RNU2}-D17S858/D17S859 (12EB/d).

3.3.2 Joining long range restriction maps at \textit{EDH/D17S855} and \textit{IA1.3B/RNU2} using an interstitial exon trap product

Similarities between the long-range restriction maps around \textit{EDH/D17S855} and \textit{RNU2} became apparent with the observation that the \textit{NotI} fragment to which probe ET-A37, \textit{IA1.3B} and \textit{RNU2} hybridised was the same length (750kb). On closer inspection it was apparent that the 380kb \textit{MluI} and the 220kb \textit{EagI} fragments to which the ET A37 probe hybridised were the same size as those hybridising to \textit{IA1.3B}. Confirmation that the fragments hybridising to \textit{IA1.3B} and ET A37 were the same was achieved by sequentially hybridising each probe to PFGE filters containing rare-cutter digested DNA samples from lymphoblastoid cell lines of different individuals. Hybridisation of either the \textit{IA1.3B} or the ET A37 probe to PFGE filters containing \textit{MluI} digested DNA samples or to PFGE filters containing \textit{NotI} digested samples produced the same results: a characteristic polymorphic pattern of hybridisation (Fig. 3.3.2). Comparing the autoradiographs from the \textit{IA1.3B} and ET A37 hybridisations revealed that the same bands hybridised to both probes, confirming their close proximity. Taking into account the observation that \textit{IA1.3B} and ET A37 appeared to reside on the same restriction fragments, the restriction maps around \textit{EDH} and \textit{RNU2} could be overlapped producing a long-range restriction map stretching over 2.5 megabase pairs from \textit{EDH} to D17S858/D17S859 (Fig. 3.3.2 map A+B).
Fig. 3.3.1 Construction of a physical map from PFGE data using probes *RNU2*, *1A1.3B* and *12E/Bd*, a fragment from the D17S858/D17S859-positive cosmid, G05149. Each probe was hybridised to the same PFGE filter resulting in autoradiographs 1, 2 and 3. Between each hybridisation the blot was stripped and checked by autoradiography for any residual signal. Arrowed bands indicate the restriction fragments which hybridise to more that one probe. Restriction maps were constructed using the information provided by each hybridisation experiment. Superimposing the maps (1+2+3) gives a rough estimation of the distance between 12E/Bd and RNU2 as 300kb. N, *NcoI*; M, *MluI*; Bs, *BssHII*; Ea, *EagI*; Nr, *NruI*; (M) and (Ea), restriction sites showing partial methylation.
**Fig. 3.3.2 Joining long range restriction maps at EDH/D17S855 and 1A1.3B/RNU2 using an interstitial exon trap product.** The long-range restriction maps generated at EDH/D17S855 (map A) and RNU2/D17S858 (map B) were superimposed following the observation that probes ET A37 and 1A1.3B hybridised to the same 220 kb EagI, polymorphic NotI and polymorphic MluI fragments (map A+B). The autoradiographs shown here are the results from sequential hybridisations of each probe ET A37 and 1A1.3B to the same PFGE filter containing MluI digested DNA samples from lymphoblastoid cell lines of different individuals. Identical polymorphic patterns of hybridisation were observed with each probe, confirming they both hybridise to the same fragment.
3.3.3 Comparison of the PFGE-generated physical map to PACs in the region

The location of genes and DNA markers around \textit{EDH} suggested by the PFGE analysis was confirmed by analysis, carried out by other members of the laboratory, of PAC clones isolated from the region (Brown \textit{et al.} in press). One PAC (103014), from which exon-trap products ET-A37 and ET-A38 were produced, hybridised to a promoter probe from the \textit{1A1.3B} gene and to 855RF, confirming the proximity of these two markers.

3.4 Physical mapping of the region \textit{PPY} to \textit{GP2B}.

This distal part of the \textit{BRCA1} region was characterised initially by high density genetic mapping and radiation hybrid analysis (Abel \textit{et al.} 1993; Anderson \textit{et al.} 1993) to give a rough chromosomal order of genes and polymorphic markers. They suggested the order of genes and markers in the region as CEN-\textit{EDH}-D17S855-D17S859-D17S858-[\textit{PPY}-D17S78]-D17S183-EPB3-GP2B-D17S579-TEL (Anderson \textit{et al.} 1993). Initial mapping with somatic cell hybrid panel generated in this laboratory wrongly placed D17S183 proximal to D17S78 which in turn was assigned to a location distal to D17S579 (Black \textit{et al.} 1993). These discrepancies were found to be due to the presence of an extra chromosomal fragment in hybrid KLT12 which had not been taken into account.

Analysis by PFGE in this project gave further definition to the map of the region and resolved discrepancies in the results of previous mapping efforts by providing the exact location of the \textit{PPY},EPB3 and \textit{GP2B} genes as well as the location of the D17S78 and D17S183 polymorphic markers used in genetic mapping studies. In addition, the PFGE analysis shown here highlighted the probable location of previously unmapped genes through the identification of CpG islands.
3.4.1 Single copy DNA probes generated in the region

Probes used in PFGE analysis of the region are detailed below (Table 3.4.1).

<table>
<thead>
<tr>
<th>PROBE NAME</th>
<th>PROBE DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPY</td>
<td>PCR generated probe from 3' end of published sequence (accession number M111726)</td>
</tr>
<tr>
<td>p131</td>
<td>single copy probe which recognises the MspI polymorphism at D17S78 (Luty et al. 1988)</td>
</tr>
<tr>
<td>EPB3</td>
<td>PCR generated probe from 3' end of published sequence</td>
</tr>
<tr>
<td>GP2B</td>
<td>PCR generated probe from 3' end of published sequence</td>
</tr>
<tr>
<td>4A5/A</td>
<td>Probe from one side of the NotI site in linking clone 4A5 (Borrow et al. 1990)</td>
</tr>
<tr>
<td>4A5/B</td>
<td>Probe from other side of the NotI site in linking clone 4A5 (Borrow et al. 1990)</td>
</tr>
<tr>
<td>B43</td>
<td>Probe from cosmid SCG43 containing the polymorphic marker D17S183 (HGM 10, A2491)</td>
</tr>
</tbody>
</table>

Table 3.4.1. Details of single copy probes from the PPY to GP2B region used in PFGE Southern analysis.
3.4.2 Results of PFGE Southern hybridisations.

Each probe was hybridised in turn to one PFGE Southern filter containing digested genomic DNA from a B cell line. This allowed direct comparisons of the resulting autoradiographs and identified occurrences of different probes hybridising to the same genomic restriction fragments. The hybridisation data obtained from the analysis of the resulting autoradiographs (not shown) is summarised in table 3.4.2 below.

<table>
<thead>
<tr>
<th>PROBE</th>
<th>FRAGMENT SIZE (Kb)</th>
<th>NotI</th>
<th>MluI</th>
<th>NruI</th>
<th>NotI+MluI</th>
<th>NotI+NruI</th>
<th>MluI+NruI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPY</td>
<td>480 (400)</td>
<td>450</td>
<td>(400)</td>
<td>(300)</td>
<td>450 (400)</td>
<td>(300)</td>
<td></td>
</tr>
<tr>
<td>p131</td>
<td>480 (300)</td>
<td>450</td>
<td>(400)</td>
<td>(160)</td>
<td>450 (400)</td>
<td>(160)</td>
<td></td>
</tr>
<tr>
<td>4A5/B</td>
<td>480 (300)</td>
<td>250</td>
<td>(400)</td>
<td></td>
<td>unclear</td>
<td>(v. small)</td>
<td>unclear</td>
</tr>
<tr>
<td>4A5/A</td>
<td>150 (560)</td>
<td>250</td>
<td>(560)</td>
<td></td>
<td>150 (560)</td>
<td></td>
<td>250</td>
</tr>
<tr>
<td>EPB3</td>
<td>150 (400)</td>
<td>250</td>
<td>(400)</td>
<td></td>
<td>150 (400)</td>
<td></td>
<td>250</td>
</tr>
<tr>
<td>GP2B</td>
<td>~30 (560)</td>
<td>250</td>
<td>~30</td>
<td></td>
<td>~30 (560)</td>
<td></td>
<td>250</td>
</tr>
<tr>
<td>B43</td>
<td>~10 (560)</td>
<td>250</td>
<td>~10</td>
<td></td>
<td>~10 (560)</td>
<td></td>
<td>250</td>
</tr>
</tbody>
</table>

Table 3.4.2 Results of PFGE Southern hybridisation analysis using probes from the PPY to GP2B region. (), partial restriction fragments as a result of methylation of restriction sites.

Both PPY and p131 hybridised to the same 480 kb NotI and 450 kb NruI fragments. The linking clone probe 4A5/B hybridised to the same 480 kb NotI fragment but to a smaller 250 kb NruI fragment. The adjacent clone 4A5/A, from the other side of the NotI site in this linking clone, hybridised to the same 250 kb NruI fragment as 4A5/B, but to a smaller 150 kb NotI fragment. Since these linking clones have been constructed so that they span NotI sites, these hybridisation results suggest the 480 kb NotI and 150 kb NotI fragments are adjacent to one another on the chromosome. Probe EPB3 gave identical hybridisation results to 4A5/B, suggesting it lies in close proximity to 4A5/B on the next adjacent NotI fragment to PPY and p131. The probes GP2B and B43 reside on the same
Physical mapping

250kb NruI fragment as EPB3 and linking clone probes 4A5/A and B. Both GP2B and B43 hybridised to different NotI fragments of around 30 kb and 10 kb respectively.

Throughout the region all the probes appeared to detect more than one MluI fragment. This is probably due to a high degree of methylation of the MluI sites which predominantly reside in the chromosomal regions between CpG islands. The methylation has made these sites resistant to restriction cleavage with the result of many partial restriction fragments being produced. As a result, precise localisation of the MluI sites was not possible with the results given in table 3.4.2.

3.4.3 Construction of a long-range restriction map of the region

The hybridisation results above were used to construct a long-range restriction map of the region (Fig. 3.4.3). Although the results suggested that B43 and GP2B reside on the same 250kb NruI fragment as the EPB3 and 4A5 probes, but on different NotI fragments, their exact positions on the map could not be determined from these hybridisation results alone. The most likely location of GP2B is indicated and takes into account its hybridisation to a NotI fragment distinct to the other probes in the region and results of radiation hybrid and genetic mapping which suggest it resides distal to D17S183 and EPB3.

Southern hybridisation and restriction analysis of cosmid SCG43, from which probe B43 was obtained, revealed a clustering of NotI sites in this region and hybridisation to 4A5/A, B43 and 4A5/B (see chapter 4). This suggests the likely location of the B43 probe proximal to EPB3 between 4A5/A and B on a small 10 kb NotI fragment as shown (Fig. 3.4.3). The 150 kb and 480 kb NotI fragments to which probes 4A5/A and 4A5/B hybridise are likely to be situated either side of this small fragment. In the construction of the linking clone from this region, restriction digestion with NotI most likely resulted in the excision of the small NotI fragment prior to cloning into the phage linking clone vector.
Fig. 3.4.3 Long-range restriction map of the region PPY-GP2B. Results from Southern hybridisations with probes from the region (table 3.4.3) were used to construct the map. All MluI sites appear to be methylated in this region, as indicated by brackets. N, NotI; M, MluI; Nr, NruI.
3.5 Summary of physical mapping analysis of the BRCA1 region

The physical mapping of the BRCA1 region carried out in this project is summarised in Fig. 3.5. Although restriction site polymorphisms and differences in methylation could exist between the B cell line used in the PFGE analysis here and the cell lines used to construct the genomic clone libraries, the distances between markers and genes suggested by this analysis appear to agree with genomic clone contig analysis carried out in this project (chapter 5), by other members of the laboratory (M. Bonjardim unpublished; Brown et al. in press) and by other members of the consortium (Albertsen et al. 1994; Neuhausen et al. 1994).

Two gaps in the final long-range restriction map remain at either side of the EDH-RNU2-D17S859 restriction map. One, distal to the RARA-IGFBP-4 region, was due to the absence of restriction fragments hybridising to probes from this region in addition to probes from the EDH region. It is likely that the distance between IGFBP-4 and EDH is several hundred kilobase pairs, as suggested by a YAC and P1 contig including this region (Albertsen et al. 1994). A further restriction map including genes and markers between these two genes would be required to complete the physical map in this area.

The second gap in the map, between EDH-RNU2- and PPY/GP2B, could not be resolved from the PFGE data alone. The occurrence of MluI, NotI, NruI, BssHII and EagI sites between PPY and D17S858/D17S859 in the cell line DNA used in this study resulted in the absence of restriction fragments hybridising to probes from both these regions. A linking clone spanning the NotI site which could help to link up these two regions could not be isolated. However, a cosmid contig was constructed in this work linking D17S78 to D17S858/D17S859 (chapter 5). It suggested the distance between these two regions was less than 100 kb.

The physical analysis of the region has identified the precise localisation of genes and markers throughout the region and highlighted the likely location of unmapped and unidentified genes through the identification of CpG islands. It has aided genomic cloning efforts by orientating genomic clone contigs and giving rough estimates of the distances required to complete the coverage of the BRCA1 region. By referring newly isolated clones back to the restriction map of the region by PFGE analysis, clones not mapping to the BRCA1 region have been quickly identified and as a result excluded from further study.
Fig. 3.5 Physical map of the \textit{BRCA1} region on chromosome 17 (drawn to scale).

The map was generated by pulsed-field gel electrophoresis analysis of the genomic DNA from a lymphoblastoid cell line. Genetic distance, measured in centimorgans, has been converted into physical distance, measured in kilobase pairs. The putative locations of CpG islands suggested by clustered sites for rare-cutter endonucleases are shown. The PFGE data alone could not link the \textit{PPI} long-range restriction map to that of \textit{EDH} to D17S858, as indicated by the dotted line between these two regions. However, a cosmid contig constructed between D17S76 and D17S858 (see chapter 5) has indicated \textit{PPI} to be roughly 60kb distal to D17S858 in the position shown. The positions of \textit{Eagl} and \textit{BstHI} sites around \textit{KARA} and \textit{PPI} were not clear from PFGE studies and have been omitted from the map. N, \textit{NcoI}; M, \textit{MluI}; Bs, \textit{BstHI}; Ea, \textit{Eagl}. 
CHAPTER 4 INVESTIGATION OF FEATURES OF THE PHYSICAL MAP AROUND \textit{BRCA1} IDENTIFIED BY PFGE ANALYSIS

During the construction of the PFGE physical map (chapter 3), a number of interesting features of the \textit{BRCA1} region were discovered. Numerous putative CpG islands were identified by chromosomal DNA analysis (Fig. 3.5) which indicated several unmapped and possibly novel genes were situated throughout the region. Most of the CpG islands identified to date have been clearly associated with genes, particularly where transcription begins (Bird 1986; Bird 1987). The work described in this chapter carried out early on in the project concerns the identification of the previously unmapped human ribosomal RNA transcription upstream binding factor gene (\textit{hUBF}) at one such putative CpG island.

The second section of this chapter investigates another feature of the \textit{BRCA1} region, the tandem repeat array at the \textit{RNU2} locus. The \textit{RNU2} locus encodes human U2 RNA, an abundant small nuclear RNA (snRNA) in the nucleus of all eukaryotic cells. This snRNA as well as snRNAs U1, U4 and U5 exist as components of small nuclear ribonucleoprotein particals (snRNPs) which together with other associated protein subunits and an additional group of non-snRNP proteins, make up the spliceosome. This large RNA-protein complex catalyses the removal of introns from nuclear pre-mRNA (reviewed in (Lamond 1993; Hodges and Beggs 1994)).

Human U2 genes are located on a 6 kb DNA fragment which is tandemly repeated around 10-20 times per haploid genome (Van Arsdell and Weiner 1984). Such tandem repetition allows high levels of gene expression and is a feature of other genes such as large ribosomal RNAs and histones which require more than one gene copy to meet the needs of the cell. Investigation of the \textit{RNU2} locus carried out in this chapter revealed polymorphic patterns of hybridisation were observed when the \textit{RNU2} probe and probes nearby were used in the PFGE Southern analysis of rare-cutter digested genomic DNA samples from several different lymphoblastoid cell lines (Fig. 4.2.2). Further investigation on the extent and nature of the variation at this locus within genomic DNA samples from a number of different lymphoblastoid cell lines was also investigated (section 4.2.2).

4.1 Analysis of a CpG island at D17S183.

The polymorphic marker D17S183, isolated in our laboratory from a chromosome 17-specific cosmid SCG43 (M Yagle HGM 10, A2491), was found to reside at the distal end of the \textit{BRCA1} region by physical mapping studies carried out in this project (Fig. 3.4.3
and Fig. 3.5) and by genetic mapping studies of other groups (Anderson et al. 1993; Bowcock et al. 1993). PFGE analysis of genomic DNA in this region revealed a putative gene-associated CpG island at this locus (4.1.1). Cosmid SCG43 was therefore used to screen a cDNA library in order to isolate the genes thought to reside in the region.

4.1.1 Identification of a CpG island cluster

PFGE analysis of genomic DNA in this region revealed a non-methylated cluster of rare-cutter restriction endonuclease sites (Fig. 4.1.1(i)). An *NruI* site and two closely situated *NotI* sites interspersed with a partially methylated *MluI* site were identified within the genomic DNA of the B cell line analysed. The cosmid SCG43 which maps to this site was found to contain a total of 4 *NotI* sites (Fig. 4.1.1(ii)). Since the recognition site for *NotI* consists entirely of Cs and Gs, most *NotI* sites tend to occur predominantly within CpG islands (Larsen et al. 1992). Hence it was decided to investigate this region further with the aim of isolating the gene or genes associated with the putative CpG islands.

4.1.2 Isolation of corresponding cDNA clones

The two largest *NotI* fragments from cosmid SCG43 were excised from a preparative gel, radioactively labelled and used to probe a λgt11 placental cDNA library (Clontech) (see materials and methods). A total of five duplicated positively hybridising spots were visible on the library autoradiographs out of a total of 8 x 10⁵ plaque-forming units screened. The positives were picked and passed through two further rounds of screening followed by subcloning of the single positive phage clones into pBluescript plasmid vector (see materials and methods). Further analysis was carried out on two independent clones, 1 and 4 which appeared to have insert sizes of 1.5 kb and 2.5 kb respectively. Both appeared to hybridise to the original SCG43 probes (data not shown). Northern analysis of clone 1 showed it detected a ubiquitously expressed 2.2kb message (Fig. 4.1.2), whereas clone 4 detected a weakly expressed 4.5 kb message predominantly in thymus (data not shown). This suggested these clones corresponded to different genes, since they appeared to hybridise to different mRNA species.

Sequencing of the ends of each of the clones was carried out by automated sequencing using vector primers. Approximately 300 bp of sequence obtained from each end of each clone was used to search DNA sequence databases for homology to pre-existing sequences. These searches revealed 100% identity of clone 1 to the cDNA sequences from the human nucleolus organiser region autoantigen NOR-90 and the human...
Fig. 4.1.1. (i), Genomic PFGE map of the D17S183 region. The site of a putative CpG island and the location of cosmid SCG43 are shown. Sites for NotI (N), NruI (Nr) and MluI (M) are clustered together in the genomic DNA. (ii), Restriction digestion of cosmid SCG43 with NotI. Restriction digestion with NotI and electrophoresis through an ethidium-stained gel, revealed 5 NotI fragments (see arrows) due to the presence of 4 NotI sites within the cosmid insert. Hybridisation with pCOS8 vector DNA to a Southern filter from this gel revealed the vector band (v).

Fig. 4.1.2. Northern analysis of a cDNA clone isolated from a placenta cDNA library using two large NotI fragments from cosmid SCG43 as probes. Clone 1 detects a ubiquitously expressed 2.2kb mRNA corresponding to the ribosomal RNA transcription upstream binding factor, hUBF. 1, spleen; 2, thymus; 3, prostate; 4, testis; 5, ovary; 6, small intestine; 7, colon; 8, peripheral blood leukocyte. Bottom panel: actin control hybridisation to the same northern filter.
Features of the physical map

ribosomal RNA transcription upstream binding factor (hUBF). The NOR-90 and hUBF proteins appear to be derived from a common precursor hUBF mRNA via alternative splicing, as suggested by immunoprecipitation analysis coupled with sequencing of the corresponding cDNA clones (Chan et al. 1991). The function of hUBF/NOR-90 is to control rRNA synthesis by binding an upstream control element and a core element of rRNA genes which has the effect of enhancing transcription by RNA polymerase I.

Sequence analysis and database searches revealed no such homology to clone 4 which failed to reveal homology to anything in the database, providing further evidence that it corresponded to a different gene from the region.

4.1.3 Localisation of the human upstream binding factor (hUBF) gene

Confirmation that this gene mapped to this region of chromosome 17q21 was achieved by hybridisation of a PCR-generated probe from the gene to the chromosome 17-specific hybrid mapping panel and the SCG43 cosmid (Fig. 4.1.3). A forward primer at position 1359 and a reverse primer at position 2480 of the human autoantigen NOR-90 sequence (accession number 56687) were used in a PCR reaction with the clone 1 cDNA as a template to generate a 1.1kb probe.

Hybridisation of this probe to a somatic cell hybrid mapping panel filter (Fig. 4.1.3(i)) and to an extended panel of CMGT hybrids (not shown) showed this probe was present in the chromosome 17-only hybrid TRID62 and the chromosome 17q-only hybrid, PCTBA1.8 as well as in the APL t(15;17)15q+ and NF1 t(1;17) translocation hybrids. It also appeared to be present in the CMGT hybrids KLT8, PLT6B, KLT12 PLT20 and PLT15, but was absent in the APL17q-, PLT8 and KLT3 hybrids. This pattern of hybridisation was identical to probe B43 from the D17S183 region (see table 3). The same probe also hybridised strongly to the D17S183 cosmid SCG43, but not to cosmids from elsewhere on chromosome 17 (Fig. 4.1.3 (ii)). Hybridisation analysis of clone 4 produced identical results (data not shown).
Fig. 4.1.3 Localisation of the human upstream binding factor (hUBF) gene. (i) Hybridisation to the chromosome 17 somatic cell hybrid mapping panel. The hUBF probe hybridises to a large 18 kb BamHI band in the same hybrids as probes from D17S183 and EPB3 (see table 3). The smaller hybridising band is due to the mouse homologue of the UBF gene which is clearly visible in the mouse cell line, EL4 and in each of the other hybrids which have been constructed with a mouse background.

Fig. 4.1.3 Localisation of the human upstream binding factor (hUBF) gene. (ii) Hybridisation to the D17S183 cosmid, SCG43. Hybridisation results clearly show specific hybridisation of the hUBF gene to cosmid SCG43. No hybridisation was observed with other cosmids from elsewhere on chromosome 17.
4.2 Length heterogeneity of the RNU2 gene cluster and identification of rare-cutter restriction fragment polymorphisms

4.2.1 Polymorphic hybridisation patterns around the RNU2 locus

Hybridisation of a probe from the RNU2 locus (a gift from A Weiner, Yale University) and probes adjacent to it, to PFGE Southern filters, containing genomic DNA from a number of different lymphoblastoid cell lines digested with BssHII, MluI or NotI restriction endonucleases, revealed distinctive polymorphic patterns of hybridisation (Fig. 4.2.2). Huge variation in the sizes of the hybridising bands between the different cell lines were observed. Two different-sized hybridising bands were observed in the majority of the cell line DNA samples corresponding to two different alleles within the cell line. Hybridising bands in the DNA samples from the cell lines corresponding to four related individuals 3543, 3550, 3566 and 3677 appeared to be of similar sizes, indicating they shared the same alleles.

4.2.2 Size variation of the intact U2 gene array at the RNU2 locus

The variation of the rare-cutter restriction fragment sizes at the RNU2 locus observed in the above Southern analyses was thought to be due to previously documented variation in the number of U2 repeat units of the U2 gene array (Van Arsdell and Weiner 1984). The exact size of the array in the different cell lines was therefore investigated to see if this would explain the restriction fragment length variation.

The U2 gene array consists of 10-20 repeat units flanked by junction fragments JR and JL as shown (Fig. 4.2.1(i)). These details were kindly provided by A. Weiner, Yale University and have since been published (Pavelitz et al. 1995). The size of the intact arrays in each of a number of human lymphoblastoid cell lines was calculated by digesting their genomic DNA with EcoRI, which does not cut within the U2 repeat unit but which does cut at the end of each attached junction fragment, as shown (Fig. 4.2.1(i)). This produced a large EcoRI fragment for each allele containing the tandem array together with its attached junction fragment. The size of each tandem array was obtained by electrophoretically separating the EcoRI-digested genomic DNA by PFGE followed by Southern analysis using a 3.7kb PvuII-HindIII fragment from the repeat unit (a gift from A Weiner, Yale University) (Fig. 4.2.1(ii)).
Fig. 4.2.2 PFGE Southern hybridisation results with the RNU2 probe. A polymorphic pattern of hybridisation is observed with PFGE Southerns containing BssHII digested genomic DNA (A) and MluI digested genomic DNA (B). H, D, F, E, lymphoblastoid cell lines from unaffected individuals; BT20, HBL100, MCF-7, MDA MB 157, MDA MB 231, SKBR3, T47D, ZR75, ZR 75-1, breast tumour cell lines; 3543, 3550, 3566, 3677, 487,489, lymphoblastoid cell lines from affected individuals from chromosome 17q-linked families.
**Fig. 4.2.1(i) Map of the human U2 repeat unit with its flanking junction fragments.** The U2 gene array consists of over 10 copies of repeat units flanked by junction fragments JR and JL, as shown. EcoRI sites are present at each end of the entire array, as indicated by the black arrows. E, EcoRI; H3, HindIII; P, PvuII; LTR, viral long terminal repeat; CT, dinucleotide repeat of (CT70); U2, U2 snRNA coding region. (adapted from Pavelitz et al 1995)

**Fig. 4.2.1(ii) The variation in size of the U2 tandem array in different cell lines.** High molecular weight genomic DNA from 9 different lymphoblastoid cell lines (1-9), including that used in the construction of the PFGE map of *BRCA1* (8) was digested with EcoRI and electrophoretically separated by PFGE, blotted and probed with the 3.7kb *PvuII-HindIII* fragment of the U2 repeat unit (see (i)). Intact arrays from both alleles in each cell line are clearly visible as single bands ranging from ~50 kb to 260 kb in size. The smear below represents background hybridisation to bulk EcoRI fragments.
The results indicate a huge variation in the size of the intact array, both between the different alleles of each cell line and between the cell lines themselves. The sizes of the different alleles ranged from around 50 kb to 260 kb, suggesting a range in the number of repeat units per haploid genome from 10 to over 40 copies; much greater than previous estimates (Fig. 4.2.1(ii)).

4.3 Discussion of the features of the \textit{BRCAl} region identified by PFGE analysis

The approach used in this project to search the regions of CpG islands for genes has successfully isolated cDNA clones which potentially correspond to two different genes. Both appear to map to the same \textit{Not} I fragment at the D17S183 locus.

The discrepancy between the number of \textit{Not}I sites found at D17S183 by PFGE analysis of chromosomal DNA and the number found by restriction enzyme analysis of the cosmid from the same region could be explained in several ways. The most likely explanation for the occurrence of two additional \textit{Not}I sites at this location in the cosmid which were not detected in the genomic DNA analysis is that the \textit{Not}I fragments resulting from cleavage at these sites were too small to be detected by PFGE. The discrepancy might also be explained by sequence differences between the cell line used to construct the cosmid library and the genomic B cell line used in the PFGE analysis due to polymorphism at this locus. Alternatively, it is possible that the two further \textit{Not}I sites are present in the genomic DNA at this site but were not cleaved by the \textit{Not}I endonuclease due to the presence of methyl groups. The same \textit{Not}I sites were accessible to the enzyme in the cosmid DNA due to the removal of methyl groups which occurs during the cloning process. Further analysis of this genomic region is required in order to determine the exact number of CpG islands at this locus and to identify any further genes localising to the region.

The ribosomal RNA transcription upstream binding factor gene, \textit{hUBF} identified at the D17S183 CpG island was an unlikely candidate for \textit{BRCAl} since it appears to have a critical housekeeping role in all cells. It was difficult to see how disruption of this gene in the germline DNA of individuals would lead to breast cancer. Any pathological association this gene may have is more likely to be with autoimmune responses. The UBF protein in addition to RNA polymerase I and other proteins of the transcription machinery appear to be the targets of autoantibodies produced in the autoimmune disease, scleroderma (Rodriguez-Sanchez \textit{et al.} 1987). A second reason for the elimination of this gene and the other putative gene at this locus as \textit{BRCAl} candidates was their location. A crossover event in an affected 34-year old female member of a chromosome 17q-linked...
breast cancer family was discovered which suggested $BRCA1$ was centromeric to D17S78, a marker which is proximal to D17S183 (Simard et al. 1993).

The variability of the tandem array at the $RNU2$ locus shown here is an interesting feature of this region of chromosome 17q21. It was first visualised by PFGE Southern analysis of rare-cutter digested genomic DNA with probes from around $RNU2$ (Fig. 4.2.2). Further work showed that genomic DNA from number of individuals showed a wide variation in size of the tandem array between different alleles of each cell line and between the cell lines themselves (Fig. 4.2.1(ii)). This served as a useful signpost in the PFGE physical map as probes giving such hybridisation results could be quickly localised. However, the vast amount of variation in the hybridising restriction fragments observed with this method precluded its use in screening for cytological changes such as large deletions and rearrangements in this part of the $BRCA1$ region within the genomic DNAs of affected breast cancer families. Any changes around the $RNU2$ locus within these individuals would be indistinguishable from normal variation at this locus.

This size variation of the $RNU2$ locus is most likely to be due to variation in the number of repeat units in each tandem array, rather than polymorphisms within the repeat unit on the basis of earlier work which suggested homogeneity of the U2 repeat unit (Van Arsdell and Weiner 1984). Large numbers of homogeneous U2 snRNA genes at this locus are required in order to maintain high expression levels of U2 snRNA which is crucial to cell survival. The mechanisms involved in the variation of the $RNU2$ locus are unclear but may involve the recombinogenic properties of the LTR or the large (CT) dinucleotide repeat within each repeat unit or the U2 tandem array itself (discussed in (Pavelitz et al. 1995)).
CHAPTER 5 GENOMIC CLONING IN THE BRCA1 REGION

Following the narrowing of the genetic region containing BRCA1 to 1.0-1.5cM between the marker D17S776 (Goldgar et al. 1994) and D17S78 (Simard et al. 1993), the positional cloning effort began to concentrate on genomic contig construction. Initial work utilised yeast artificial chromosome clones (YACs) which, because of their large insert capacity, presented a quick way to complete the genomic clone coverage of the region. Contigs of YAC clones were assembled throughout the region from several YAC libraries with the aid of the many polymorphic markers, genes and anonymous DNA probes from the region (Albertsen et al. 1994; Jones et al. 1994). Unfortunately, problems of chimaerism, instability and rearrangements prevented the use of the YAC contigs in the direct isolation and mapping of genes housed within the BRCA1 region.

Work described in this chapter concentrated on the distal half of the BRCA1 region between 1A1.3B and D17S78. This involved the conversion of YAC clones which had been isolated by other members of the laboratory into ordered chromosome 17-specific cosmid pools (section 5.1). This avoided the analysis of other genomic regions present in the chimaeric YAC clones and isolated non-deleted copies of the DNA that was unstable in the YACs. The extent of the genomic clone coverage of this half of the BRCA1 region was assessed using the PFGE long-range restriction map and revealed a 300 kilobase pair gap between the cosmid pools (section 5.2). In order to fill the hole in the genomic clone contig of the region, two cosmid walks were initiated; one progressing towards the telomere from the proximal end of the hole at RNU2, and the other progressing towards the centromere from the distal end of the hole at D17S858/D17S859 (sections 5.3.1 and 5.3.2). After several steps, Southern hybridisation analysis of the cosmids at the ends of each walk suggested that the two walks had met. (section 5.4) Analysis of the cosmids by fluorescence in situ hybridisation confirmed that the cosmid walk was contiguous and the genomic coverage of the region was complete (section 5.5).

5.1 Conversion of YAC clones to cosmid contigs

Several YACs were isolated with probes and PCR primers from known genes and anonymous markers in the region from 1A1.3B to D17S78 (Table 5.1, data provided by M.Bonjardim and M.Brown). Of the YACs tested, all appeared either
Table 5.1 Results of YAC analysis in the IA1.3B-D17S78 region. Each YAC, indicated in the first column, has been tested for each marker in the top row. Presence of a marker is indicated as +. YAC 26D6 lost markers D17S858 and D17S859 on deletion, as indicated by brackets. (from Jones et al 1994)

<table>
<thead>
<tr>
<th>YAC name</th>
<th>Library a</th>
<th>Probe</th>
<th>size (kb)</th>
<th>Features b</th>
<th>D17S855</th>
<th>IA1.3B</th>
<th>RNU2</th>
<th>D17S858</th>
<th>D17S859</th>
<th>PPY</th>
<th>D17S78</th>
</tr>
</thead>
<tbody>
<tr>
<td>106G4</td>
<td>ICRF</td>
<td>IA1.3B</td>
<td>700</td>
<td>unstable</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12H4</td>
<td>ICRF</td>
<td>IA1.3B</td>
<td>500</td>
<td>chimaeric</td>
<td>+ +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A167E6</td>
<td>Wash</td>
<td>RNU2</td>
<td>200</td>
<td>chimaeric</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>176B12</td>
<td>CEPH17</td>
<td>RNU2</td>
<td>800</td>
<td>nd</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B196</td>
<td>Wash</td>
<td>RNU2</td>
<td>150</td>
<td>nd</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174A12</td>
<td>CEPH17</td>
<td>YAC</td>
<td>700</td>
<td>chimaeric</td>
<td>+ +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>175G10</td>
<td>CEPH17</td>
<td>YAC</td>
<td>nd</td>
<td>chimaeric and deleted</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26D6</td>
<td>CEPH</td>
<td>PPY/p131</td>
<td>250-400</td>
<td>unstable</td>
<td>(+) (+)</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B260E11</td>
<td>Wash</td>
<td>PPY</td>
<td>&lt;100</td>
<td>nd</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>172H5</td>
<td>CEPH17</td>
<td>YAC</td>
<td>450</td>
<td>nd</td>
<td>+ +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>175H2</td>
<td>CEPH17</td>
<td>PPY</td>
<td>350</td>
<td>nd</td>
<td>+ +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109E5</td>
<td>ICRF</td>
<td>PPY/p131</td>
<td>nd</td>
<td>nd</td>
<td>+ +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a ICRF, ICRF YAC library; Wash, Washington School of Medicine library; CEPH17, CEPH chromosome 17 selected mega YAC library.
b Chimaerism was determined by in situ analysis of YAC Alu PCR probes.
c Alu PCR probe from YAC.
chimaeric, rearranged or contained deletions. For the purposes of completing the
genomic coverage of the 1A1.3B-D17S78 region, the YAC 12H4, which contains 1A1.3B
and RNU2, and a preparation of YAC26D6 prior to deletion which contained D17S858,
D17S859, PPY and D17S78 were used to screen a gridded flow-sorted chromosome 17
cosmid library (Lehrach 1990).

5.1.1 Generation and ordering of a cosmid pool from YAC 12H4

A pool of Alu-PCR products from YAC 12H4 was used to probe the cosmid library and
identified over 200 strong positive clones. The majority of these clones corresponded to
the repeated array of the U2 RNA unit (see section 4.2) (Van Arsdell and Weiner 1984),
as revealed by hybridisation of a duplicate library filter with the RNU2 probe (D.Black
data not shown). Those clones not corresponding to the RNU2 locus were analysed
further. Some of them hybridised to a full length 1A1.3B cDNA probe and one of these,
cosmid B0176 appeared to contain both 1A1.3B and RNU2, confirming earlier genomic
dNA PFGE analysis of the close proximity of these genes (see section 3.3). A weakly
hybridising RNU2 positive clone, E1132 was found to be negative for 1A1.3B and Alu-
PCR fingerprinting analysis suggested it did not overlap with cosmid B0176 (D.Black
data not shown). Hybridisation of three large EcoRI fragments from E1132 to the cosmid
library filters identified two further cosmids, C1198 and A04131, both of which failed to
hybridise to the RNU2 probe. It was thought that these new cosmids were situated
adjacent and distal to the RNU2 locus on the opposite side to 1A1.3B (Fig.5.1.1(i)). This
was confirmed by the identical Southern hybridisation results produced when a
radioactively labelled repeat free fragment from C1198 and the RNU2 probe were
sequentially hybridised to the same PFGE Southern filter containing genomic DNA
digested with a combination of rare-cutter restriction endonucleases (Fig. 5.1.1(ii)).
Fig. 5.1.1(i) Confirmation of the location of C1198 adjacent to the RNU2 locus by PFGE analysis. A PFGE Southern filter, containing genomic DNA from a B cell line digested with a combination of rare-cutter restriction endonucleases, was hybridised in turn with the RNU2 probe (A) and a fragment from cosmid C1198 (B). The identical patterns of hybridisation observed suggest cosmid C1198 is adjacent to the RNU2 locus at the location shown below.

Fig. 5.1.1 (ii) Ordering of the cosmid pool from YAC 12H4. Information on the order and overlap of the cosmids was provided by PCR and Southern analyses (D. Black, M. Brown data not shown). PFGE Southern analysis results shown above suggest cosmid C1198 lies adjacent and distal to the RNU2 locus.
5.1.2 Generation and ordering of a cosmid pool from YAC 26D6

Cosmid clones corresponding to the chromosome 17 region contained in YAC 26D6 were identified by hybridising a pool of Alu-PCR products from YAC26D6 to the gridded cosmid library filters. A total of 18 positively-hybridising cosmid clones were isolated, and analysis by PCR (H. Nicolai data not shown) revealed which markers were present in each of the cosmids, giving a rough preliminary ordering. Subsequent Southern hybridisation analysis of the cosmid set, using fragments from three different cosmids as probes to reveal overlaps, further refined the order and suggested the markers D17S858/D17S859 were roughly 60 kb away from PPY/D17S78 (Fig. 5.1.2(i)). Using these results, an ordered contig consisting of 10 of these cosmids was constructed (Fig. 5.1.2(ii)). The order and extent of overlap of the cosmids in the contig were confirmed by PstI restriction fragment size comparisons (D.Black, data not shown).

5.2 Assessment of the gap between cosmid contigs

An accurate scaled long-range restriction map incorporating IAl.3B, RNU2 and probe 12E/Bd from the D17S858/D17S859-positive cosmid, G05149 (Fig. 3.3.1) was used to estimate the distance between C1198, the most distally extending cosmid from the RNU2 contig and G05149, the most proximally extending cosmid from the PPY contig. The map incorporated all of the restriction fragment sizes from the PFGE Southern analysis in this region and was therefore able to pinpoint the exact locations of these probes throughout the 750 kb NotI fragment to which they all mapped. It revealed the distance between the cosmid G05149 probe 12E/Bd and RNU2 as around 300 kb (Fig. 5.2). Considering cosmid C1198 appeared to be immediately adjacent to RNU2 (Fig. 5.1.1(ii)), this implied the distance between C1198 and G05149, was also around 300 kb.
1. Hybridisation of repeat free fragments from cosmids G05149 and B07136, suggesting overlap between these two cosmids.

2. Hybridisation of repeat free fragments from cosmids D1198 and D07136, B07136, C07101, and G05149, showing strong hybridisation to G05149, B07136, C07101, and itself.

3. Hybridisation of a large fragment from cosmids DO1179 and D1198, showing strong hybridisation to G12160, D09100, A01183, and D11181.

Taken together, these results indicate an overlap between the PPY/D17S78 positive cosmids, A01183, D01179, D1198, and C07101, which in turn overlap with the D17S858/859 positive cosmids B07136 and G05149.

Fig. 5.1.2 (i) Southern analysis of cosmids isolated with the 26D6 YAC.

Fig. 5.1.2 (ii) Ordering of the cosmid pool from YAC 26D6.

The order of the cosmids and the extent of overlap between neighbouring cosmids was deduced from PCR analysis and the results of Southern analyses shown above.
Fig. 5.2 The gap between the cosmid contigs around RNU2 and PPY.

5.3 Construction of a 300 kb cosmid walk in the BRCA1 region

The walking strategy employed the dual opposed SP6 and T7 promoters flanking the cloning site in the cosmid vector Lawrist 4 (D. Nizetic, unpublished data) to create probes from each end of the cosmid (see materials and methods). Neighbouring cosmids were identified when each end-specific probe was hybridised in turn to the cosmid library, as shown (Fig 5.3) (Jones et al. 1994). Hybridisation of the riboprobe to the cosmid in the library from which it was generated, served as an internal positive control. At each step, newly isolated cosmids were checked by Southern analysis to ensure they overlapped with previous cosmids in the walk (see Fig. 5.3.1 and 5.3.2).

5.3.1 Generation of a cosmid walk from D17S858/D17S859 in the direction of the centromere

This walk commenced at cosmid G05149 (see Fig. 5.1.2(i)) and proceeded proximally for seven steps, covering an approximate distance of 180 kb, based on the average size of a cosmid insert and an estimated 10-30 kb overlap between each cosmid in the walk. The cosmids isolated at each step in the walk are indicated in Fig. 5.3.1. In step 1, riboprobes from both ends of cosmid G05149 were pooled and hybridised simultaneously to the cosmid library filters. The positively-hybridising cosmid clones were identified and checked by Southern analysis with probes from cosmid G05149 to ensure they were genuine overlapping cosmids (see step 1 autoradiograph results, Fig. 5.3.1).
Fig. 5.3. Hybridisation of riboprobes from each end of cosmid B09127 to filters containing gridded cosmids from the flow-sorted chromosome 17 cosmid library. The T7 riboprobe extends furthest into the gap as indicated by its hybridisation with new cosmids, C05123 (5) and C05173 (4), whilst the SP6 riboprobe hybridised to previous cosmids in the walk, E1132 (2) and A04131 (3). Both riboprobes hybridised to the cosmid B0927 (1) from where they originated, serving as an internal positive control. (from Jones et al 1994)
Fig. 5.3.1 Construction of a cosmid walk from D17S858/D17S859 to the centromere.
The walk commenced from cosmid G05149 and proceeded by sequential hybridisations of cosmid riboprobes onto the gridded cosmids library filters (see section 5.3 text). Southern analysis was carried out at each step in the walk (1-7) to ensure newly-isolated cosmids hybridised to the most proximal end of the previous cosmid in the walk. The dotted lines indicate the cosmids which hybridised to each riboprobe. A positive control (+) and a cosmid from elsewhere on chromosome 17 (-) were included in each Southern filter for comparison.
The most proximally extending of these new cosmids were identified through their failure
to hybridise with a probe from cosmid B07136, which overlaps G05149 at its distal end.
One of these proximally-extending cosmids, F0829, was used in the next step in the walk.

The T7 riboprobe from F0829 hybridised to G05149, C0530, A0748 and F1051 (see Fig.
5.3.1), whilst the SP6 riboprobe from F0829 isolated several new cosmids including
D0571 and G1151 (autoradiograph not shown). Cosmid G1151 was used for the next
step. Its T7 riboprobe showed specific hybridisation to cosmids F0829, C0530, A0748
and F1051, which confirmed the position of cosmid G1151 in the walk. The SP6 end of
G1151 was used to extend the walk proximally.

In the subsequent steps in the walk, riboprobes from each end of the "walking" cosmid
were hybridised in turn to the cosmid library. In each case, one of the riboprobes
hybridised to cosmids which had been isolated in the previous step in the walk. This
served to confirm the position of each cosmid in the walk. The riboprobe from the
opposite end of the walking cosmid identified new cosmids which in turn were used for
the next step in the walk.

5.3.2 Generation of a cosmid walk from RNU2 in the direction of the telomere

This walk consisted of five steps covering an approximate distance of 120 kb. It began at
cosmid C1198, one of the cosmids isolated with fragments from the RNU2-positive
cosmid, E1132 (see section 5.1.1) (Fig. 5.3.2). The T7 riboprobe from C1198 detected
over 50 positive cosmid clones in the library including E1132 (data not shown). These
cosmid were assumed to contain tandem repeat units from the RNU2 locus, indicating
Fig. 5.3.2 Cosmid walk from \(RNU2\) to the telomere. The walk from the \(RNU2\) locus began with cosmid C1198 (section 5.1.1). Riboprobes from each end of C1198 were hybridised in turn to the gridded cosmid library filters and the positively-hybridising clones identified. The cosmids hybridising to the more distal SP6 riboprobe, as shown by the vertical dotted line, were checked by Southern analysis with the original C1198 SP6 riboprobe (see step 1). One of the cosmids showing specific hybridisation was chosen for the next step in the walk. Each step 1-6 was carried out in a similar way. In each Southern a positive control (+) was included. Cosmids not hybridising to the riboprobe (-) were excluded from further study.
that this end of C1198 was the wrong one to extend the walk distally towards D17S858 and the telomere. The SP6 riboprobe from C1198 hybridised to two new cosmids, B09127 and C0835. The walk proceeded distally by hybridising each riboprobe sequentially to the cosmid library and checking new cosmids by Southern analysis as before.

5.4 Meeting of the cosmid walks between \( \text{RNU2} \) and D17S858/D17S859

At the fifth step in the \( \text{RNU2} \)-telomere cosmid walk, hybridisation of the cosmid G124 T7 riboprobe to the gridded chromosome 17 cosmid library detected several cosmids which had been isolated at the seventh step in the \( \text{PPY} \)-centromere walk. This finding indicated that the two walks had met. To further confirm the overlap of the two walks, Southern filters containing different digests of two of the cosmids from the proximal end of the \( \text{PPY} \)-centromere walk, D12132 and C02179 and the \( \text{RNU2} \)-telomere walk cosmid G124 were hybridised with the G124 T7 riboprobe (Fig. 5.4). A clear single hybridising band was observed in each cosmid. The absence of any background bands in the autoradiograph served as an internal negative control, indicating that the G124 T7 probe did not contain repetitive elements and the single hybridising bands observed in each cosmid track were due to specific hybridisation to the probe.

5.5 Confirmation of the cosmid walk integrity and physical mapping by FISH

Analysis of combinations of cosmids from the \( \text{1A1.3B} \) to D17S78 region by FISH (Senger et al. 1993) (data provided by P.Gorman) confirmed the ordering of the cosmids as C1198-G05149-D11181 and C1198-G124-G05149 (Fig. 5.5). In addition, FISH analysis on chromatin released from nuclei (Senger et al. 1994), confirmed the overlap of the two cosmid walks at cosmids G124 and C02179 (data not shown). This provided further evidence that the cosmid walk was contiguous and that the order of the cosmids within the walk is correct.
Fig. 5.4 Southern hybridisation results indicating overlap between cosmid G124, from the distal end of the RNU2-telomere walk, and cosmids C02179 and D12132 from the proximal end of the PPY-centromere walk. A single restriction fragment (shown by the arrows) from each cosmid digest has hybridised to the cosmid G124 T7 riboprobe at roughly the same intensity as that shown by cosmid G124 itself, the positive control. This specific hybridisation pattern indicates cosmid G124 overlaps with both C02179 and D12132 and that the two cosmid walks have met.
Fig. 5.5 Confirmation of the cosmid walk between RNU2 and D17S78 by FISH analysis. Panel (A) shows cosmid C1198, labelled red, D11181, labelled green and G05149, labelled red giving an order red-red-green. In panel (B), the red-labelled C1198 cosmid, the red-labelled D11181 cosmid and the green-labelled G05149 give an order, red-green-red. Panel (C) shows G124 (green), C1198 (red) and G05149 (green) arranged in the order red-green-green. These results suggest the ordering of the cosmids within the walk shown above is correct and the cosmids are contiguous.

(from Jones et al 1994, FISH analysis carried out by P.Gorman)
CHAPTER 6 LOCALISATION OF GENES BETWEEN IA1.3B TO D17S78 AND CONSTRUCTION OF A DETAILED TRANSCRIPT MAP OF THE REGION

Following the construction of a complete cosmid walk from IA1.3B to the polymorphic marker, D17S78 (detailed in chapter 5), this project progressed to the transcriptional analysis of the contig and the isolation of genes housed within the constituent clones. The first part of this work required the location of the likely sites of transcriptional activity within the contig. Analysis by rare-cutter restriction enzyme digestion of the minimal set of cosmids spanning the region identified several sites of putative CpG islands; stretches of DNA which are relatively high in G+C content (see section 6.1) (Jones et al. 1994). Previous studies have shown a strong association of these sequences with genes, particularly with the regions where transcription begins (reviewed (Bird 1987)). A parallel survey for evolutionarily conserved EcoRI restriction fragments, carried out by Julie Chambers in the laboratory, showed that some of these putative islands were also conserved in several different species. By incorporating this information, the likely locations of genes throughout the contig were identified.

Further work, carried out in this project and by other members of the laboratory, involved the isolation of genes within the cosmid contig. A number of gene isolation techniques were employed including direct selection and exon trapping as well as conventional screening for cDNA clones in random-plated cDNA libraries. Several cDNA clones corresponding to the dual specificity VH1-related VHR phosphatase gene, HDSPH, were isolated which were analysed in detail (section 6.2). These cDNA clones and others isolated throughout the region were precisely localised on the contig (section 6.3). Northern analysis of the clusters of cDNA clones carried out in this project indicated they were separate gene loci (section 6.3). Sequence homology searches carried out in this work and by other members of the laboratory revealed that a number of the genes isolated in the contig had homology to other previously isolated genes.

6.1 Identification of putative CpG islands in the region and the likely locations of genes in the cosmid contig.

The minimal set of overlapping cosmids in the contig (indicated in bold, Fig. 6.1(ii)) were screened by restriction enzyme analysis for the presence of CpG islands (Linsay and Bird 1987). Each cosmid was digested with the following combinations of enzymes:

(a) EcoRI only
(b) EcoRI + SacII
(c) EcoRI + BssHII
Transcriptional analysis

The digested DNA samples were separated electrophoretically side by side on an ethidium-stained agarose gel. The EcoRI restriction fragments which were cut by either SacII or BssHII or both were identified (Fig. 6.1(i)). The recognition sites of both SacII and BssHII enzymes contain only Gs and Cs and are predominantly located within CpG islands (Larsen et al. 1992). Hence, the location of their recognition sites indicates the likely location of CpG islands. The EcoRI fragments which were cut by both enzymes were thought to be highly likely to contain CpG islands and their locations were indicated on the contig map (Fig. 6.1(ii)).

6.2 Isolation of the VHR phosphatase gene, HDSPH

As part of a large-scale cDNA isolation exercise, the experiments described in this thesis were designed to isolate cDNA clones specific to the cosmid walk between 1A1.3B and D17S858/859. Selected cDNA libraries enriched with transcripts from the region were generated from small groups of four or five cosmids by hybrid selection with several different cDNA library sources including foetal brain, foetal liver, adult skeletal muscle (Stratagene) and the breast cancer cell line, ZR 75 (a gift from Dr M. Parker) (Lovett et al. 1991; Korn et al. 1992; Lovett 1994). After selection, the cDNA population was cloned into a plasmid vector and transformed into E.coli cells. Around 500 colonies for each library were picked into microtitre plates and spotted onto filters. Selected cDNA clones were hybridised back to the sublibrary in order to organise the positive clones into sets of cDNA contigs. This work was carried out by Bernhard Korn at the Deutsches Krebsforschungszentrum, Heidelberg and by Beatrice Griffiths in the laboratory.

The resultant gridded library filters were used for hybridisation to the CpG island-containing fragments, evolutionarily-conserved fragments, whole cosmids and single copy probes from the contig, identifying numerous cDNA clones. This chapter describes the isolation and identification of cDNA clones from the human dual-specificity VH1-related phosphatase gene together with the localisation of cDNA contigs throughout the region.

A fragment from cosmid G01152 of the cosmid walk was used to screen a gridded selected cDNA library which had been constructed from cosmids C02179, A1028, G01152 and B07165. Several positive clones were identified and a number of clones including I19g2, J13g2, G7g1, G4g1 and L1g2 (marked by arrows, Fig. 6.2.1(i)) were picked for further analysis. The insert sizes of these cDNA clones ranged from 650 bp (I19g2) to 850 bp (G7g1) (data not shown). Northern analysis using the I19g2 insert probe detected a ubiquitously expressed transcript of approximately 4.5kb (see Fig. 6.4†).
Fig. 6.1(i) CpG island screen of the cosmids from RNU2-D17S858/859 by restriction enzyme analysis. Each cosmid was digested with EcoRI (R), EcoRI+BssHII (RB), EcoRI+SacII (RS) and the digested samples electrophoretically separated on ethidium-stained gels, as shown.

- \( \text{EcoRI fragment cut by BssHII} \)
- \( \text{EcoRI fragment cut by SacII} \)
- \( \text{EcoRI fragment cut by both BssHII and SacII} \)
- \( \text{restriction fragments due to vector} \)
Fig 6.1(ii) Transcriptional analysis of the cosmid contig between RNU2 and D17S858/859. The locations of 8 newly isolated genes, represented by black boxes are shown. The Southern hybridisation data which enabled the localisation of each of the gene loci is given in figures 6.3(i)-(v). The approximate sites of putative CpG islands, identified by rare-cutter restriction enzyme analysis of cosmid DNA (see Fig 6.1(i)), are indicated by open boxes. The regions of the cosmid contig showing evolutionary conservation are indicated by horizontal black bars (data provided by J. Chambers). The minimal overlapping cosmid set used in these analyses are in bold.
Figure 6.2.1(i) Isolation of *HDSPH* cDNA clones from a gridded selected cDNA library. A fragment from cosmid G01152 was hybridised to filters containing the gridded selected cDNA library constructed using cosmids from the *RNU2-D17S78* cosmid walk. Clones G7g1, I19g2, J13g2 and L1g2, marked by arrows were picked for further analysis.
Hence, it was decided to screen a full bacteriophage \( \lambda \)ZAP placenta cDNA library (Stratagene) to isolate the full length transcript.

Several \( \lambda \)ZAP placenta cDNA clones corresponding to the I19g2 probe were isolated. Preliminary sequence analysis of the clone ends was performed using vector primers. Two of the clones I191.2 and I191.4, both of which contained inserts of 2.5 kb, appeared to overlap to a large degree, as shown in Fig. 6.2.1(ii). Comparisons of their sequences to updated EMBL and Genbank DNA sequence databases failed to indicate any significant homologies to pre-existing sequences. However, sequence analysis of the overlapping 3.5 kb cDNA clone P62.a2 revealed 100% homology between one end of this clone and the published coding region of the HDSPH gene, as shown in Fig. 6.2.1(ii) (Ishibashi et al. 1992). The remaining sequence contained in the P62.a2 clone and clones I191.2 and I191.4 corresponded to the 3’ untranslated region of the HDSPH gene which had not been published or deposited into the database.

6.2.2 Complete sequence of HDSPH

The three clones P62.a2, I191.2 and I191.4 were sequenced end to end using vector primers initially and subsequently with primers designed approximately every 300bp from the new sequence. Each clone was sequenced on both strands from each end in both directions. The resulting sequence, given in Fig. 6.2.2 represents the entire HDSPH cDNA including the published sequence beginning 20 bp upstream of the coding start site up to nucleotide 841 (shown in the shaded area, Fig. 6.2.2) together with a further 3.5kb of sequence up to and including the polyadenylation signal, obtained from the three placental cDNA clones. Further analysis of this gene to determine the genomic structure was carried out to enable SSCP mutational analysis (chapter 7).

6.3 Localisation of clusters of cDNA clones throughout the region

The methods used to obtain cDNA clones from the HDSPH gene were employed to isolate cDNA clones corresponding to the rest of the cosmid walk between 1A1.3B and D17S78 and succeeded in generating several contigs of cDNA clones. The majority of this gene cloning work was carried out by other members of the laboratory and therefore is not described in detail.
Fig. 6.2.1(ii) cDNA clones corresponding to the HDSPH gene. Clones P62.a2, I191.1 and I191.4 overlap with each other and the published sequence as shown. The intron/exon structure of the gene and the primers used in the SSCP analysis are indicated (refer to chapter 7).
Figure 6.2.2 Nucleotide and deduced amino acid sequence of the dual-specificity VH1-related VHR phosphatase cDNA. The sequence was obtained from cDNA clones P62.a2 and 1191.1 together with the published sequence, indicated by the shaded area. (Ishibashi et al 1992). The poly (A)+ addition sequence is indicated by asterisks.
In order to understand the chromosomal arrangement of these transcripts and to verify their location within the \textit{BRCA1} region, members of each contig of cDNA clones were hybridised back to Southern filters containing digested DNA from each of the cosmids in the minimal overlapping cosmid set. The positions of each of the gene loci throughout the contig map are indicated in Fig. 6.1(ii). The locations of the \textit{HERV-K} related gene, two novel genes, \textit{ET6B} and \textit{8Eg2}, a novel G protein gene, \textit{HAL64} and \textit{HDSPH} were obtained from the results of the Southern hybridisations, as shown in Figures 6.3(i), 6.3(ii), 6.3(iii), 6.3(iv) and 6.3(v) respectively. The remaining genes indicated on the contig map (Fig. 6.1(ii)) were located by other members of the laboratory using PCR-based and sequencing methods.

\textbf{6.4 Identification of separate gene loci by northern analysis}

A single cDNA clone from each of the putative gene loci was hybridised to a northern filter containing a number of different human tissues (Fig. 6.4). One of the cDNAs located at the proximal end of the walk at cosmid C1198 was found to hybridise to a large message of around 9.5 kb, clearly visible in brain, liver and kidney. Sequencing of this clone and others from the region, carried out by other members of the laboratory, showed it had strong homology to the \textit{HERV-K} family of human endogenous retroviral elements (reviewed in (Leib-Mosch \textit{et al.} 1990)). The other cDNA clones from the region detected a range of different sized messages on the northern filters, and a wide variety of different expression patterns. A clone with identity to \textit{HDSPH} and another known as BCC10 which had no database homologies appeared to have a ubiquitous expression pattern, whilst other clones such as BCC8, BCC7 and \textit{8Eg2} appeared to be expressed in a more specific manner.
Fig. 6.3(i) Localisation of DNA probes corresponding to elements of a human endogenous retrovirus.
The cosmid fragment probe C1198f, with high homology to POL and ENV genes of the HERVK10 retrovirus is situated at the proximal end of the cosmid contig. The 9P cDNA, containing sequence with high homology to a retroviral LTR element is situated further distally at the end of cosmid B09127. A restriction map around the two loci has been constructed using the data provided in the autoradiograph results above. B, BamHI; E, EcoRI; X, XbaI
Fig. 6.3 Localisation of ET6B and 8Eg2 gene loci by Southern analysis of cosmids between RNU2 and D17S858/859.

The autoradiograph results indicate the ET6B cDNA clone hybridises to B09127 and the ends of A04131 and C1198. The 8Eg2 cDNA hybridises to a 25kb EcoRI fragment in C05123 and B09127, but does not hybridise to cosmids C1198 or E1132 (not shown), placing it further distal in the contig. The T7 riboprobe at the end of B09127 hybridises to the same 25kb EcoRI fragment as 8Eg2 (data not shown), suggesting that this fragment is at the end of cosmid B09127. E, EcoRI; X, XbaI.
Fig. 6.3(iv) Localisation of a novel G protein, HAL64 to the cosmid contig. The cDNA probe from this novel gene hybridises strongly to a 8.5kb EcoRI fragment in cosmids C05123, B09128, C0128, H03106, B0576 and to a smaller fragment in F0520, as shown. This indicates that HAL64 maps to the end of cosmid F0520 in the region shown.
Fig. 6.3(v) Localisation of cDNA clones corresponding to the *HDSPH* gene. Probes from cDNA clones II91.4 and L14g1, from the selected cDNA library were hybridised to Southern filters containing digested DNA from the cosmids between *RNU2* and D17S858/859. Strong hybridisation occurred with G01152 and B07165 cosmids only, as shown in the autoradiographs above. The L14g1 cDNA, containing exons 2 and 3 of *HDSPH*, hybridised strongly to the same region as the II91.4 probe, but more weakly to a second region further proximal, as indicated. A restriction map was constructed from the autoradiograph results.

E, *EcoR*I; B, *BamHI*; X, *XbaI*
Fig. 6.4. Northern analysis of cDNA clones from the RNU2-D17S858/D17S859 region. Each panel represents hybridisation of a cDNA clone from one of the putative gene loci in the region to a northern filter containing poly (A)+ RNA from a variety of human adult tissues (Clontech). Each of the clones, with exceptions of BCC8 and HDSPH cDNA clones, detects a different sized transcript with a different expression pattern. This suggests that each of the cDNA clones correspond to separate gene loci. Both BCC8 and HDSPH clones detect a ubiquitously expressed 4.5 kb mRNA, indicating they correspond to the same gene. However, independent sequence analysis of clones from BCC8 and HDSPH suggested they correspond to 2 separate genes (data not shown). 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas; 9, spleen; 10, thymus, 11, prostate, 12, testis; 13, ovary; 14, small intestine; 15, colon; 16, peripheral blood leukocyte.
6.5 Discussion of the transcriptional analysis and gene isolation work

The localisation of putative CpG island sites and regions which are evolutionarily conserved within the cosmid contig was a valuable exercise which allowed the gene isolation efforts to concentrate on the regions most likely to contain genes. In order to increase the sensitivity of the cosmid CpG island screen, both restriction enzymes SacII and BssHII were used since variability of island sequences means that some CpG islands do not contain SacII sites whilst others may not contain BssHII sites. The results of the screening for CpG islands through the analysis of cosmid DNA shown here differed from the results from the PFGE analysis of chromosomal DNA carried out earlier in the thesis work (refer to Fig. 3.5). An increased number of putative CpG islands within the 1A1.3B-D17S78 region was identified by the cosmid screening method; probably a result of the loss of methylation which occurs in cloned DNA. In chromosomal DNA, methylation blocks C-G enzyme sites in non-island DNA. Hence, the number of CpG islands identified by the cosmid screening method was thought to be an overestimate. A comparison of the locations of putative CpG islands with the locations of cDNA clone contigs shown in Fig. 6.1(ii) revealed that in fact five out of six of the putative CpG islands identified by the cosmid screening coincided with the locations of transcripts. Northern analysis confirmed that the cDNA contigs at each of these locations corresponded to separate gene loci. Each of these putative gene loci appeared to exhibit a different expression pattern and a different sized message, as shown in Fig. 6.4.

Sequencing and analysis carried out in this project (section 6.5) and by other members of the laboratory revealed several interesting features of the genes within the 1A1.3B-D17S78 region. The cDNA clones and genomic DNA from cosmid C1198 adjacent to the RNU2 locus revealed nearly 100% homology with LTR, POL and ENV genes of the human endogenous retrovirus, HERV-K10 (Ono et al. 1986) (data not shown). Northern analysis using clones from this region (as shown in Fig. 6.3) revealed a 9.2 kb message; the same size as that of full length HERV-K10 sequence. It is possible that this stretch of retroviral sequence is linked to the ancestral proviral insertion event which led to the expansion of the U2 repeat unit (Pavelitz et al. 1995) (refer to chapter 4). Interestingly, sequence analysis of the U2 repeat unit in baboon genomic DNA has also shown a 6 kb retroviral sequence including ENV, POL and GAG sequences between two LTRs (Pavelitz et al. 1995). This suggests that the HERV-K10 sequence found at this location is not mere coincidence. Further investigation into its involvement in breast tumorigenesis was merited considering previously observed involvements of retroviral sequences in carcinogenesis (see section 1.5.7) (reviewed (Leib-Mosch et al. 1990)).
The other genes discovered in this region also warranted further investigation into their involvement in breast tumour development due to their location within the limits defining the \textit{BRCA1} region. The \textit{HAL64} G protein gene appeared to be closely related to the highly conserved family of GTP-binding proteins with homology to the \textit{ras} oncogene superfamily and the heterotrimeric G protein \(\alpha\) subunits (Kahn \textit{et al.} 1993) (D Black data not shown). Hence, considering its oncogenic potential, further investigation of this gene seemed necessary. The localisation of the \textit{HDSPH} (section 6.4) within the cosmid walk was also of interest, considering its potential tumour suppressing role involving the antagonism of protein tyrosine kinases. Further investigation of this gene in breast tumour development is detailed in chapter 7.

The success in the isolation of the large number and variety of genes discovered within this region between \textit{1A1.3B} and \textit{D17S78} may have been due to the wide variety of techniques used in the gene isolation process. A large number of different cDNA library sources were used, probably increasing the likelihood of isolating most of the genes in this region, including those with high tissue specificity.
CHAPTER 7 SEARCHING FOR DISRUPTIVE MUTATIONS IN THE
GERMLINE DNA OF AFFECTED FAMILY MEMBERS AND SPORADIC
TUMOURS

A number of different techniques were employed in this project and by other members of
the laboratory to find the genetic defect which is responsible for breast tumorigenesis in
both inherited and sporadic breast tumours. In an initial attempt to locate the disease
gene, a search was carried out in the germline DNA of affected individuals from
chromosome 17q-linked families for constitutional chromosomal abnormalities (section
7.1). Observations of such defects, which include translocations and deletions, have been
instrumental in the identification of many disease susceptibility genes (reviewed (Collins
1992)).

Cytogenetic abnormalities involving large chromosomal regions have indicated the likely
locations of disease genes. Examples include the interstitial deletions of chromosome
13q in RB patients with congenital abnormalities (Yunis et al. 1981), translocations
involving chromosome 17q11.2 in NF-1 patients (Rey et al. 1987; Ledbetter et al. 1989)
and deletions of 5q in familial adenomatous polyposis (Herrera et al. 1986). Such
changes, and other more subtle alterations which are difficult to detect cyogenetically,
can often be observed with the use of pulsed-field gel electrophoresis (PFGE).
Experiments described here search for similar abnormalities in the germline DNA of
individuals affected with breast cancer using this technique.

Searches for smaller constitutional rearrangements within the DNA of affected
individuals using conventional Southern analyses have sometimes identified the disease
genes themselves. The detection by Southern analysis of deletions spanning the
Duchenne muscular dystrophy locus within the DNA of affected DMD males quickly led
to the identification of the gene responsible (Monaco et al. 1985; Monaco et al. 1986).
Similarly, the detection of interstitial deletions in WT patients by transcripts from the WT
region identified the WT-1 gene (Call et al. 1990). Transcripts from the NF1 genes were
also identified by Southern analysis (Wallace et al. 1990). The main advantage of this
method is that no detailed knowledge of the structure and sequence of a gene is required
and a preliminary screen can be carried out with a probe of interest immediately after its
isolation. Deletions and insertions may be detected by the presence of junction fragments
or changes in band intensities in the resulting autoradiographs. Point mutations may also
be detectable if they alter restriction sites. Hence, as part of the first step in the
mutational analysis stage of this project, a series of Southern analyses using cDNA
probes and anonymous DNA fragments from the region was carried out in an attempt to
Mutational analysis

discover any small chromosomal abnormalities which may have caused oncogenesis in one or two individuals (section 7.2).

Although the above methods presented a quick way of scanning the BRCA1 region for disease-causing chromosomal aberrations, they were unlikely to detect subtle single base changes. The technique of single-stranded conformation polymorphism analysis (SSCP) (Orita et al. 1989) was therefore employed to search the genes throughout the BRCA1 region for single base alterations which might have a disruptive effect on protein function. The work described in section 7.3 involves the investigation for such alterations within the candidate VH1-related phosphatase gene, HDSPH.

7.1 Searching for gross mutations by PFGE analysis of genomic DNA from affected family members

During the course of this project, several probes suitable for genomic Southern analyses were developed. They consisted of anonymous DNA fragments from cosmid, YAC and PAC clones as well as inserts from cDNA clones. Each of these probes was hybridised to PFGE Southern filters containing rare-cutter digested genomic DNA from the following samples:

(i) Genomic DNA from the lymphoblastoid cell lines derived from 43 affected individuals with breast cancer from 17q-linked families (Feuteun et al. 1993) (kindly provided by G. Lenoir, International Agency for Research on Cancer, Lyon)

(ii) Genomic DNA from the lymphoblastoid cell lines derived from 9 individuals from 5 17q-linked families, including BOV3, BC3 and BOV2 (Spurr et al. 1993) collected by the ICRF.

(iii) Genomic DNA from a selection of breast tumour cell lines.

(iv) Genomic DNA from the lymphoblastoid cell lines from random unaffected individuals with no history of cancer (provided by Centre d'Etude Polymorphisme Humain (CEPH)).

Probes from throughout the 3.5Mbp region from RARA to D17S78 were hybridised to NotI, MluI, BssHII, EagI and SacII digested genomic DNA. The majority resulted in a monomorphic pattern of hybridisation (as shown in Fig. 7.1(i) and Fig. 7.1(ii)), with the
Fig 7.1(i) Searching for cytogenetic rearrangements within the \textit{BRCA1} region by PFGE analysis. Panels 1-3 show the results of Southern analysis of rare-cutter digested genomic DNA samples with the 855RF probe, situated at the \textit{BRCA1} locus (see map).

Panels 1+2: autoradiographs resulting from hybridisation of PFGE Southern containing the same samples from affected members of breast/ovarian cancer families (see (i) text) digested with \textit{NotI} (1) and \textit{BssHII} (2) failed to reveal any band shifts indicative of cytogenetic rearrangements.

Panel 3: autoradiographs resulting from hybridisation of PFGE Southern containing a different set of samples from breast/ovarian cancer families (see (ii) text) digested with \textit{MluI} (a) and \textit{NotI} (b) also failed to show any aberrant bands.
Fig 7.1(ii) Searching for cytogenetic rearrangements by PFGE analysis.

Southern analysis of genomic DNA samples (see text) using probes from \textit{EDH} and \textit{PPY}, now known to reside either side of \textit{BRCA1} (see map), failed to reveal any aberrant bands, as shown in panels 1-3. The same samples were also digested with other enzymes and hybridised with the both \textit{PPY} and \textit{EDH} with similar results (not shown).

nc, normal controls; *, DNA samples from affected individuals from small breast cancer families.
exception of \textit{1A1.3B} and \textit{RNU2} probes which detected a distinctive polymorphism, as
detailed in chapter 4. Occasional differences in the sizes of hybridising bands between
different cell lines were observed but appeared to be due to differences in methylation
and occurred in the normal control samples as well as in the affected individuals.

7.2 Rapid surveying for mutations by conventional Southern analysis

Transcript probes from a number of genes within the \textit{BRCA1} region were hybridised to
digested genomic DNA from breast tumour cell lines and \textit{17q}-linked affected individuals.
Hybridisations of these probes to \textit{PstI}, \textit{XbaI}, \textit{EcoRI}, \textit{TaqI}, and \textit{BamHI}- digested DNA
detected a monomorphic pattern of hybridisation, as shown in Fig. 7.2(i) and Fig. 7.2(ii).

7.3 Analysis of the candidate gene \textit{HDSPH} for causative mutations

Since Southern hybridisation techniques failed to identify any abnormalities within this
gene in the genomic DNA of affected breast cancer family members or breast tumour cell
lines, it was decided use single stranded conformation polymorphism analysis to search
for subtle mutations which may have been overlooked. Preliminary mutational analysis
was performed on genomic DNA from affected members from \textit{17q} linked breast/ovarian
cancer families (see section 7.1(ii) above) and on the genomic DNA from a number of
individuals from smaller breast cancer families in which linkage had not been established.
In addition, a number of breast tumour DNA samples and paired tumour/blood DNA
samples were also analysed using this technique.

To enable SSCP analysis to be performed on genomic DNA, the intron/exon structure of
the gene had to be determined and primers flanking each exon designed. This would
facilitate the analysis of each entire exon and the intron/exon splice sites. Therefore, the
genomic region from the \textit{HDSPH} locus was cloned into a plasmid vector. The cosmid
\textit{B07165}, which contains the \textit{HDSPH} gene was digested with \textit{BamHI} and \textit{EcoRI} and the
entire digestion mixture was subcloned into pBluescript vector (see materials and
methods). Transformed colonies were screened with insert probes from P62.a2 and
\textit{I191.1} cDNAs (see section 6.2). Sequence analysis of the positive genomic subclones
was carried out by Fiona Harris at the CRC Beatson Institute. Comparison of the
genomic subclone sequences to the \textit{HDSPH} cDNA sequence (Fig. 6.2.2) revealed the
sites of intron/exon boundaries (Fig. 6.2.1(ii)). Primers were designed from the introns to
enable each exon to be amplified such that intron/exon junctions would be included in the
PCR product. The primers and PCR conditions are shown below in table 7.3 below.
Fig 7.2 Mutational analysis by conventional Southern analysis (i) HDSPH gene. A probe from the 3' region of the gene was used to detect rearrangements within the gene (see map) in the genomic DNA of affected members of chromosome 17q-linked breast/ovarian cancer families. Autoradiograph results from hybridisation of Southern containing PstI and XbaI-digested DNA shown above failed to show any indication of chromosomal rearrangements in this region within these samples.
Fig 7.2(ii) Mutational analysis by conventional Southern analysis (ii) 1A1.3B and IGFBP-4 genes. Probes from throughout both genes were used to hunt for chromosomal rearrangements within the genomic DNA of breast tumour cell lines. Hybridisations of Southern blots containing the samples digested with a range of restriction enzymes failed to show any aberrant bands when compared to the normal controls (nc), as shown.
### Table 7.3 PCR primers and conditions for analysis of HDSPH

<table>
<thead>
<tr>
<th>Exon</th>
<th>Primer set</th>
<th>PCR conditions</th>
<th>Product size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HDP-W5</td>
<td>standard conditions (55°C annealing temp) see chapter 2</td>
<td>185 bp</td>
</tr>
<tr>
<td></td>
<td>5' CAGAATGACCCGGACGCCTA 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HDP-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5' CATGTCGGGCTGTCGACGC 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>HDP-Y5</td>
<td>as above</td>
<td>190 bp</td>
</tr>
<tr>
<td></td>
<td>5' GCCTCCCATCTTTTCTGGT 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HDP-Y3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5' CCCACGGCGGATCCCCCAT3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>HDPZ1</td>
<td>as above</td>
<td>200 bp</td>
</tr>
<tr>
<td></td>
<td>5' CCCACGGACCTCTGCAGCGGAG 3'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysis of exon 3 of the *HDSPH* gene in the genomic DNA of affected family members and in 84 tumour DNA samples failed to reveal any band shifts indicative of single base pair alterations (Fig 7.3). Preliminary and less complete analysis of exons 1 and 2 also failed to show any differences (data not shown). However, a number of tumour samples appeared to show absence of one or two bands or even a total absence of the PCR-SSCP product altogether (Fig. 7.3). The PCR-SSCP analysis was repeated in these tumour samples and compared to the SSCP analysis of the corresponding genomic DNA derived from the lymphoblasts of the same individual. In each case, the tumour DNA sample lacked bands that were present in its blood DNA counterpart (Fig. 7.3).
affected individuals from breast/ovarian cancer families (see (ii), text) and small breast only cancer families

---

Fig 7.3 Genomic SSCP analysis of the VH1-related VHR phosphatase gene within the DNA of affected members of breast/ovarian cancer families and early onset breast tumours. No aberrant bands were visible in the germline DNA of affected family members (panel A). A number of tumour DNA samples appeared to lack bands (see arrows, panels B and C), indicative of allele loss or a mutation within the gene sequence. Analysis of paired tumour/blood DNA samples revealed abnormalities in tumour DNA which were not visible in their blood DNA counterparts (panel C). (-), negative control (no DNA); nc, normal control.

---
7.4 Discussion of the mutation analysis results

The search for gross chromosomal abnormalities by the PFGE studies detailed in section 7.1 failed to find any aberrant hybridising bands indicative of large regions of deletions, translocations or other large chromosomal defects. This was typical of the findings of other groups which indicated that the genetic defects causing breast and ovarian tumorigenesis were more subtle. Southern analysis carried out in this work also failed to reveal any defects within the genomic DNA of chromosome 17q-linked individuals using probes from throughout the IGFBP-4, 1A1.3B and HDSPH genes. These genes have since been "eliminated" as candidates for BRCA1 by extensive mutational analysis using PCR-SSCP and sequencing methods (Nicolai et al in preparation; (Campbell et al. 1994; Kamb et al. 1994) and section 7.3 results).

It is unlikely that Southern analysis techniques will be applied to the mutational analysis of the recently published BRCA1 gene, since mutations in it appear to involve subtle alterations such as single base pair deletions and insertions which would be overlooked (Castilla et al. 1994; Friedman et al. 1994; Futreal et al. 1994; Simard et al. 1994; Shattuck-Eidens et al. 1995).

The analysis of the HDSPH gene by SSCP was halted following the publication of the BRCA1 gene in October, 1994 (Miki et al. 1994). At the same time, another paper was published by the same group which detailed the mutational analysis of the HDSPH gene in the cDNA and genomic DNA of affected individuals carrying BRCA1 predisposing alleles from 8 breast cancer families, 24 sporadic breast tumours and 12 breast cancer cell lines (Kamb et al. 1994). The preliminary mutational analysis detailed in this chapter is consistent with the published results which suggested this gene is not altered in breast tumours and is not the cause of breast tumorigenesis in the majority of familial breast cancer cases.

The most likely explanation for the missing bands observed in the tumour samples (Fig. 7.3) is that the DNA concentration in these tumour samples is different from the rest, resulting in the production of different conformers. The total absence of product in one of the samples is most likely to be due to an extremely low DNA concentration. The concentrations of these DNA samples should be re-assessed and the PCR-SSCP repeated if the concentration needs to be adjusted. In addition, a control PCR-SSCP analysis could be performed on the entire sample set using primers from elsewhere in the genome to check whether these abnormalities are confined to the BRCA1 region.
A second explanation for the absent band observed in the tumour samples could be that one allele of the *HDSPH* gene is missing as part of a large region of allele loss within these tumours. A study of loss of heterozygosity in these tumour samples using polymorphic markers throughout the *BRCA1* region would have to be carried out to confirm this. In addition, sequence analysis would have to be performed on each allele at the *HDSPH* locus within these tumour samples in order to ascertain whether the absent bands revealed in the SSCP analysis were due instead to mutations within the *HDSPH* gene.

Further investigation of the *HDSPH* gene is required to complete the mutational analysis and to be certain that *HDSPH* does not play a role in tumorigenesis or tumour development. The SSCP analysis detailed here and in the published work (Kamb et al. 1994) have employed only a single set of SSCP running conditions. Previous investigations of this technique have clearly demonstrated a variation in sensitivity which correlates with alterations in the conditions of the SSCP analysis (Michaud et al. 1992; Grompe 1993). Hence, further analysis of this gene would have to be performed under a range of different conditions in all the available samples. This would also have to include an analysis of both 5' and 3' untranslated regions to check for alterations which might have an effect on expression, processing, stability or localisation of the protein product.
CHAPTER 8 DISCUSSION

The work described in this thesis presents a characterisation of the chromosomal region in which the familial breast/ovarian cancer gene, \textit{BRCAl} is located. The information provided in this work has been important in the positional cloning of this much-publicised gene and has influenced the work of others both in the laboratory and in other groups in the field. The physical mapping data presented here represents a refinement of the mapping efforts published so far. It has aided the construction of genomic clone contigs and has provided some insight into the physical features of the genomic region surrounding \textit{BRCAl} which may have a bearing on \textit{BRCAl}-mediated oncogenesis. The gene isolation and localisation work detailed here contributes further information on the chromosomal region and has identified a number of genes which are worthy of further investigation in their own right.

The isolation of \textit{BRCAl} was published at the end of this thesis work in October 1994 by a large collaboration of groups headed by Dr Mark Skolnick (Miki \textit{et al.} 1994). This group employed the same strategy as described in this thesis but they had the added advantages of a far greater source of genetic data from extensive and well-characterised pedigrees and a large number of people and resources employed on the project. Two recombinations which were not made available, were identified in their families which further narrowed the \textit{BRCAl} region from 1.3Mb to 600kb between the markers \textit{D17S1321} and \textit{D17S1327} (Fig. 1.5.1(i)). The group concentrated on this small region and by a "brute force" effort, isolating every gene localised from the area through direct selection and exon-trapping techniques and analysing each one by SSCP and sequencing, eventually identified the culprit. It was interesting to note that the efforts of our laboratory had succeeded in isolating \textit{BRCAl}, but the analysis of this gene was still only in the preliminary stages (Brown \textit{et al.} in press).

\textbf{The \textit{BRCAl} gene}

Preliminary analysis of this novel gene (Miki \textit{et al.} 1994) has revealed interesting facts which question previous hypotheses regarding the gene's function and its role in oncogenesis.

Many of the previously isolated cancer genes have been shown to be transcription factors (see table 1.2.1). One expectation is that the \textit{BRCAl} protein too interacts directly with DNA and controls transcription. An initial analysis of the \textit{BRCAl} cDNA sequence at the NH$_2$ terminus revealed it corresponded to a RING finger, a motif related to the zinc...
finger (reviewed (Freemont 1993)). Other proteins containing this motif are involved in the regulation of development and cell differentiation. Knowing the role of zinc fingers in DNA binding and their abundance in transcription factors, this implicated BRCA1 as a putative transcription factor. However, studies of other proteins containing the RING finger motif have failed to show evidence of specific DNA binding (Freemont 1993).

However, zinc-dependent DNA binding by RING fingers have been observed \textit{in vitro} (reviewed (Freemont 1993)). Other evidence exists which suggests the RING finger may also be involved in protein-protein interaction. Obviously, further work is required to elucidate the role of the RING finger in BRCA1 and to establish whether it does have a DNA binding function.

\textbf{Mutational analysis}

The initial mutational analysis of this gene revealed five sequence alterations in eight kindreds (Miki \textit{et al.} 1994). Subsequent analyses of breast and ovarian kindreds identified an inherited germline mutation in approximately 50\% of cases (Simard \textit{et al.} 1993; Castilla \textit{et al.} 1994; Friedman \textit{et al.} 1994)\textit{Castilla, 1994 #266; Friedman, 1994 #291; Simard, 1994 #340}. A recent collaborative study of 327 unrelated patients from breast/ovarian families described 38 distinct BRCA1 mutations in 80 individuals (Shattuck-Eidens \textit{et al.} 1995) An initial survey of the BRCA1 mutations appeared to suggest an almost uniform distribution throughout the entire gene, with the possible exception of a slight clustering around the RING finger. These findings suggest that the very 3' end of BRCA1 in addition to the RING finger at the 5' region play important roles in normal gene function. Interestingly, the mutation studies indicated a number of common mutations, found to be present on the same haplotype, suggesting common ancestry (Simard \textit{et al.} 1994; Black \textit{et al} unpublished data; Goldgar pers comm).

The majority of the mutations resulted in the loss or premature termination of protein synthesis. These observations of "loss of function" mutations indicated a tumour suppressor role for BRCA1 and supported the "two hit" model of oncogenesis, proposed following frequent observations of allelic loss in the BRCA1 region within tumours and the finding that the LOH in familial tumours was attributable to loss of the wild-type allele (Simard \textit{et al.} 1993). However, the expectation that BRCA1 would also play a role in the more common, sporadic form of the disease was not borne out. A study of 44 breast tumours exhibiting allelic loss of the BRCA1 region revealed only 4 BRCA1 mutations. Re-analysis of these samples together with DNA from blood samples of the same individuals revealed these mutations were germline and not somatic. Another study
Discussion

of over 150 breast tumour samples has also failed to find any somatic BRCAI mutations (Nicolai unpublished data). An obvious interpretation of these findings is that BRCAI does not play a role in the great majority of breast cancers and that there is another gene on chromosome 17q that does. Evidence in support of this comes from several allelic loss studies which have suggested the presence of another breast cancer gene on chromosome 17q distal to BRCAI (Saito and Streuli 1991; Eccles et al. 1992; Cornelis et al. 1993; Jacobs et al. 1993; Lindblom et al. 1993). However, further large-scale mutational analysis of sporadic breast tumours which employ a range of techniques will have to be performed to be certain that the lack of discoveries of mutations in sporadic tumours is not due simply to a lack of sensitivity in the mutation hunting techniques used.

Another interesting discovery regarding the mutational analysis of BRCAI is the finding of small numbers of ovarian tumours containing sporadic BRCAI mutations (Hosking et al. 1995; Merajver et al. 1995). This presents the possibility that BRCAI-associated tumorigenesis is a complex process moderated differently in breast and ovarian tissue.

Findings of this thesis

The physical map of the BRCAI region described in chapter 3 together with the identification and localisation of several novel and previously unmapped genes throughout the region (chapter 6) have provided a detailed physical characterisation of the region surrounding BRCAI. During the course of this work, several interesting features of the region were discovered which may be involved in breast and ovarian tumorigenesis and development.

1A1.3B and BRCAI

One such finding was the location of exon-trap products and genes around the D17S855 marker, now known as the location of BRCAI. Genomic PFGE analysis of the region suggested the exon trap product ET A37, subsequently found to correspond to exon 13 of BRCAI, and 1A1.3B were in close proximity (see section 3.3.2) (Brown et al. in press). This finding sparked further investigation of the region involving Southern, PCR and sequence analysis of BRCAI- and 1A1.3B-containing PAC and cosmid clones. These studies, carried out by other members of the laboratory, revealed that BRCAI and 1A1.3B are extremely closely situated, with the putative first exon of BRCAI lying head to head with the 5' end of the 1A1.3B gene with a maximum distance of 295 base pairs between them (Brown et al. 1994). Work is now underway to further characterise the 5' regions of
these two genes and to discover whether the promoter and enhancer motifs found in the small region between them function to coregulate their expression.

A model of dis-coordinate expression was proposed which predicted an increased expression of $1A1.3B$ in tumours with a corresponding decrease in $BRCA1$ expression. This effective down-regulation of $BRCA1$ would be consistent with a tumour suppressor model for this gene and could result in tumorigenesis without a requirement for mutations within the coding region of $BRCA1$. This model could therefore provide an explanation for the absence of $BRCA1$ mutations in the majority of breast and ovarian tumours. Further work is now being carried out in the laboratory to investigate the 5' end of $BRCA1$ and the mechanisms of regulation of $BRCA1$ and $1A1.3B$ in normal and tumour tissue.

Other genes identified in the region surrounding $BRCA1$

During the search for $BRCA1$, a number of genes were mapped to the region (see chapter 6). Several of these genes have potential to cause tumorigenesis or aid tumour development. Loss of function mutations within the dual-specificity phosphatase gene could have promoting effects on tumour development by allowing the uninhibited activity of tyrosine and serine kinases. In addition, several novel genes which could potentially have tumour suppressing functions were located in the $BRCA1$ region. Since these genes all reside within a region which is commonly deleted in breast and ovarian tumours, the implications for tumorigenesis and development must be investigated. It would be useful to assess the expression levels of these genes in breast and ovarian tumours in comparison to normal tissue to determine whether any loss of function has occurred which could have an effect on tumorigenicity.

The observed extensive variation of the $RNU2$ locus (detailed in chapter 4) is worthy of further investigation in the light of previous studies which have suggested this locus and the sites of other snRNA genes are responsible for the specificity of virally induced chromosome fragility (Lindgren et al. 1985) (discussed (Pavelitz et al. 1995)). Hence, an analysis of $RNU2$ to assess its effects on the surrounding chromosomal stability may reveal whether it plays any role in tumour development.
Future work regarding BRCA1

The discovery of BRCA1 and preliminary investigations concerning this gene have highlighted a need for a number of areas of research. Now that the gene is known, a large-scale epidemiological study can be carried out to determine the extent of BRCA1-mediated oncogenesis in the population. The structure and function of this novel gene and its involvement in cell growth needs to be elucidated in order to determine the mechanisms of BRCA1-mediated oncogenesis. Such studies may help to reveal the molecular basis for the heterogeneity of penetrance, organ of origin of the cancer, age of onset and laterality of the lesions. Once these aspects are better understood, improved treatment regimes and novel therapies can be designed.

- **Mutational analysis and future wide-scale screening.** Further wide-scale mutational analysis is required in order to build an accurate picture of the spectrum of BRCA1 mutations in breast and ovarian tumours. This will aid genotype/phenotype correlations and provide recommendations for suitable future screening procedures.

The observation of a large number of different mutations almost uniformly distributed throughout BRCA1, coupled with the extensive heterogeneity of the disease suggests the development of a diagnostic test for the wide-scale screening for BRCA1 mutations will not be easy (discussed (Vogelstein and Kinzler 1994)). A sensitive, quick and economically viable testing protocol will have to be devised before the mutational analysis is extended to the at-risk population in general. One possibility is an adaptation of the protein truncation test (PTT) (Roest et al. 1993). This test is combines RT-PCR, transcription and translation and selectively detects translation-terminating mutations. It is capable of rapidly detecting gross rearrangements, interstitial gene deletions and duplications as well as mutations affecting splicing and point mutations resulting in early termination within large multi-exonic genes. This approach has proved successful in identifying NF1 mutations (Heim et al. 1994). However, refinement of the technique is required to enable it to reliably detect missense mutations.

- **Structure of the BRCA1 gene.** An indepth analysis of the protein structure of the BRCA1 product is required to help determine the molecular consequences of the different mutations observed throughout the gene and to provide clues to its function.

- **Interaction of the BRCA1 protein with other cellular components.** Studies are needed to elucidate the function of the RING finger and to establish whether it does
bind DNA. If, as is suspected, it is involved in protein-protein interaction, the proteins which interact with the \textit{BRCA1} product must be isolated and characterised and the consequences of their interactions determined.

- **Regulation of \textit{BRCA1}**. Further work is required to determine the mechanisms of \textit{BRCA1}-mediated oncogenesis and to discover whether \textit{1A1.3B} plays a role in the process. This will involve RNase protection experiments to determine the exact locations of transcriptional start sites of these genes and also their expression patterns in breast and ovarian tumours. In addition, it would be interesting to investigate in further detail the promotor elements governing \textit{BRCA1} and \textit{1A1.3B} expression and to establish the effects of their manipulation in tumour cell lines.

The physiological significance, if any, of the alternate splicing of this gene also needs to be addressed. It will be interesting to investigate whether it regulates the function of the RING finger in a manner similar to \textit{WT-1}, in which alternate splicing of exons near the zinc finger domain alter its the DNA binding properties (Bickmore \textit{et al.} 1992).

- **Mouse studies**. Recent studies have shown \textit{BRCA1} and its surrounding genes to be conserved in the mouse (B. Griffiths unpublished data; P Futreal pers comm). This will allow comparisons to be made between the human and mouse \textit{BRCA1} gene structure. It also presents the possibility of creating mouse models to investigate, for example, the effects of introducing \textit{BRCA1} mutations, or the removal of the \textit{BRCA1} gene altogether.

**The future of positional cloning**

The work described in this thesis was part of a large positional cloning effort aimed at the isolation of the breast and ovarian cancer susceptibility gene, \textit{BRCA1}. Like other positional cloning projects, this involved initial chromosomal assignment of the gene by linkage analysis followed by successive narrowing of the candidate interval through the identification of sites of recombination in families in which the disease gene was segregating.

This approach may not be necessary in the hunt for disease genes in the future. A few disease genes have already been identified by a "candidate gene" approach whereby a previously isolated gene bearing the expected traits of the disease gene is located in the chromosomal region showing linkage to the disease. Investigation of this gene for
disease-causing mutations has revealed it to be the culprit locus. In these cases, construction of a physical map, the compilation of genomic clone contigs and involved transcriptional analyses was not necessary.

The association between the TP53 gene and the Li-Fraumeni cancer syndrome (Malkin et al 1990) was discovered in this way. Other genes responsible for other diseases have been identified through a combination of positional and functional approaches. The recent identification of the HNPCC locus, hMSH2 is one example (Fishel et al. 1993; Leach et al. 1993).

It is likely that the accumulation of information on genes from a diverse range of sources into readily accessible centralised databases will accelerate the identification of the genes responsible for other diseases. Together with resources such as readily available genomic clones spanning the genome, vast numbers of polymorphic markers and well defined physical and genetic maps, which are being developed by the Human Genome Project, an increasing number of disease genes will be identified without a requirement for exhaustive genomic clone contig construction and transcriptional analyses.
REFERENCES


differentially expressed during early mesodermal patterning in mouse embryos.

*Development** 116: 1123-1126.


polypeptide (PPY) and the anonymous marker D17S78 are within 45kb of each other on chromosome 17q21. *Genomics* **21**: 458-460.


chromosome 17q in 109 breast carcinomas indicate at least two distinct target regions. Oncogene 8: 781-785.


References


References


References


References


References


References


