ANALYSIS OF THE BREAST CANCER SUSCEPTIBILITY GENE (BRCA1) REGION IN HUMAN AND MOUSE

by

Julie Anne Chambers

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

Somatic Cell Genetics
Imperial Cancer Research Fund
London

Cancer Genetics Laboratory
Division of Medical and Molecular Genetics
Guy's Hospital
London

and

Department of Genetics and Biometry
University College
London

March 1998
ABSTRACT

Breast cancer is one of the most common diseases affecting women. A human breast cancer susceptibility gene, BRCA1, was localised to chromosome 17q in 1990, and worldwide efforts have focused on the isolation of this gene. This thesis describes the identification and characterisation of candidate genes within the region, using a number of different techniques. Following the isolation of the BRCA1 gene, breast and breast/ovarian cancer families were analysed for mutations in this gene, either by direct sequencing of the entire coding region or by screening for specific mutations using allele specific oligonucleotide hybridisation. These techniques identified mutations in the BRCA1 gene in five families.

The BRCA1 gene lies within a duplicated region on chromosome 17. Pseudo-copies of the 5' exons of BRCA1 lie head to head with the NBR1 gene, whose function is unknown. Analyses of the BRCA1 and NBR1 genes in other species, in particular the mouse, are presented. Isolation of murine cDNA clones shows that NBR1 is highly conserved between the human and the mouse, whereas BRCA1 is less well conserved, but has several domains that are highly homologous with the human. The structure of the NBR1 gene is also well conserved, with exon/intron boundaries being largely the same in the two species. Analysis of mouse genomic clones reveals that this region is not duplicated in the mouse and that the Brca1 and Nbr1 genes lie head to head less than 1 kb apart. Studies on this promoter region, using a luciferase reporter gene assay system, are presented and suggest that the Brca1 gene and the Nbr1 gene may share a bidirectional promoter.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>1</td>
</tr>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>3</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>11</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>14</td>
</tr>
<tr>
<td><strong>Chapter One: Introduction</strong></td>
<td>15</td>
</tr>
<tr>
<td>1.1 The multistep nature of cancer</td>
<td>15</td>
</tr>
<tr>
<td>1.2 Inherited susceptibility to cancer</td>
<td>17</td>
</tr>
<tr>
<td>1.3 Breast cancer susceptibility genes</td>
<td>21</td>
</tr>
<tr>
<td>1.3.1 The \textit{BRCA1} gene</td>
<td>22</td>
</tr>
<tr>
<td>1.3.2 The \textit{BRCA2} gene</td>
<td>22</td>
</tr>
<tr>
<td>1.3.3 The \textit{TP53} gene</td>
<td>23</td>
</tr>
<tr>
<td>1.3.4 The \textit{ATM} gene</td>
<td>23</td>
</tr>
<tr>
<td>1.3.5 The \textit{PTEN/MMAC1} gene</td>
<td>24</td>
</tr>
<tr>
<td>1.3.6 Male breast cancer susceptibility genes</td>
<td>24</td>
</tr>
<tr>
<td>1.4 Positional cloning of the \textit{BRCA1} gene</td>
<td>25</td>
</tr>
<tr>
<td>1.4.1 Linkage analysis</td>
<td>25</td>
</tr>
<tr>
<td>1.4.2 Analysis of known genes within the region</td>
<td>28</td>
</tr>
<tr>
<td>1.4.3 Creation of a physical map</td>
<td>28</td>
</tr>
<tr>
<td>1.4.4 Construction of a transcript map</td>
<td>28</td>
</tr>
<tr>
<td>1.5 The \textit{BRCA1} gene</td>
<td>33</td>
</tr>
<tr>
<td>1.5.1 Features of the \textit{BRCA1} gene</td>
<td>34</td>
</tr>
<tr>
<td>1.5.2 Mutation analysis of the \textit{BRCA1} gene</td>
<td>34</td>
</tr>
<tr>
<td>1.5.3 Clinical features of \textit{BRCA1} tumours</td>
<td>41</td>
</tr>
<tr>
<td>1.5.4 The role of \textit{BRCA1} in sporadic breast and ovarian cancer</td>
<td>41</td>
</tr>
<tr>
<td>1.5.5 Structure of the \textit{BRCA1} genomic region</td>
<td>44</td>
</tr>
<tr>
<td>1.5.6 Regulation of \textit{BRCA1} gene expression</td>
<td>44</td>
</tr>
<tr>
<td>1.6 The mouse \textit{Brca1} gene</td>
<td>48</td>
</tr>
<tr>
<td>1.6.1 Conservation of the \textit{BRCA1} gene in mice</td>
<td>50</td>
</tr>
<tr>
<td>1.6.2 Animal models of \textit{BRCA1}-induced tumorigenesis</td>
<td>51</td>
</tr>
</tbody>
</table>
1.7 Functions of the BRCA1 gene ................................................................. 51
  1.7.1 Subcellular localisation of the BRCA1 protein .............................. 52
  1.7.2 The BRCA1 RING finger may act as a protein-protein
      interaction domain ............................................................................. 53
  1.7.3 A role for BRCA1 in transcription ............................................. 55
  1.7.4 BRCA1 and cell cycle regulation .............................................. 55

Aims Of This Thesis ....................................................................................... 60

Chapter Two: Materials And Methods ....................................................... 61
  2.1 Preparation of DNA .......................................................................... 61
      2.1.1 Preparation of genomic DNA from blood samples and
           cell lines ...................................................................................... 61
      2.1.2 Preparation of genomic DNA from tissue samples .......... 62
      2.1.3 Preparation of plasmid/cosmid DNA by alkaline lysis ....... 62
      2.1.4 Preparation of high quality plasmid/cosmid DNA .......... 63
      2.1.5 Small-scale preparation of plasmid/cosmid DNA .............. 63
      2.1.6 Preparation of bacteriophage lambda DNA ...................... 63
      2.1.7 Preparation of oligonucleotides .......................................... 64
  2.2 Preparation of RNA .......................................................................... 64
      2.2.1 Preparation of RNA from cell lines .................................. 64
      2.2.2 Preparation of RNA from whole tissues ............................ 64
  2.3 Quantitation of Nucleic Acids .......................................................... 65
      2.3.1 Quantitation using a spectrophotometer ............................ 65
      2.3.2 Quantitation by gel analysis .............................................. 65
  2.4 Size separation of DNA by agarose gel electrophoresis............... 65
  2.5 Restriction enzyme digestion of DNA ............................................. 66
  2.6 Amplification of DNA by the Polymerase Chain Reaction (PCR).... 66
      2.6.1 PCR reaction ......................................................................... 66
      2.6.2 Purification of PCR products ............................................ 67
  2.7 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) ....... 67
  2.8 Hybridisation of DNA to Southern blots ........................................ 68
      2.8.1 Southern Transfer of DNA .............................................. 68
      2.8.2 Radioactive labelling of DNA probes ............................... 68
      2.8.3 Prehybridisation and hybridisation of DNA probes to
           filters ......................................................................................... 69
      2.8.4 Prehybridisation and hybridisation of oligonucleotide
           probes to filters ....................................................................... 70
Chapter Three: Isolation Of Candidate Genes Within The \textit{BRCA1} Region On Chromosome 17q

3.1 Identification of transcripts by searching for cross-species homology

3.1.1 Cosmid C1198 transcripts
3.1.2 Cosmid G01152 transcripts ........................................................87
3.1.3 Cosmid B07165 transcripts .........................................................90
3.1.4 Cosmid G1151 transcripts ..........................................................90
3.1.5 Cosmid A1028 transcripts ..........................................................93
3.2 Analysis of clones from mini-libraries created by direct selection .................................................................94
  3.2.1 Transcripts from cosmids F0520 and F0345 ............................96
  3.2.2 Cosmid G124 transcripts ............................................................96
  3.2.3 Cosmid C02179 transcripts ........................................................99
3.3 Summary of transcripts identified .......................................................99

Chapter Four: Mutation Analysis Of The BRCA1 Gene In Breast And
Breast/Ovarian Cancer Families ...........................................................................102
  4.1 Mutation analysis in breast and breast/ovarian cancer families with strong evidence for linkage to BRCA1 .........................................................102
  4.2 Screening for common BRCA1 mutations in breast and breast/ovarian cancer families .................................................................107
  4.3 Summary of mutations identified in breast and breast/ovarian cancer families ..................................................................................................109

Chapter Five: Isolation Of The Murine Homologues Of The BRCA1 And
NBR1 Genes .................................................................................................................112
  5.1 Conservation of the BRCA1 and NBR1 genes in other species............112
  5.2 Isolation of the murine homologue of the NBR1 gene .......................114
    5.2.1 Isolation of murine Nbr1 cDNA clones ......................................114
    5.2.2 Isolation of murine Nbr1 genomic clones .................................116
    5.2.3 Comparison of the human and mouse NBR1 genes ....................121
  5.3 Isolation of the murine homologue of the BRCA1 gene ....................124
    5.3.1 Isolation of murine Brca1 cDNA clones .................................124
    5.3.2 Isolation of genomic clones of the murine Brca1 gene ............126
    5.3.3 Comparison of the human and mouse BRCA1 genes ..............128
  5.4 Summary of the conservation observed between the human and
mouse BRCA1 and NBR1 genes .......................................................................128

Chapter Six: Analysis Of The Promoter Regions Of The Murine Brca1
And Nbr1 Genes .................................................................................................131
  6.1 Localisation of the murine Nbr1 gene adjacent to the Brca1 gene ......131
6.2 Analysis of the intergenic region between the Brca1 and Nbr1 genes................................................................................................................... 133
6.3 Analysis of the activities of the murine Brca1 and Nbr1 gene promoters using a luciferase reporter gene assay system...................... 138
6.4 Summary of the results from studying the Brca1/Nbr1 intergenic region ................................................................................................................. 142

Chapter Seven: Discussion ........................................................................................ 144
7.1 Findings of this thesis ................................................................................ 144
  7.1.1 Isolation of candidate cDNAs within the BRCA1 gene region .............. 144
  7.1.2 Mutation analysis of the BRCA1 gene .............................................. 145
  7.1.3 Conservation of the BRCA1 region in the mouse ............................ 146
  7.1.4 Analysis of the promoter regions of the Brca1 and Nbr1 genes .......... 147
7.2 Implications for future work ............................................................................ 149

Appendix A: Primers used for PCR and sequencing during the mutation analysis of the BRCA1 gene in breast and breast/ovarian cancer families ....... 152

Appendix B: Nucleotide sequence of the murine Nbr1 mRNA (accession number U73039) ........................................................................................................... 155

Appendix C: Primers used for determining the mouse Nbr1 exon/intron structure .............................................................. 158

References ..................................................................................................................... 160
**LIST OF FIGURES**

**Figure 1.1** A model to illustrate the multistep nature of colorectal cancer       18
**Figure 1.2** Knudson's two hit model of tumorigenesis                        19
**Figure 1.3** Mechanisms of loss of heterozygosity                          20
**Figure 1.4** The positional cloning strategy for the isolation of a disease susceptibility gene                                             26
**Figure 1.5** Genetic map of the BRCA1 region                              27
**Figure 1.6** Method for generating a cDNA library by direct selection        31
**Figure 1.7** Schematic diagram summarising the reported germline mutations in the BRCA1 gene                                      38
**Figure 1.8** Schematic diagram of the genomic region between the BRCA1 and the NBR1 genes                                      45
**Figure 1.9** Schematic diagram of the BRCA1 cDNA illustrating the functional domains                                             54
**Figure 1.10** Regulation of BRCA1 during the cell cycle                     56
**Figure 3.1** Diagram showing the cosmid contig between the RNU2 locus and the PPY gene on chromosome 17q21                   86
**Figure 3.2** Hybridisation of cosmid fragments to zoo blots               89
**Figure 3.3** Hybridisation of cosmid fragments to multiple tissue Northern blots                                             91
**Figure 3.4** Schematic diagram illustrating the isolated cDNA clones from the DLG3 gene                                           95
**Figure 3.5** Hybridisation of GAP3 mini-library clones to multiple tissue Northern blots                                       98
**Figure 3.6** Transcript map of the region between the RNU2 locus and the PPY gene on chromosome 17                             100
**Figure 4.1** Identification of the 185delAG mutation in the BOV3 breast/ovarian cancer family by direct sequencing          105
**Figure 4.2** Pedigree of the BOV3 family indicating the individuals known to have the 185delAG mutant allele and those having the wild-type allele        106
**Figure 4.3** Results from screening for the 185delAG mutation by ASO       108
**Figure 5.1** Conservation of the BRCA1 and the NBR1 genes                 113
**Figure 5.2** Schematic diagram of the mouse Nbr1 cDNA illustrating the coverage of the isolated cDNA clones                     115
**Figure 5.3** Hybridisation of mouse Nbr1 to a Northern blot                117
Figure 5.4 Schematic diagram illustrating the genomic sequence covered by Nbr1 cosmid and bacteriophage lambda clones ................................................. 120
Figure 5.5 Comparison of the human and mouse NBR1 predicted amino acid sequences .............................................................................................................. 123
Figure 5.6 Comparison of the 5' ends of the human and mouse NBR1 genes ............................................................................................  125
Figure 5.7 Schematic diagram illustrating the genomic sequence covered by Brca1 cosmid and bacteriophage lambda clones ................................................ 127
Figure 5.8 Comparison of the human and mouse BRCA1 predicted amino acid sequences .............................................................................................................. 129
Figure 6.1 Partial map of cosmid MCHCA1, indicating the relative positions of the Brca1 and Nbr1 genes ................................................................................................. 132
Figure 6.2 Comparison of the BRCA1 genomic region in the human and the mouse ...................................................................................................................... 134
Figure 6.3 Analysis of the expression of Nbr1 exon 1a and exon 1b transcripts by RT-PCR ..................................................................................................................... 136
Figure 6.4 Diagram illustrating the sequence between the Brca1 and Nbr1 genes ............................................................................................................................... 137
Figure 6.5 Diagram illustrating the constructs used in the luciferase reporter gene assays ........................................................................................................ 140
Figure 6.6 Analysis of the Brca1 and Nbr1 promoter regions using a luciferase reporter gene assay system .............................................................................. 141
LIST OF TABLES

Table 1.1 Functions of tumour suppressor genes ................................................... 16
Table 1.2 Advantages and disadvantages of techniques for gene isolation ...... 32
Table 1.3 Human and mouse homologies for tumour suppressor genes .......... 49
Table 1.4 Similarities between the BRCA1 and BRCA2 genes......................... 58
Table 3.1 Summary of results of hybridisations to zoo blots ......................... 88
Table 3.2 Clones from the GAP3 mini-library that were picked for analysis ................................................................. 97
Table 4.1 Breast and breast/ovarian cancer families analysed by sequencing for BRCA1 mutations .................................................... 103
Table 4.2 BRCA1 mutations detected in ICRF families ................................ 110
Table 5.1 Exon/intron structure of the mouse Nbr1 gene ............................. 122
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>adenomatous polyposis coli</td>
</tr>
<tr>
<td>ASO</td>
<td>allele specific oligonucleotide</td>
</tr>
<tr>
<td>A-T</td>
<td>ataxia telangiectasia</td>
</tr>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CCM</td>
<td>chemical cleavage by mismatch</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DGGE</td>
<td>denaturing gradient gel electrophoresis</td>
</tr>
<tr>
<td>DHFR</td>
<td>dihydrofolate reductase</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene diamine tetra acetic acid</td>
</tr>
<tr>
<td>FAP</td>
<td>familial adenomatous polyposis</td>
</tr>
<tr>
<td>HA</td>
<td>heteroduplex analysis</td>
</tr>
<tr>
<td>HCl</td>
<td>hydrochloric acid</td>
</tr>
<tr>
<td>KCl</td>
<td>potassium chloride</td>
</tr>
<tr>
<td>K$_2$HPO$_4$</td>
<td>di-potassium hydrogen orthophosphate</td>
</tr>
<tr>
<td>KH$_2$PO$_4$</td>
<td>potassium di-hydrogen orthophosphate</td>
</tr>
<tr>
<td>KOAc</td>
<td>potassium acetate</td>
</tr>
<tr>
<td>LOH</td>
<td>loss of heterozygosity</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>Na$_2$HPO$_4$</td>
<td>di-sodium hydrogen orthophosphate</td>
</tr>
<tr>
<td>NaH$_2$PO$_4$</td>
<td>sodium di-hydrogen orthophosphate</td>
</tr>
<tr>
<td>NaOAc</td>
<td>sodium acetate</td>
</tr>
<tr>
<td>NaOH</td>
<td>sodium hydroxide</td>
</tr>
<tr>
<td>NF</td>
<td>neurofibromatosis</td>
</tr>
<tr>
<td>(NH$_4$)$_2$SO$_4$</td>
<td>ammonium sulphate</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localisation signal</td>
</tr>
<tr>
<td>MEN</td>
<td>multiple endocrine neoplasia</td>
</tr>
<tr>
<td>MgCl$_2$</td>
<td>magnesium chloride</td>
</tr>
<tr>
<td>MgSO$_4$.7H$_2$O</td>
<td>magnesium sulphate 7-hydrate</td>
</tr>
<tr>
<td>PAC</td>
<td>P1-derived artificial chromosome</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PSM</td>
<td>phage storage medium</td>
</tr>
<tr>
<td>PTT</td>
<td>protein truncation test</td>
</tr>
<tr>
<td>RACE</td>
<td>rapid amplification of cDNA ends</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SSCP</td>
<td>single strand conformation polymorphism analysis</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>VHL</td>
<td>von Hippel-Lindau</td>
</tr>
<tr>
<td>WT</td>
<td>Wilms' tumour</td>
</tr>
<tr>
<td>YAC</td>
<td>yeast artificial chromosome</td>
</tr>
</tbody>
</table>
Work in this thesis is included in the following publications:


ACKNOWLEDGEMENTS

I would like to thank my supervisor Ellen Solomon for all her advice and guidance throughout my thesis. I would also like to thank Sue Povey for her support.

I have been privileged to work in an excellent laboratory and I am immensely grateful to all members of the laboratory past and present: Karen Jones, Kathy Howe, Beatrice Griffiths, Fanli Xu, Melissa Brown, Hans Nicolai, David Grimwade, Aurelie Catteau, Lynne Davies, Caroline Whitehouse and Jo Selfe for sharing their expertise and for making it such an enjoyable time.

Finally, I would like to thank Mark for his continual encouragement and patience.
CHAPTER ONE: INTRODUCTION

Cancer results when a cell continues to divide and proliferate inappropriately, ignoring normal growth controls that are regulated by many genes with diverse functions. Cancer cells overcome these controls through mutations in the genes responsible for these controls. Hence, cancer is a genetic disease since it results when genetic events occur enabling the cell to overcome or lose the normal restraints. Genetic mutations leading to cancer can occur in a wide range of genes which are often categorised as either proto-oncogenes or tumour suppressor genes, depending on their normal function.

Proto-oncogenes have growth promoting functions in normal cells and when mutated they contribute to tumour formation through 'gain of function' mutations that cause overactivation of the gene and/or its product. Oncogenes can be activated by structural alterations, gene amplifications, or by loss of control mechanisms.

In contrast, tumour suppressor genes normally function as negative regulators of cell growth and differentiation. In tumours, they are inactivated as a result of 'loss of function' mutations. More than ten tumour suppressor genes have now been identified and have diverse functions ranging from a role in DNA repair (TP53) (Selivanova and Wiman, 1995) to cell structural components (NF2) (Thomas et al., 1994) (Table 1.1).

1.1 The multistep nature of cancer

It is now clear that many steps are required for the progression from normal tissue through benign hyperplasia to a metastatic tumour. This progression is caused by genetic events or mutations, in line with the observation that cancer incidence increases dramatically with age. The best-characterised example of this multistep nature is colon cancer (Vogelstein and Kinzler, 1993). Colon tumours evolve through defined morphological stages, and hence it has been possible to identify the mutations acquired at each step. Mutations in the APC gene are required to initiate tumour development and result in the formation of benign adenomas. If this is followed by mutations in other growth control genes such as the RAS, DCC and TP53 genes, benign adenomas will progress to carcinomas.
<table>
<thead>
<tr>
<th>Name of gene</th>
<th>Chromosomal location</th>
<th>Familial Syndrome</th>
<th>Associated tumours</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB</td>
<td>13q14</td>
<td>familial retinoblastoma</td>
<td>retinoblastomas</td>
<td>inhibitor of transcription, cell cycle regulation</td>
</tr>
<tr>
<td>TP53</td>
<td>17p13</td>
<td>Li-Fraumeni Syndrome</td>
<td>soft tissue sarcomas, breast cancer, brain tumours, leukaemia</td>
<td>transcription factor, apoptosis, response to DNA damage</td>
</tr>
<tr>
<td>CDKN2 (p16)</td>
<td>9p21</td>
<td>familial melanoma</td>
<td>malignant melanomas</td>
<td>cyclin-dependent kinase inhibitor</td>
</tr>
<tr>
<td>WT1</td>
<td>11p13</td>
<td>Wilms' tumour</td>
<td>nephroblastomas</td>
<td>transcription factor</td>
</tr>
<tr>
<td>APC</td>
<td>5q21</td>
<td>familial adenomatous polyposis</td>
<td>colorectal tumours</td>
<td>signal transduction, cell adhesion</td>
</tr>
<tr>
<td>NF1</td>
<td>17q11.2</td>
<td>neurofibromatosis type I</td>
<td>neurofibromas</td>
<td>GTPase-activating protein, signalling pathways</td>
</tr>
<tr>
<td>NF2</td>
<td>22q12</td>
<td>neurofibromatosis type II</td>
<td>schwannomas, meningiomas</td>
<td>putative membrane-organizing protein, involved with cytoskeleton</td>
</tr>
<tr>
<td>VHL</td>
<td>3p25</td>
<td>von Hippel-Lindau disease</td>
<td>renal cell carcinomas, retinal and cerebellar haemangioblastomas</td>
<td>transcription elongation</td>
</tr>
</tbody>
</table>

Table 1.1 Functions of tumour suppressor genes.
Further genetic alterations will cause progression to a metastatic tumour (Figure 1.1).

1.2 Inherited susceptibility to cancer

It is now widely accepted that cancer predisposition can be an inherited trait. Indeed, familial clustering has been reported for virtually every form of human cancer. Features of an inherited cancer syndrome include clustering of cancers in a family (two or more first degree relatives with tumours), a young age at diagnosis and multiple primary tumours. Inherited predisposition is due to a germline mutation, usually in a tumour suppressor gene. However, both alleles of the tumour suppressor gene must be inactivated for tumour development.

Knudson proposed a 'two hit model' of tumorigenesis based on his studies of retinoblastoma (Knudson, 1971). The model states that 'two hits' are required for tumorigenesis. In familial cancers the first hit is a mutation in the germline; in sporadic cancers the first hit is a somatic mutation. The second hit would be somatic inactivation of the second copy of the same gene in both familial and sporadic cases (Figure 1.2). Hence, the susceptibility to cancer would be dominantly inherited, but oncogenesis itself would be recessive. The inactivation of the second allele is often observed as loss of heterozygosity and there are various mechanisms that result in homozygosity for the mutant allele. These may include: mitotic non-disjunction with loss of the wild-type chromosome or reduplication of the mutant chromosome, mitotic recombination, or more regionalised events such as gene conversion, deletion or mutation (Cavenee et al., 1983), (Figure 1.3). Hence, loss of heterozygosity (LOH) is often indicative of a tumour suppressor gene in the region of loss.

Several tumour suppressor genes are now known and many of these have been identified through their role in hereditary cancers. Patients with von Hippel-Lindau disease develop renal cell carcinomas due to mutations in the VHL gene, which is a negative regulator of a transcription elongation factor (reviewed in Krumm and Groudine, 1995). Malignant melanoma may be inherited due to germline mutations in the CDKN2 gene. The CDKN2 gene product, p16, is an inhibitor of the cyclin dependent kinase, CDK4, and hence involved in regulation of the cell cycle (Hussussian et al., 1994, Kamb et al., 1994b). Wilms' tumour is a childhood cancer caused by mutations in the transcription factor, WT1 (Call et al., 1990, Wang et al., 1993). Other familial cancer syndromes include
Figure 1.1 A model to illustrate the multistep nature of colorectal cancer (adapted from Fearon and Vogelstein, 1990).
FAMILIAL CANCER

1st hit

inherited germline mutation in one allele

2nd hit

somatic mutation in second allele

tumour

SPORADIC CANCER

1st hit

somatic mutation in first allele

2nd hit

somatic mutation in second allele

tumour

Figure 1.2 Knudson's two hit model of tumorigenesis. Cancer may be inherited or sporadic. In the inherited form, the first hit is an inherited mutation in a cancer-predisposing gene, and the second hit is a somatic mutation leading to tumour development. In sporadic cases both the first and second hits are somatic mutations.
Figure 1.3 Mechanisms of loss of heterozygosity (adapted from Cavenee et al, 1983).
neurofibromatosis types I and II, which are due to mutations in genes involved in interactions with the RAS proto-oncogene and the cytoskeleton, respectively (Thomas et al., 1994, Wallace et al., 1990). Hence, tumour suppressor genes are involved in a variety of processes including regulation of proliferation, differentiation, apoptosis and response to genetic damage (Table 1.1).

Whilst most inherited cancer syndromes are due to germline mutations in a tumour suppressor gene, there is one exception. Multiple endocrine neoplasia types 2A and 2B (MEN2A and MEN2B) are due to germline mutations in the RET proto-oncogene, a receptor tyrosine kinase (Eng et al., 1994, Mulligan et al., 1993). In contrast to the tumour suppressor genes implicated in the syndromes discussed above, mutations in RET involve gain of function and one mutational event is sufficient to lead to tumour formation.

1.3 Breast cancer susceptibility genes

Breast cancer is the most common cancer in women, with about one in ten women being affected. Approximately 5-10% of cases are inherited (Claus et al., 1991) and whilst this appears to be a low percentage, it constitutes a significant number of cancer cases, due to the disease being common. Indeed, of the known risk factors for breast cancer, a positive family history seems to be the most significant. Studies show that the risk of breast cancer is greater for a woman with several affected relatives than for a woman with one affected relative, and that risk increases with decreasing age of the index case (Claus et al., 1990). Therefore, it is important to study the inherited disease, not only because genes involved in familial disease are often also involved in the sporadic form, but to allow predictive testing and genetic counselling for those at risk and ultimately the elucidation of gene functions and hence, effective therapies.

Various genes have been implicated in breast cancer susceptibility. Epidemiological data show that a large proportion of inherited breast cancer is likely due to mutations in two highly penetrant genes, BRCA1 and BRCA2 (Easton et al., 1993, Wooster et al., 1994). The remaining susceptibility may be accounted for by mutations in other genes with lower penetrance, such as the genes involved in Li-Fraumeni syndrome (TP53), Ataxia telangiectasia (ATM) and Cowden disease (PTEN/MMAC1) (Lloyd and Dennis, 1963, Srivastava et al., 1990, Swift et al., 1991). In addition, there may be as yet unidentified genes which, when mutated, also confer breast cancer susceptibility.
1.3.1 The \textit{BRCA1} gene

Epidemiological studies predict that approximately 30\% of inherited breast cancer is due to the inheritance of mutations in the breast cancer susceptibility gene linked to chromosome 17q21 (Ford \textit{et al.}, 1995, Hall \textit{et al.}, 1990). This gene has been named the \textit{BRCA1} gene (Solomon and Ledbetter, 1991). Mutations in this gene are believed to carry a risk for breast cancer of 85\% and of 63\% for ovarian cancer by age 70 (Easton \textit{et al.}, 1995). Gene carriers may also be at risk of colon and prostate cancer (Ford \textit{et al.}, 1994). The \textit{BRCA1} gene is the main subject of this thesis and will be comprehensively discussed in subsequent sections.

1.3.2 The \textit{BRCA2} gene

Whilst mutations in \textit{BRCA1} are likely to be responsible for the majority of breast/ovarian cancer families, they may account for just half of breast cancer-only families, leading to the suggestion that there may be other high penetrance breast cancer susceptibility genes (Easton \textit{et al.}, 1993). Evidence in support of this came from the localisation of a second susceptibility gene, \textit{BRCA2}, to chromosome 13q12-13 (Wooster \textit{et al.}, 1994). This is a region where loss of heterozygosity has been observed in breast tumours and therefore could harbour a tumour suppressor gene (Collins \textit{et al.}, 1995, Devilee \textit{et al.}, 1989). Furthermore, a family with breast cancer exhibiting LOH at chromosome 13q showed loss of the wild-type allele in tumours (Collins \textit{et al.}, 1995). The \textit{BRCA2} gene has now been isolated and mutations found in breast cancer families (Gayther and Ponder, 1997, Tavtigian \textit{et al.}, 1996, Wooster \textit{et al.}, 1995). It is a large gene, encoding a protein of 3418 amino acids, and is expressed ubiquitously (Tavtigian \textit{et al.}, 1996). Although the precise function of BRCA2 is not yet known, similarities to BRCA1 are being found and will be discussed below.

The risk of breast cancer in female mutation carriers is probably similar to that of \textit{BRCA1}, although the risk of ovarian cancer is less (Easton, 1997, Easton \textit{et al.}, 1997, Wooster \textit{et al.}, 1995). In addition, mutations in \textit{BRCA2} confer susceptibility to male breast cancer and may also increase the risk of prostate, pancreatic and laryngeal cancers (Easton \textit{et al.}, 1997, Wooster \textit{et al.}, 1995).
1.3.3 The TP53 gene

That genes involved in sporadic breast cancer may also be involved in familial breast cancer is well illustrated by the TP53 gene. This is one of the most frequently mutated genes in cancer (Hollstein et al., 1991) and therefore it has been extensively studied, with the result that its functions are beginning to be understood. p53 is involved in the response to DNA damage (Kastan et al., 1991). After exposure to DNA-damaging agents, levels of p53 protein increase, inducing p21\textsuperscript{CIP1/WAF1}, which in turn blocks cell cycle progression (El-Deiry et al., 1993). Therefore, upon DNA damage, p53 is involved in the decision whether to exit the cell cycle and repair the damage, or to follow the path of apoptosis (cell death), by mechanisms that are not yet fully understood (Shaw et al., 1992, reviewed in Levine, 1997). Consequently, loss of this cell cycle checkpoint allows a cell to replicate and divide with damaged DNA and may also elevate the overall mutation rate, resulting in tumorigenesis. Germline mutations in this gene have been identified in families affected with Li-Fraumeni Syndrome (Malkin et al., 1990, Srivastava et al., 1990). This syndrome is characterised by children with soft tissue sarcomas, who often have a parent with cancer, particularly breast cancer in the mothers (Li and Fraumeni, 1969).

1.3.4 The ATM gene

Ataxia telangiectasia (A-T) is an autosomal recessive disorder with symptoms such as an unsteady gait (ataxia) reflecting cerebellar degeneration, dilated blood vessels (telangiectases) in the eyes and acute sensitivity to radiation. A proportion of inherited breast cancer may be due to germline mutations in the ataxia telangiectasia gene, ATM, since heterozygotes (about 1% of the population) have been reported to have approximately three times the risk seen in the normal population (Easton, 1994). In addition, loss of heterozygosity on chromosome 11q22-q23, where the ATM gene is located, has been observed in breast tumours (Carter et al., 1994, Gudmundsson et al., 1995, Hampton et al., 1994, Kerangueven et al., 1997, Koreth et al., 1997, Laake et al., 1997, Winqvist et al., 1995). The ATM gene was cloned in 1995 and has homology to phosphatidylinositol-3' kinases (PI-3 kinases) that are involved in signal transduction, meiotic recombination and cell cycle control (Savitsky et al., 1995). Like p53, the ATM gene product is also involved in a DNA damage checkpoint and indeed is believed to function upstream of p53, since cells from patients with A-T do not show an increase in p53 levels in response to ionising radiation and
do not arrest in G1 or G2 stages of the cell cycle, in contrast to control cells (Kastan et al., 1992).

Various germline mutations in the ATM gene have now been reported in a number of A-T patients (Byrd et al., 1996b) and so mutation analysis is also being performed on breast cancer patients. Surprisingly, ATM mutations seem to be less common in breast cancer patients than might have been expected (Fitzgerald et al., 1997, Vorechovsky et al., 1996b), although a Swedish population showed a higher frequency of ATM carriers among breast cancer patients than would be expected in the rest of the population (Vorechovsky et al., 1996a). Clearly further studies need to be carried out to determine the proportion of breast cancer patients carrying mutations in the ATM gene.

1.3.5 The PTEN/MMAC1 gene

Cowden disease is a rare autosomal dominant cancer syndrome with a high risk of breast and thyroid tumours (Lloyd and Dennis, 1963). Two independent groups have recently identified a strong candidate gene, PTEN/MMAC1, on chromosome 10q23.3 (Li et al., 1997, Steck et al., 1997). Mutations in the gene were found in a number of cell lines and primary tumours, including breast tumours (Li et al., 1997, Steck et al., 1997). Subsequently, germline mutations in PTEN/MMAC1 have been found in patients with Cowden disease, confirming that this is the susceptibility gene (Liaw et al., 1997, Nelen et al., 1997). The proportion of breast cancer caused by mutations in the PTEN/MMAC1 gene is, however, not yet known. A recent study suggested that neither LOH nor linkage to this region of chromosome 10 was common in breast cancer, suggesting that the Cowden disease gene may not be a major breast cancer susceptibility gene (Kerangueven et al., 1997).

1.3.6 Male breast cancer susceptibility genes

Male breast cancer is rare, but some cases are inherited due to germline mutations in the BRCA2 gene (see above and Thorlacius et al., 1995). Mutations in the BRCA1 gene are unlikely to be involved, since male breast cancer families show strong evidence against linkage to BRCA1 (Stratton et al., 1994). Mutations have been found in the androgen receptor gene of some patients, suggesting that mutations in this gene may also confer cancer susceptibility in a few cases (Lobaccaro et al., 1993, Wooster et al., 1992).
1.4 Positional cloning of the *BRCA1* gene

In the absence of any functional information, the *BRCA1* gene has had to be cloned by relying solely on the basis of its chromosomal position, a method known as 'positional cloning'. This technique involves a number of steps from the chromosomal localisation of the disease susceptibility gene to its isolation (Figure 1.4).

1.4.1 Linkage analysis

The first step is the genetic linkage analysis of the disease phenotype in members of large affected families. Polymorphic markers from different chromosomal regions are used to identify segregation of the disease with a common marker(s) in the affected family members. The markers used may be restriction fragment length polymorphisms (RFLPs), variable number of tandem repeat sequences (VNTRs) or microsatellites. Once a chromosomal location has been identified, this region is refined by looking for recombination events in affected individuals to narrow the region containing the disease gene.

The first breakthrough in identifying the *BRCA1* gene came in 1990 when linkage to chromosome 17q21 was found and confirmed (Hall *et al.*, 1990, Narod *et al.*, 1991). Polymorphic markers from chromosome 17q were used to narrow the region containing the *BRCA1* gene. This was complicated by a number of factors: 1) breast cancer is a common disease, and therefore phenocopies may be present within families (individuals who have the disease but not due to mutations in *BRCA1*); 2) breast cancer is a genetically heterogeneous disease, since other genes may be involved in addition to *BRCA1* (*TP53, ATM*); 3) the *BRCA1* gene may not be fully penetrant, and therefore individuals who carry the mutant allele may not develop the disease.

However, despite these problems, worldwide efforts succeeded in narrowing the region first between the markers D17S250 and D17S579 (Easton *et al.*, 1993) and subsequently between THRA1 and D17S78 (Bowcock *et al.*, 1993, Simard *et al.*, 1993) (Figure 1.5). Further recombinants placed the gene distal to D17S857, and then to D17S776, reducing the region to about 1-1.5 Mb (Goldgar *et al.*, 1994, Kelsell *et al.*, 1993). No more recombinants were reported and the construction of physical and transcript maps began.
Figure 1.4 The positional cloning strategy for the isolation of a disease susceptibility gene. A disease locus is localised to a chromosomal region, either by a cytogenetic abnormality indicating the chromosome involved, or by genetic linkage analysis of multiple affected families. Finer genetic mapping is performed to reduce the locus to as small a region as possible, and then a physical map is created using cosmids, YACs, BACs or PACs. If any cloned genes are known to map to this region, they can be assessed as candidates, otherwise novel genes must be identified. Any candidates must be screened for mutations in affected individuals in order to identify the disease susceptibility gene.
Figure 1.5. Genetic map of the BRCA1 region. Linkage to D17S74 was first reported by Hall et al in 1990. 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 1-1.5 megabase pairs. The polymorphic marker D17S855 is now known to reside within the BRCA1 gene. Abbreviations for gene markers are: EDH, oestradiol dehydrogenase; NBR1, a novel B-box protein; RNU2, a component of the small nuclear ribonucleoprotein complex; PPY, pancreatic polypeptide.
1.4.2 Analysis of known genes within the region

At this stage, genes known to map to the region may be considered as potential candidates for the disease gene. In the absence of known genes, or if they prove not to be the gene of interest, a physical map must be constructed and new genes isolated. A number of candidate genes were known to map to the initial BRCA1 region, but these were subsequently found by linkage analysis to lie outside the region. Hence, it was necessary to construct a physical and transcript map.

1.4.3 Creation of a physical map

The construction of a physical map of the region is the next step, to generate genomic clones that can be used as tools in the isolation of candidate genes. Yeast artificial chromosomes (YACs) have been widely used in physical mapping (Burke et al., 1987). However, YACs have a number of disadvantages in that they are often chimaeric and may be unstable, deleting portions of the insert. Consequently, YACs are frequently supplemented with other vectors such as cosmids, bacteriophage P1 clones (Sternberg, 1990), bacterial artificial chromosomes (BACs) (Shizuya et al., 1992) and P1-derived artificial clones (PACs) (Ioannou et al., 1994). These vectors have overcome many of the disadvantages of YACs, but the size of insert in these clones is smaller. Whilst YACs can accommodate inserts between 100-2000 kb, cosmids contain only 35-40 kb. P1 clones have inserts in the range of 70-100 kb, BACs up to 300 kb and PACs 100-300 kb. Cosmids and PACs are probably most useful in complementing YACs, since they carry antibiotic resistance, and hence can be positively selected (unlike BACs), and they can be transformed into host E. coli cells by electroporation (unlike P1 clones that require packaging of phage particles) (reviewed in Monaco and Larin, 1994).

A number of groups have generated physical maps of the BRCA1 region (Albertsen et al., 1994, Brown et al., 1995, Couch et al., 1995, Jones et al., 1994, Neuhausen et al., 1994).

1.4.4 Construction of a transcript map

Crucial to the positional cloning strategy is the efficient and comprehensive isolation of transcripts in the disease region. Since often little is known about the target gene, all genes in the region must be identified and analysed. This is the
rate-limiting step in the positional cloning strategy, and hence many techniques are being developed, improved and used to speed up this process. In addition to the more traditional methods of gene isolation (including screening cDNA libraries and Northern blots with complex probes), several other techniques have been developed (reviewed in Brennan and Hochgeschwender, 1995, Parrish and Nelson, 1993). Some of the more popular techniques are described below.

**Cross-species homology**

Coding sequences are more often evolutionarily conserved between species than are non-coding regions. Hence, looking for cross-species homology can indicate the presence of a gene. This is carried out by hybridising fragments of genomic clones to Southern blots containing DNA from a range of species, known as 'zoo blots'. This technique has been extremely useful in the isolation of several genes, including the Wilms' tumour gene (Call et al., 1990), the Duchenne muscular dystrophy gene (Monaco et al., 1986), the neurofibromatosis type I gene (Wallace et al., 1990) and the X-linked adrenoleukodystrophy gene (Mosser et al., 1993). However, the disadvantages of this technique are that it is labour-intensive and that it will not detect genes that have diverged during evolution.

**Detection of CpG islands**

CpG islands are clusters of non-methylated CpG dinucleotides in a relatively G-C rich region, of approximately 1-2 kb and are frequently found at the 5' end of genes (Bird, 1986). These islands can be detected by cutting with methylation-sensitive restriction enzymes, and hence can be a useful indicator of the presence of a gene.

**Exon amplification**

A newer technique that has been developed to detect exons within genomic DNA is exon amplification, also called exon trapping (Buckler et al., 1991a). This method relies on the splice acceptor and donor sequences that flank exons. Fragments of genomic DNA are cloned into a plasmid vector within an intron flanked by 5' and 3' splice sites of the HIV-1 tat gene. COS7 cells are transfected with these constructs and the resulting RNA transcripts are processed in vivo. Splice sites of any exons present in the genomic DNA can pair with the HIV splice sites so that mature RNA will contain the previously unknown exons. These can be amplified by reverse transcriptase polymerase chain reaction (RT-
PCR). This method should be capable of detecting most of the exons within a given fragment. The limitations of this technique are that the PCR products may contain intronic sequence, due to cryptic splice sites present in the sequence and intronless or single intron genes will be missed.

**Direct selection**

Another recent but popular technique is direct selection (Korn *et al.*, 1992, reviewed in Lovett, 1994). The principle behind this method is that cDNAs specific to a genomic region can be selected from a mixture of cDNAs. The method is outlined in Figure 1.6 and involves the hybridisation of PCR amplified cDNAs to cosmids. The cosmids are previously blocked to remove repetitive sequences. The cDNA/cosmid hybrids are eluted and reamplified by PCR. This yields a set of cDNAs enriched for the region of interest. The cDNAs will be specific for the tissue or tissues from which the original cDNAs were derived. The main limitation of this method is that the starting cDNA must contain the targeted cDNA. This is the major limitation of all expression-dependent techniques. In addition, there are PCR limitations of this technique, such as a preference for smaller products and hybridisation limitations such as failure of short exons to hybridise, or hybridisation of cDNAs to related genes or pseudogenes. On the other hand, it has several advantages, in that it works well with large target sequences, is very sensitive and can sample many tissues in one experiment.

**Direct sequencing**

Another technique that has recently become feasible is the large-scale sequencing of genomic clones with the use of computer programs to identify potential exons.

Since all the techniques have different advantages and disadvantages (Table 1.2) and no one technique will identify all transcripts, a combination of methods is usually employed. As transcripts are isolated, each one must be screened for mutations in affected family members. Only when disease-causing mutations are identified and shown to segregate with the disease has the target gene been found.

As expected, a number of techniques were employed in isolating candidates for the *BRCA1* gene, with some groups choosing to focus on one technique such as direct selection (Hattier *et al.*, 1995, Jacob *et al.*, 1995, Osborne-Lawrence *et al.*,...
Figure 1.6 Method for generating a cDNA library by direct selection (adapted from Korn et al, 1992).
<table>
<thead>
<tr>
<th>Technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>cross-species homology</td>
<td>-not expression-dependent</td>
<td>-labour intensive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-will not detect divergent genes</td>
</tr>
<tr>
<td>detection of CpG islands</td>
<td>-not expression-dependent</td>
<td>-not all genes are associated with CpG islands</td>
</tr>
<tr>
<td></td>
<td>-can detect divergent transcripts</td>
<td></td>
</tr>
<tr>
<td>exon amplification</td>
<td>-not expression-dependent</td>
<td>-may include intronic sequence due to cryptic splice sites</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-will not detect intronless or single intron genes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-verification may be difficult</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(most genes are not expressed in most tissues)</td>
</tr>
<tr>
<td>direct selection</td>
<td>-large target size possible</td>
<td>-starting cDNA must contain target cDNA</td>
</tr>
<tr>
<td></td>
<td>-very sensitive</td>
<td>-PCR favours short fragments</td>
</tr>
<tr>
<td></td>
<td>-verification easier</td>
<td>-PCR may incorporate sequence errors</td>
</tr>
<tr>
<td></td>
<td>-rapid</td>
<td>-hybridisation to short exons may fail</td>
</tr>
<tr>
<td></td>
<td>-insensitive to introns and cryptic splice sites</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-can detect rare and divergent transcripts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-can sample many tissues in a single experiment</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-may obtain hybridisation to related genes or pseudogenes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-low level repeats, mitochondrial DNA and ribosomal sequences may complicate analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-may not be comprehensive</td>
</tr>
</tbody>
</table>

Table 1.2 Advantages and disadvantages of techniques for gene isolation.
1995), direct screening of cDNA libraries (Albertsen et al., 1994, Friedman et al., 1995a) or exon amplification (Brown et al., 1995). Others used a variety of methods, including exon amplification, direct selection, evolutionary conservation and identification of CpG islands (Brody et al., 1995, Jones et al., 1994). What has become evident is that no technique identified all transcripts and several techniques identified unique transcripts missed altogether by other methods, reinforcing the importance of not relying on a single method for gene isolation. Direct selection appears to have been particularly efficient compared to direct screening of cDNA libraries and genomic sequencing (Harshman et al., 1995).

Whilst some genes were identified by several different groups, there was a perhaps surprising lack of overlap. This may be due to the region having not been exhaustively searched, to technical difficulties, or to some genes not being expressed in the tissues examined. In one particular study, comparison of the candidate cDNA fragments identified, excluding known or homologous genes, with other reports showed that only five overlapped (Harshman et al., 1995). The authors suggest a number of reasons for this observation: i) the sequences deposited in the Genbank database may not be the total sequence length identified; ii) the candidate cDNA isolation was not complete; or iii) a large proportion of the isolated cDNAs do not represent expressed sequences from the region. An incomplete transcript map is the most likely explanation.

However, more than 25 genes have been isolated and mapped to this region, including the NBR1 gene, a candidate for the ovarian tumour antigen, CA125, a dual-specificity serine/tyrosine phosphatase, a human ADP ribosylation factor, homologues of the Drosophila tumour suppressor, dlg-A and a homologue of the yeast PRP22 mRNA splicing factor (Brody et al., 1995, Campbell et al., 1994b, Friedman et al., 1995a, Ishibashi et al., 1992, Jones et al., 1994, Kamb et al., 1994a, Mazoyer et al., 1995, Osborne-Lawrence et al., 1995, Smith et al., 1996a, Smith et al., 1995). Other cDNAs were identified which had no homology to known proteins and these may be novel genes.

1.5 The BRCA1 gene

Neuhausen et al identified a recombinant in a large breast/ovarian cancer family, placing the gene between D17S1321 and D17S1325 and reducing the 1-1.5 Mb region to approximately 600 kb (Neuhausen et al., 1994), (Figure 1.5). Isolation of
transcripts in this region identified a gene with putative disease-causing mutations in four breast/ovarian cancer families (Futreal et al., 1994b, Miki et al., 1994). These mutations included deletions, insertions, nonsense and missense mutations. Whilst the missense mutation could not categorically be claimed to be disease-causing, the other mutations were strongly in favour of this gene being a very plausible candidate for BRCA1. In addition, affected members from another family were shown to express only one allele at the mRNA level, while both alleles were present at the genomic level, and they were therefore, presumed to have a regulatory mutation. All the mutations fulfilled the necessary criteria for a disease gene: co-segregation of the mutant allele with the disease and absence of the mutant alleles in controls. Subsequently, other groups confirmed that this candidate is indeed the BRCA1 gene, reporting BRCA1 mutations in a large number of breast and breast/ovarian cancer families worldwide (see below).

1.5.1 Features of the BRCA1 gene

BRCA1 is a large gene, covering about 81 kb, which has now been completely sequenced (Miki et al., 1994, Smith et al., 1996b). The 7.8 kb mRNA contains 22 exons encoding a predicted protein of 1863 amino acids. Northern mRNA analysis shows that this transcript is widely expressed and most abundant in testis and thymus, with intermediate expression in breast and ovary. A RING finger domain was identified near the amino terminal of the predicted protein. This is a type of zinc finger found in a number of different proteins and may mediate protein-protein interactions (Freemont, 1993, Saurin et al., 1996). The sequence gave no other clues to the function of this gene, although it appears to be conserved in a number of species including sheep, pig, rabbit and mouse (Miki et al., 1994).

1.5.2 Mutation analysis of the BRCA1 gene

Identification of mutation carriers can, in principle, allow early detection, treatment or prevention of the disease. It will also provide information on the proportion of breast, ovarian and any other cancers that are due to germline BRCA1 mutations. For these reasons many techniques have been developed for detecting mutations.
Methods for detecting mutations

The challenge is to find a technique that is fast, efficient, and inexpensive. Several factors affect the speed and efficiency of a technique, including the size and structure of the gene, and the nature and distribution of mutations throughout the gene.

There are two modes of mutation detection: the scanning mode and the diagnostic mode. The scanning mode searches for unknown mutations revealed by aberrant migration (denaturing gradient gel electrophoresis, heteroduplex analysis, single strand conformation polymorphism analysis) or by DNA/DNA cleavage (chemical cleavage by mismatch). The diagnostic method screens for a specific mutation (allele specific oligonucleotide hybridisation, restriction enzyme digestion). Common methods, along with their advantages and disadvantages, are discussed below (reviewed in Gayther and Ponder, 1997).

- **DGGE** (Denaturing gradient gel electrophoresis). This method is based on the principle that the melting of DNA duplexes is dependent on the strength of base pairing and therefore, the sequence (Myers et al., 1985). Hence, the mobility of a mutant duplex will differ from that of a wild-type duplex and from that of a wild-type/mutant duplex. This technique is highly sensitive (near 100%) but requires specialised equipment.

- **HA** (Heteroduplex analysis). In a non-denaturing gel, a heteroduplex with a single mismatch will have a different mobility to the corresponding duplex without the mismatch and this allows their separation (White et al., 1992). Whilst this technique is rapid and easy to perform, it has relatively poor sensitivity for detecting point mutations.

- **SSCP** (single strand conformation polymorphism analysis). Single-stranded DNA samples differing in one or more bases will adopt different conformations in a non-denaturing environment, and hence the species can be separated by non-denaturing gel electrophoresis (Orita et al., 1989). This method is simple and can detect about 80% of mutations; it can also be enhanced by separating the same fragments under a combination of different gel conditions (Michaud et al., 1992).

- **CCM** (chemical cleavage by mismatch). DNA fragments containing mismatches can be cleaved chemically and separated by gel electrophoresis
Chapter One: Introduction

(Cotton et al., 1988). This method has several advantages: it is able to scan longer fragments than the above techniques; it has high sensitivity; and it gives an indication of the approximate location of the mutation. However, it requires the use of toxic chemicals, and is labour-intensive.

- **PTT** (protein truncation test). This technique is based on reverse transcriptase polymerase chain reaction (RT-PCR), followed by in vitro transcription and translation of the PCR products (Roest et al., 1993). PCR products carrying truncating mutations will be indicated by shorter polypeptides. This method gives an approximate location for the mutation, but has the disadvantage that it will only detect truncating mutations.

- **ASO** (allele specific oligonucleotide hybridisation). This technique is used to screen for a known mutation. It relies on the principle that two perfectly complementary oligonucleotides will bind more strongly than two with a mismatch. After hybridisation, washing will distinguish between perfect and imperfect matches, and hence wild-type and mutant sequences. This is a rapid and efficient way of screening for known mutations, although it may require optimisation and will not detect new mutations (Conner et al., 1983).

- **Restriction enzyme digestion.** If a mutation creates or destroys a known restriction enzyme recognition site, then the absence or presence of the mutation can be detected by digestion with this enzyme. This is a rapid method, since it only requires a PCR reaction and a enzyme digestion, but it can only detect a known mutation and only then if it creates or abolishes a recognition site.

- **Direct sequencing.** Most of the scanning techniques do not precisely define the location and nature of the mutation, and therefore sequencing is often the final step in mutation detection. Direct sequencing of the coding region is perhaps the most sensitive method of identifying mutations and so it is sometimes used in preference to the scanning methods. However, it is very time-consuming and expensive, and is therefore not suitable for analysis of large numbers of samples.
Mutations in the BRCA1 gene

Following the initial isolation and mutation analysis of the BRCA1 gene (Miki et al., 1994), several other reports quickly confirmed BRCA1 as the breast cancer susceptibility gene, by the identification of further germline mutations (Castilla et al., 1994, Friedman et al., 1994, Futreal et al., 1994b). At present more than 250 mutations have been reported and the following observations have been made (Couch et al., 1996, Gayther and Ponder, 1997, Shattuck-Eidens et al., 1995, Xu and Solomon, 1996):

Spectrum of BRCA1 germline mutations
Over 200 different mutations have been identified in breast and ovarian cancer families; more than half are unique and they are scattered throughout the length of the gene (Figure 1.7). The majority (~87%) of mutations are insertions, deletions, or splicing mutations causing frameshifts, or nonsense mutations, all of which are predicted to lead to a truncated protein. As a result, PTT has become a popular technique for screening for these truncating mutations (Hogervorst et al., 1995, Lancaster et al., 1996, Plummer et al., 1995).

A number of missense alterations have also been detected, but in the absence of a functional test, the significance of these is unclear. However, two missense mutations change critical cysteines in the RING finger, and hence are predicted to be disease-causing (Friedman et al., 1994). Other missense alterations may be considered likely to be disease-causing if they segregate with disease in the family and are not detected in the general population, but it is still possible that they may be rare variants, having no effect on the function of BRCA1. This problem will only be resolved when functional assays are available.

Regulatory mutations in BRCA1 have been found in several breast and/or ovarian cancer families (Miki et al., 1994, Serova et al., 1996, Xu et al., 1997b). These mutations are characterised by an unstable or loss of transcript, such that only one allele is present at the mRNA level. For most of these mutations the precise cause has not been identified, but is presumably due to an alteration in one of the regulatory regions of the gene, in the introns or the promoter. Recently, the causative mutation has been identified in one family with a regulatory mutation (Swensen et al., 1997). Affected individuals have a 14 kb deletion at the 5’ end of BRCA1, removing the promoter region and hence preventing transcription from this allele. A 1 kb deletion, including exon 17, that
Figure 1.7 Schematic diagram summarising the reported germline mutations identified in the BRCA1 gene (adapted from The Breast Cancer Information core: http://www.nhgri.nih.gov/Intramural_research/Lab_transfer/Bic/). The BRCA1 cDNA is illustrated as a rectangle with vertical lines separating the exons, some of which are numbered. Exons 1a, 1b and 4 have been omitted, since these are non-coding. Each symbol represents a different mutation. Frameshift and nonsense mutations are depicted by □, missense mutations by ○, and splice mutations by ( ). Unique mutations are illustrated in yellow, recurrent mutations in green. The number of times a recurrent mutation has been reported is shown below the symbol. The two common mutations are shown in black.
leads to an unstable transcript, has also been reported (Puget et al., 1997). Both these mutations are thought to have arisen from Alu-mediated recombination. Since the BRCA1 gene region has a high density (42%) of Alu elements (Smith et al., 1996b), it is possible that other Alu-mediated rearrangements may exist. Indeed, one group has now found BRCA1 genomic deletions ranging from 510 bp to ~14 kb in 14 families, comprising 36% of BRCA1 mutations in Dutch families (Petrij-Bosch et al., 1997). Hence, genomic deletions may be relatively common and patients with regulatory mutations or those with evidence for linkage, but with no detectable mutation, should be investigated for these rearrangements.

The frequency of BRCA1 mutations in breast/ovarian cancer families has now been examined in a number of populations and has shown that the proportion of high-risk families attributable to BRCA1 mutations varies widely among different populations. For example, BRCA1 mutations are found in 75% of ovarian cancer families in Russia, whereas less than 20% of high-risk families in Japan carry BRCA1 mutations (Gayther et al., 1997, Inoue et al., 1995, Matsushima et al., 1995; reviewed in Szabo and King, 1997). In most populations mutations in BRCA1 and BRCA2 together account for about 6-10% of all breast and ovarian cancer, while in high-risk families they account for about 70%. This still leaves about 30% of familial cancer unaccounted for and has led to the suggestion that further BRCA susceptibility genes may yet be found (Hakansson et al., 1997, Rebbeck et al., 1996, Serova et al., 1997). Preliminary support for this comes from studies which have shown LOH of chromosome 8p in breast tumours and linkage to this region has been found in breast cancer families that do not show linkage to BRCA1 or BRCA2 (Kerangueven et al., 1995, Seitz et al., 1997).

Common BRCA1 mutations
Whilst more than 130 unique mutations have been reported, two mutations together account for about 20% of all mutations: these are the 185delAG mutation in exon 2 and the 5382insC mutation in exon 20 (Figure 1.7). Both mutations cause a frameshift that results in premature termination. These common mutations could be due to multiple occurrences of the specific mutation, or they could be due to a founder effect. Individuals reported to carry these recurrent mutations almost all share the same haplotype and many are of Ashkenazi Jewish origin (Friedman et al., 1995b, Neuhausen et al., 1996, Simard et al., 1994), suggesting that a founder effect is responsible. Founder effects have also been reported in other populations (Johannsson et al., 1996).
Studies of these mutations in the general Ashkenazi Jewish population have shown that the carrier frequency for 185delAG may be approximately 1% (Struwing et al., 1995). This is in contrast to the frequency of BRCA1 mutations estimated for Caucasians, which is 1 in 833 (Ford et al., 1995). The frequency in the Ashkenazi Jewish population would predict that this mutation would account for 16% of breast cancer and 39% of ovarian cancer; indeed approximately 20% of Ashkenazi women with early-onset breast cancer have been found to carry the 185delAG mutation and this was not restricted to those with a family history (Fitzgerald et al., 1996, Offit et al., 1996, Roa et al., 1996, Struwing et al., 1995). A common mutation in BRCA2, 6174delT, has also now been reported in the Ashkenazi population, occurring at a frequency of about 1.5% (Roa et al., 1996). However, although the frequencies of the 185delAG and 6174delT mutations are similar, the 6174delT mutation is found significantly less often in high-risk families, indicating that this mutation may have lower penetrance than 185delAG (Levy-Lahad et al., 1997, Roa et al., 1996).

A recent study suggested that together the three mutations, BRCA1 185delAG, BRCA1 5382insC and BRCA2 6174delT, account for 62% of ovarian cancer and 30% of early-onset breast cancer among Ashkenazi Jewish women (Abeliovich et al., 1997). This raises the issue of whether or not it is feasible and/or appropriate to screen this population for these mutations. However, despite their high frequencies, caution must be exercised: a negative result may not be meaningful, as other mutations occur in this population, so the individual could still be carrying BRCA1 or BRCA2 mutations; a positive result should also be interpreted with caution, since the precise penetrance and therefore, the risks attached to the mutations are not known, and hence carriers of the mutation may not develop breast or ovarian cancer.

Genotype/phenotype correlations

In an attempt to explain the variation in the proportion of breast and ovarian cancer cases observed between BRCA1 families, many groups have looked for correlations between mutations and disease phenotypes. Gayther et al have reported that mutations in the 3' third of the gene are associated with a lower proportion of ovarian cancer (Gayther et al., 1995). This could be due to the residual protein domains having some function that protects against ovarian cancer but not breast cancer. This observation was also noted by other investigators, although it did not reach significance (Shattuck-Eidens et al., 1995).
However, a number of other groups have failed to find this association (Berman et al., 1996, Serova et al., 1996, Stoppa-Lyonnet et al., 1997).

Sobol et al suggested another genotype/phenotype correlation: that mutations in the N-terminal RING finger or the C-terminal region may lead to more highly proliferating tumours than mutations in the central portion of the gene (Sobol et al., 1996). Further studies are required to confirm these findings.

1.5.3 Clinical features of BRCA1 tumours

Studies of the histological characteristics of breast and ovarian tumours with BRCA1 mutations have been carried out, to determine the most appropriate therapeutic strategies. In general, breast cancers with BRCA1 mutations are of a higher grade than tumours with no BRCA1 mutations (Eisinger et al., 1996, Lakhani et al., 1997). This was attributed to more pleomorphism, higher mitotic index and less tubule formation, all indicative of highly proliferating tumours (Lakhani et al., 1997). BRCA1 tumours differed from sporadic tumours, but also differed from BRCA2 tumours. Interestingly, a study of ovarian cancers with germline BRCA1 mutations showed that these tumours appeared to have a more favourable clinical course than sporadic tumours (Rubin et al., 1996).

1.5.4 The role of BRCA1 in sporadic breast and ovarian cancer

Support for this gene playing a role in both familial and sporadic breast cancer came from LOH studies which showed that this region is frequently lost in breast tumours (Caligo et al., 1996, Futreal et al., 1992, Godwin et al., 1994, Kirchweger et al., 1994, Lindblom et al., 1993, Saito et al., 1993). Hence, somatic mutations were expected in sporadic breast and ovarian tumours. Surprisingly, no somatic mutations have been found in sporadic breast tumours and only five in ovarian tumours (Futreal et al., 1994b, Hosking et al., 1995, Merajver et al., 1995). This observation suggests either that BRCA1 does not play a role in sporadic tumorigenesis, or that it does play a role but is inactivated by a different mechanism.

Evidence against a role for BRCA1 in sporadic breast tumorigenesis

The absence of somatic mutations clearly questions the role of BRCA1 in the development of sporadic tumours. At least three regions of LOH on chromosome 17q have been observed in breast and ovarian cancers (Cornelis et al., 1993,
Kirchweger et al., 1994, Nagai et al., 1995, Niederacher et al., 1997, Saito et al., 1993), suggesting that not BRCA1 but another gene nearby is the real target. Indeed, deletion units distal and centromeric to BRCA1 have been reported for both breast and ovarian tumours (Eccles et al., 1992, Godwin et al., 1994, Jacobs et al., 1993, Kalikin et al., 1997, Tangir et al., 1996). In addition, transfer of chromosome 17q distal to BRCA1 suppresses tumorigenicity in breast cancer cells, indicating that other tumour suppressor genes may be present in this region (Theile et al., 1995). Analysis of the morphological changes in BRCA1 tumours has shown differences compared to sporadic tumours, which may indicate different processes in tumorigenesis in these tumours (see section 1.5.3, Lakhani et al., 1997).

Evidence in favour of BRCA1 playing a role in sporadic breast tumorigenesis

Whilst LOH studies may indicate the involvement of other genes on chromosome 17 in sporadic breast and ovarian cancer, LOH has been observed at the BRCA1 region of mutation carriers. This invariably involved the wild-type allele (Cornelis et al., 1995, Smith et al., 1992), implying that BRCA1, at least in these cases, is the target. One hypothesis is that the BRCA1 protein may be aberrantly localised in the cytoplasm instead of the nucleus in breast cancer cells (Chen et al., 1995), although the subcellular localisation of BRCA1 has been somewhat controversial (see below). This phenomenon has been observed for the TP53 gene in breast tumours. Thirty-seven per cent of tumours showed cytoplasmic staining (nuclear localisation is essential for normal function) in the absence of coding region mutations (Moll et al., 1992). The authors suggest that nuclear exclusion may be mediated by phosphorylation at a site near the nuclear localisation signal.

Further support for a role for BRCA1 in sporadic tumorigenesis comes from the observation that BRCA1 mRNA and protein levels are reduced or absent in breast cancer cell lines and sporadic breast tumours (Thompson et al., 1995; C. Wilson unpublished results). The mechanisms for this are unknown, although several possibilities exist.

Mutations may be present but perhaps may exist in the promoter region rather than the coding region. A number of recent reports have shown that Alu-mediated rearrangements are found in familial cases, and hence sporadic cases should perhaps also be investigated for large genomic deletions (Petrij-Bosch et al., 1997, Puget et al., 1997, Swensen et al., 1997). Other sequence alterations could
Chapter One: Introduction

occur by RNA editing, a co- or post-transcriptional process in which selected nucleotides in RNA are altered from those originally encoded in the genome. This has been observed for both the WT1 and NF1 tumour suppressor genes (Sharma et al., 1994, Skuse et al., 1996). In the case of WT1, RNA editing results in an amino acid change from leucine to proline and the protein containing leucine is a less effective transcriptional repressor (Sharma et al., 1994). In NF1 editing introduces a premature stop codon, with the protein terminating N-terminal of the domain homologous to GTPase-activating proteins (Skuse et al., 1996). In both cases editing is a normal occurrence and a means of regulating gene expression, but clearly inappropriate editing could lead to disease. Interestingly, although NF1 editing was present in all cells studied, the proportion of editing was higher in tumours (Skuse et al., 1996).

Gene expression may be affected by mechanisms that do not change the sequence, such as the presence of antisense transcripts. The Wit1 gene is divergently expressed from the WT1 gene and some Wit1 transcripts contain antisense sequences of WT1 (Campbell et al., 1994a, Eccles et al., 1994, Malik et al., 1995). Hence, it has been proposed that Wit1 may regulate WT1 expression and therefore, that abnormal Wit1 expression could also affect WT1. Consequently, fine mapping of the BRCA1 region has been carried out (see section 1.5.5).

Another mechanism of gene silencing is DNA hypermethylation, which was first described for the RB gene in sporadic cases of retinoblastoma (Ohtani-Fujita et al., 1993). The same phenomenon was observed for the VHL gene in renal carcinoma cell lines and no detectable VHL expression was found in the tumours from which these cell lines were derived (Herman et al., 1994). In addition, there are several reports of methylation of the p16 gene in human cancers silencing gene expression (Gonzalez-Zulueta et al., 1995, Herman et al., 1995, Merlo et al., 1995) and expression could be regained with the use of hypomethylating drugs (Merlo et al., 1995). The latter observation may suggest potential therapeutic approaches. There is now some evidence that BRCA1 may also be inactivated by this mechanism in up to 10% of sporadic breast tumours (Dobrovic and Simpfendorfer, 1997; A. Catteau, unpublished results). Clearly these mechanisms warrant further investigation in the case of BRCA1.
1.5.5 Structure of the BRCA1 genomic region

In order to begin to investigate whether abnormal regulation of gene expression plays a role in sporadic breast and ovarian tumorigenesis, the structure of the BRCA1 gene region has been analysed in detail. Initial studies suggested that a previously isolated candidate gene, 1A1.3B (renamed NBR1- next to BRCA1), lay adjacent to the BRCA1 gene, with their 5' ends less than 300 bp apart (Brown et al., 1994). Further analysis revealed that the situation was more complex and that this region is in fact duplicated, such that the real BRCA1 gene lies next to a partial copy of NBR1 and the real NBR1 gene lies next to a partial copy of BRCA1 (Brown et al., 1996) (Figure 1.8). In each case it is only the 5' exons that are duplicated: exons 1A, 1B and 2 of BRCA1 and exons 1A, 1B and 3 of NBR1. Comparison of the real and duplicated exon sequences has shown that they are more than 90% conserved and hence, that the duplication is likely to be a recent evolutionary event.

To determine if these duplicated exons are transcribed and whether or not they may play a role in breast cancer, studies have been carried out to look for any expression. Interestingly, it now appears that the partial copies of the NBR1 5' exons form part of another gene, named NBR2 (Xu et al., 1997a). The NBR2 gene spans about 30 kb and has an mRNA of between 1.5-2.5 kb on Northern blots and is expressed in a variety of tissues. Consequently, the BRCA1 gene actually lies adjacent to the NBR2 gene, only 218 bp apart (Figure 1.8). Whether or not the partial copy of BRCA1 is expressed is not yet known. Confirmation that the duplication is a recent event comes from studies in the mouse, showing that neither the duplication nor the NBR2 gene is present in the mouse (Chambers and Solomon, 1996; see chapters five and six).

1.5.6 Regulation of BRCA1 gene expression

Since regulatory mutations have been detected in familial cases of breast and ovarian cancer and reduced levels of BRCA1 mRNA have been detected in sporadic tumours, it is important to understand the normal regulation of BRCA1 expression. Gene expression can be controlled at a number of levels, including chromatin packaging, transcription initiation and elongation, polyadenylation, splicing and stability of the mRNA and translation initiation. Transcription is controlled by the core promoter, proximal elements and distal enhancer elements.
Figure 1.8 Schematic diagram of the genomic region between the *BRCA1* and the *NBR1* genes. The first three exons of *BRCA1* and exons 1A, 1B and 3 of *NBR1* are duplicated. The first two exons of *NBR2* are homologous to exons 1A and 3 of *NBR1* respectively. The boxes represent exons and the arrows indicate the direction of transcription. This diagram is not drawn to scale.
As a first step in assessing the factors involved in \textit{BRCA1} transcription, the transcription start sites of the gene have been mapped and two initiation sites have been found, under the control of two separate promoters, \( \alpha \) and \( \beta \) (Xu \textit{et al.}, 1995). These initiation sites are present within two alternative exons, exons 1A and 1B, separated by 277 bp.

Multiple promoters and transcription start sites give additional flexibility in the control of the expression of a gene, since a single promoter may be insufficient to accommodate all the necessary information: tissue and developmental specificity, and response to hormonal or other stimuli (reviewed in Ayoubi and Van de Ven, 1996). In addition, if an ATG translation initiation codon is present in alternative 5' exons, this can result in structurally or functionally different protein isoforms. Alternative 5' exons can also affect the level of expression by altering the stability of the RNA and its translational efficiency. Hence, the use of alternative promoters enables a gene to be expressed at the appropriate time and level in the appropriate cells. For genes that are ubiquitously expressed, often one promoter is responsible for the ubiquitous expression and an alternative promoter has a more restricted spatial or temporal pattern.

The \textit{BRCA1} \( \alpha \) and \( \beta \) promoters have been studied using reporter gene assays. In this system, a portion of the presumed promoter region is fused to a reporter gene, for example the firefly luciferase gene. The reporter gene is lacking a promoter, and therefore will only be expressed if the test fragment has promoter activity. The constructs are transfected into the selected cell line, harvested after 1-3 days and assayed for expression of the reporter gene. Studies of this kind have shown that both \textit{BRCA1} promoters have activity in breast and ovarian cells, but that promoter \( \alpha \) is much stronger than \( \beta \) (Xu \textit{et al.}, 1997c). Both transcripts are expressed in most tissues, although there appear to be higher levels of the \( \alpha \) transcript in mammary tissue (Xu \textit{et al.}, 1995).

As mentioned above, the \textit{BRCA1} gene lies head to head with the \textit{NBR2} gene, and hence these two genes could be under the control of a bidirectional promoter. There are two models for bidirectional transcription: 1) the bidirectional transcription is driven by two largely independent promoters, arranged back to back, with each gene having its own set of cis-acting elements; and 2) transcription of both genes is driven by a single promoter where the cis-elements are shared. There are examples of both models. The Surf-1 and Surf-2 genes are 97 bp and 73 bp apart in the human and the mouse respectively and share
bidirectional promoters (Lennard et al., 1994, Lennard and Fried, 1991) and the dihydrofolate reductase (DHFR) gene and the hMSH3 gene (homologous to mismatch repair proteins) are divergently expressed from a bidirectional promoter in both humans and mice (Crouse et al., 1985, Shinya and Shimada, 1994). The collagen type IV genes COL4A1 and COL4A2 are divergently expressed from a shared promoter in both the human and the mouse, but since the intergenic region itself has no transcriptional activity and efficient transcription relies on downstream elements, it may be better viewed as two overlapping promoters (Kaytes et al., 1988, Follner et al., 1997). In either case, coordinated regulation may be possible. There are also examples of cancer-predisposing genes being divergently expressed with adjacent genes: for example the ATM gene lies adjacent to a gene, E14/NPAT, whose function is unknown (Byrd et al., 1996a, Imai et al., 1996).

The BRCA1 α promoter has indeed been shown to have activity in both directions and is approximately 2.5-fold stronger in the NBR2 orientation (Xu et al., 1997c). At least some cis-elements in the intergenic region are shared, since mutation of a CCAAT box affected transcription in both directions (Xu et al., 1997c). Like other bidirectional promoters, the BRCA1/NBR2 intergenic region lacks TATA boxes but has CCAAT boxes and GC boxes, which are binding sites for the Sp1 transcription factor.

This compact organisation of genes is rare in the human genome and, as illustrated above, these divergent genes can often interact together (for example, the collagen IV genes) (Heikkila et al., 1993); they can be involved in the same pathway (for example, the chicken GPAT and AIRC genes encode two enzymes involved in the purine biosynthetic pathway) (Gavalas and Zalkin, 1995); or they may be unrelated. Therefore, it will be interesting to determine the function of the NBR2 gene and any functional similarities it may have with BRCA1. Since this genomic structure is not conserved in mice, due to the absence of the duplication in the mouse, different control mechanisms may operate in the two species. However, the mouse Brca1 gene is adjacent to the mouse Nbr1 gene and these genes appear to share a bidirectional promoter (see chapter six).

Since an appropriate hormonal environment is required for proliferation of mammary epithelial cells and is also a pre-requisite for breast carcinogenesis (Nandi et al., 1995), the effect of hormones on BRCA1 expression has been investigated. These studies reveal that BRCA1 mRNA levels increase during
puberty, pregnancy and in response to ovarian hormones in the mouse (Lane et al., 1995, Marquis et al., 1995). BRCA1 mRNA and protein levels are also elevated in response to oestrogen in human breast cancer cell lines and this induction is blocked by anti-oestrogens (Gudas et al., 1995, Spillman and Bowcock, 1996). Whether this is a direct effect on the promoter sequences, or indirect, via an intermediate protein or as a result of altering the proliferative state of the cells, is not clear. Several groups suggest that it may be an indirect effect, since cycloheximide, a protein synthesis inhibitor, blocks this induction (Marks et al., 1997, Spillman and Bowcock, 1996). However, both the α and β promoters of BRCA1 have been shown to be responsive to oestrogen (Xu et al., 1997c). Neither promoter has a classical oestrogen response element (ERE), but an alternative ERE is present within promoter β (Norris et al., 1995, Xu et al., 1997c), suggesting that at least one of the promoters may be directly responsive to oestrogen stimulation.

1.6 The mouse Brca1 gene

Isolation of homologues of a gene can often provide clues as to which regions of the gene are functionally significant. This can be beneficial to mutation analysis of disease-predisposing genes where missense mutations are detected. In the absence of a functional assay the significance of these mutations may not be clear, but if the mutated amino acid is conserved in the mouse this might imply that it is significant. The mouse is frequently the chosen species for examination, as identification of the murine gene can lead to further studies that can be easily carried out in this species due to, amongst other reasons, its small size and short reproductive cycle. The mouse provides a system in which the expression of a gene can be followed through various stages of development, and hence the role of a gene in normal growth control and differentiation, as well as tumour development, can be studied. Hence, the mouse homologues of many tumour suppressor genes have been isolated (Table 1.3).

In addition, the ability to manipulate the murine genome by gene targeting and transgenic techniques can provide a model for disease studies. Knockout mice for a number of tumour suppressor genes have now been generated and, as expected, many of the heterozygous mice are cancer-prone (reviewed in Jacks, 1996). Some of these mice very closely mimic the human disease, for example Atm deficient mice are growth-retarded, have neurological abnormalities, are sensitive to γ-irradiation and hence, recapitulate the human A-T phenotype.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Homology</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53</td>
<td>78%</td>
<td>Zakut-Houri et al., 1985</td>
</tr>
<tr>
<td>RB</td>
<td>91%</td>
<td>Bernards et al., 1989</td>
</tr>
<tr>
<td>WT1</td>
<td>95%</td>
<td>Buckler et al., 1991</td>
</tr>
<tr>
<td>APC</td>
<td>90%</td>
<td>Su et al., 1992</td>
</tr>
<tr>
<td>NF1</td>
<td>98%</td>
<td>Bernards et al., 1993</td>
</tr>
<tr>
<td>NF2</td>
<td>98%</td>
<td>Claudio et al., 1994</td>
</tr>
<tr>
<td>VHL</td>
<td>90%</td>
<td>Gao et al., 1995</td>
</tr>
<tr>
<td>CDKN1NK4A</td>
<td>65%</td>
<td>Quelle et al., 1995</td>
</tr>
<tr>
<td>ATM</td>
<td>84%</td>
<td>Pecker et al., 1996</td>
</tr>
<tr>
<td>BRCA1</td>
<td>58%</td>
<td>Abel et al., 1995, Bennett et al., 1995, Sharan et al., 1995, Szabo et al., 1996</td>
</tr>
<tr>
<td>BRCA2</td>
<td>59%</td>
<td>Connor et al., 1997, McAllister et al., 1997, Sharan and Bradley, 1997</td>
</tr>
</tbody>
</table>

1 Homologies are indicated in terms of the percentage identity between the human and mouse predicted proteins.

Table 1.3 Human and mouse homologies for tumour suppressor genes.
(Barlow et al., 1996, Elson et al., 1996). The Min mouse contains a point mutation, created by chemical carcinogenesis, in the murine Apc gene and is perhaps the most accurate of animal models (Su et al., 1992). These mice develop intestinal adenomas histologically similar to those seen in patients with familial adenomatous polyposis (FAP). These mice have also illustrated their use in identifying gene modifiers which may be responsible for the variation seen within affected families. A major modifier for the Apc phenotype, Mom-1, has been located on mouse chromosome 4 and a candidate gene has now been isolated (MacPhee et al., 1995). Furthermore, mutant mice can be bred with other mutant mice to determine co-operative effects between tumour suppressor genes (Westphal et al., 1997, Williams et al., 1994).

Hence, mouse models allow investigation into the mechanism of tumour initiation and progression due to both genetic and environmental factors. These investigations may also reveal potential therapeutic strategies and provides a system in which they can be assessed.

1.6.1 Conservation of the BRCA1 gene in mice

Homologues of the BRCA1 gene have now been isolated. Mouse Brca1 is 58% identical to human at the amino acid level (Abel et al., 1995, Bennett et al., 1995, Sharan et al., 1995). Interestingly, this homology is lower than that observed for other tumour suppressor genes, although BRCA2 shows a similar level of conservation (Connor et al., 1997), (Table 1.3). However, two regions of the gene are particularly highly conserved: the amino terminal region encompassing the RING finger (97% similar to the human protein) and an acidic sequence near the carboxy terminal (91% similar to the human), suggesting that these domains may be functionally important. These two domains are also conserved in the canine BRCA1 protein (Szabo et al., 1996). The possible functions of these domains are discussed below.

Analysing sequences in the mouse analogous to those where presumed missense mutations occurred in humans revealed that six of eight affected amino acids were conserved in the mouse (Sharan et al., 1995). In contrast, only one of seven polymorphisms studied was found to be conserved between the two species. Hence, the mouse Brca1 sequence may provide another criterion by which to determine the significance of an amino acid change in humans.
1.6.2 Animal models of BRCA1-induced tumorigenesis

To understand the biological function of BRCA1 and to investigate its role in breast and ovarian tumorigenesis, several groups have attempted to generate mice that are null for Brca1 (Gowen et al., 1996, Hakem et al., 1996, Liu et al., 1996). The homozygous mutant mice do not survive, but die during early embryogenesis, between day 7.5 and 10.5. Each group reports a slightly different time of death and phenotype for their mice and this may be a reflection of the different mutations introduced into the mice. Two groups, whose mice show a less severe phenotype, knocked out exon 11, which has been shown to be alternatively spliced in humans and mice. If this splice variant has partial functional activity, then the mutant mice may retain partial function, and hence have a slightly less severe phenotype. The embryonic lethality in the Brca1 mutant mice is most probably due to a lack of proliferation, indicating the critical role of Brca1 in early mouse development. Interestingly, Boyd et al. reported a woman with two mutant BRCA1 alleles, who developed early-onset breast cancer but has no other phenotype, clearly different to the observation in the mouse (Boyd et al., 1995).

The heterozygous mice are, however, viable and healthy, showing no signs of disease up to eleven months of age. This is intriguing, given the phenotype in human heterozygotes, although it is possible that these mice will have a late-onset phenotype. Hence, it is obvious that differences in BRCA1-induced tumorigenesis are observed between humans and mice, which may reflect different consequences in different species, or may be due to different mutations. Although the function of BRCA1 may not be exactly the same in mice and humans, interactions between gene products and the biochemical pathways involved may well be conserved. So, whilst these mice provide a model in which to study Brca1 function, caution must be taken in drawing conclusions about the human situation based on the mouse.

1.7 Functions of the BRCA1 gene

Before the isolation of the BRCA1 gene, it was anticipated that this gene would be a tumour suppressor gene and it seems that this may be the case. Germline mutations and loss of heterozygosity have been reported in a number of tumours from affected families, implying that both alleles need to be inactivated, consistent with Knudson's two-hit hypothesis for tumour suppressor genes.
Chapter One: Introduction

(Cornelis et al., 1995, Knudson, 1971, Smith et al., 1992). Further support comes from studies on breast cancer cell lines where transfection of the wild-type BRCA1 was shown to inhibit cell growth (Holt et al., 1996). In addition, transfection of antisense oligonucleotides to BRCA1 into normal and malignant human mammary cell lines results in accelerated growth (Thompson et al., 1995) and transforms mouse fibroblasts (Rao et al., 1996). These results indicate that BRCA1 may function as a negative regulator of cell growth, like other tumour suppressors.

Alternatively, BRCA1 may fall into a proposed third category of genes, namely the DNA mismatch repair genes. Germline mutations in these genes (hMSH2, hMLH1, hPMS1, and hPMS2) have been found in patients with hereditary non-polyposis colorectal cancer (HNPCC) (Leach et al., 1993, Nicolaides et al., 1994, Papadopoulos et al., 1994). As with BRCA1, somatic mutations in these genes are rarely found in sporadic tumours (Liu et al., 1995). Defective DNA repair, due to mutations in these genes, leads to genetic instability, to an acceleration in the rate of mutations in oncogenes and tumour suppressor genes, and hence to tumour development (Lazar et al., 1994). Interestingly, genomic instability, in the form of microsatellite instability and multiple mutations in the TP53 gene, has been reported in breast tumours (Glebov et al., 1994, Patel et al., 1994). Hence, it is possible that mutations in the BRCA1 gene may be causing this mutator phenotype.

In order to elucidate the precise function of this gene, various approaches can be undertaken, including searching for protein motifs, identifying interacting proteins and determining the subcellular localisation of the protein. Results from these experiments and others are now beginning to provide clues to the function of BRCA1 (reviewed in Chambers and Solomon, 1997).

1.7.1 Subcellular localisation of the BRCA1 protein

Determining the subcellular localisation of BRCA1 has proved more difficult and controversial than might have been expected. This has highlighted problems in producing good antibodies and variations in staining depending on the cell fixation technique used. Several groups have reported BRCA1 to be localised exclusively in the cell nucleus (Chen et al., 1995, Scully et al., 1996, Thomas et al., 1996), whereas others report BRCA1 to be cytoplasmic, exhibiting properties of a granin (Jensen et al., 1996). Granins are a family of acidic secretory proteins found in the secretory granules of a wide variety of endocrine cells and neurons.
The significance of the granin motif in BRCA1 has been questioned, since this motif is not conserved in the mouse, unlike the RING finger and the BRCT domain.

Evidence is now mounting for a nuclear function for BRCA1. Three potential nuclear localisation signals (NLS) have been identified at amino acids 503-508, 606-615 and 651-656, all encoded by exon 11 (Chen et al., 1996a), (Figure 1.9). When constructs containing mutant forms of the first two (503-508 and 606-615) are transfected into human cell lines, the proteins are no longer detected in the nucleus but are instead found in the cytoplasm; if the third NLS is mutated, the protein is still nuclear (Chen et al., 1996a). This suggests that NLS 503-508 and NLS 606-615 are essential for transport into the nucleus. Nuclear import requires recognition of the NLS, nuclear pore docking, translocation through the pore and release from the inner side of the pore (reviewed in Ullman et al., 1997). The NLS receptor consists of two proteins, importin α, which recognises the NLS, and importin β, which mediates interaction with the nuclear pore. Not only does BRCA1 possess two NLSs; it also interacts with importin α (Chen et al., 1996a), indicating that this is most likely the mechanism by which BRCA1 is transported into the nucleus.

However, weak cytoplasmic as well as nuclear staining has been observed (Wilson et al., 1997). Many alternatively spliced BRCA1 transcripts have been reported (Lu et al., 1996, Thakur et al., 1997, Wilson et al., 1997, Xu et al., 1997b), some of which lack all or part of exon 11 which contains the NLS. Western blots have indicated that at least some of these transcripts encode proteins and may, therefore, be responsible for the observed cytoplasmic staining. It will be interesting to discover the function, if any, of these splice variants, since the WT1 tumour suppressor gene has also been shown to have a number of splice variants with differing transactivation potentials (Wang et al., 1995).

1.7.2 The BRCA1 RING finger may act as a protein-protein interaction domain

Several motifs have been identified within the BRCA1 protein (Figure 1.9). The first to be identified was the RING finger motif, which is a type of zinc finger believed to be involved in protein-protein interactions (Saurin et al., 1996). The high homology found between the human and the mouse RING fingers (97%) suggested that this domain may have functional significance. Indeed, a novel protein, named BARD1, has been found to interact with BRCA1 via the RING
Figure 1.9 Schematic diagram of the BRCA1 cDNA illustrating the functional domains. Vertical lines indicate the divisions between the exons, some of which are numbered. The solid horizontal lines indicate the identified motifs.
finger; mutations in the RING domain abolish this interaction (Wu et al., 1996).
BARD1 also has a RING finger that is required for this interaction. The function of BARD1 is unknown, but given its interaction with BRCA1 and that they are co-expressed, it may be essential for normal BRCA1 function.

1.7.3 A role for BRCA1 in transcription

Comparison of BRCA1 with other proteins revealed a domain that is found in a large number of proteins, including a p53-binding protein and the yeast cell cycle checkpoint control protein Rad9, many of which are involved in cell cycle regulation and DNA repair. This domain has been called the BRCT (BRCA1 C-terminus) domain and is conserved between humans and mice (Callebaut and Mornon, 1997, Koonin et al., 1996), (Figure 1.9). Interestingly, this domain has also been found in the BRCA1-interacting BARD1 protein (Wu et al., 1996) and other BRCT domain-containing proteins interact with each other: for example, DNA ligase III and XRCC1 (Bork et al., 1997). The BRCT domain is found in many proteins associated with DNA damage-responsive cell cycle checkpoints, suggesting a potential function for BRCA1.

In addition, the BRCA1 BRCT domain has been shown to be capable of activating transcription, when fused to a GAL4 reporter gene and transfected into either yeast or mammalian cells (Chapman and Verma, 1996, Monteiro et al., 1996). The minimal region required is encoded by exons 21-24 (amino acids 1760-1863). Mutations, but not polymorphisms, in this region abolish this transactivation, suggesting that this may be a real phenomenon (Monteiro et al., 1996, Monteiro et al., 1997). In addition, BRCA1 has been found to associate with the RNA polymerase II holoenzyme complex that is responsible for mRNA transcription, and hence further suggests a role for BRCA1 in transcription (Scully et al., 1997a).

1.7.4 BRCA1 and cell cycle regulation

Since a number of tumour suppressor genes have been shown to be involved in regulation of the cell cycle (for example, p53), expression of BRCA1 during the cell cycle has been studied. The results show that BRCA1 mRNA and protein levels are indeed regulated during the cell cycle (Figure 1.10). Low levels are observed in G0 and early G1, with levels increasing prior to S phase entry, remaining high throughout M phase, before decreasing again in early G1 (Chen et al., 1996b, Gudas et al., 1996, Ruffner and Verma, 1997, Thomas et al., 1997,
Figure 1.10 Regulation of BRCA1 during the cell cycle. Downward arrows indicate decreased levels of BRCA1 protein and upward arrows indicate increased levels. BRCA1-P = hyperphosphorylated BRCA1.
Vaughn et al., 1996b). In addition, a second, slower-migrating band was observed on Western blots as levels of BRCA1 increased. By treatment with phosphatase, this band was shown to be a hyperphosphorylated form of the BRCA1 protein (Ruffner and Verma, 1997, Thomas et al., 1997). Further studies showed that BRCA1 is predominantly phosphorylated on serine residues and that phosphorylation follows the same pattern as expression: that is, increasing at G1/S, remaining high through M phase before being dephosphorylated in G1 (Chen et al., 1996b, Ruffner and Verma, 1997, Thomas et al., 1997). Together these results suggest that BRCA1 may play a role at the G1/S transition and/or G2/M phase.

When cells are subjected to DNA-damaging agents (in the form of hydrogen peroxide, UV or γ irradiation), hyperphosphorylation of BRCA1 is again observed and the punctate nuclear staining of BRCA1 becomes diffuse (Scully et al., 1997c, Thomas et al., 1997). This strongly implies that BRCA1 may be involved in sensing or responding to these stimuli. BRCA1 has also been observed to co-localise and interact with Rad51, the human homologue of the bacterial protein, RecA, that is involved in DNA repair (Scully et al., 1997b). These proteins interact via sequences encoded by BRCA1 exon 11 (Figure 1.9). Consequently, a role for BRCA1 in the control of recombination and genome stability has been postulated.

Further confirmation of a role for BRCA1 in the cell cycle comes from a recent report (Somasundaram et al., 1997). In this study, wild-type BRCA1 transfected into colon cancer cells inhibited S phase progression, and thus showed that BRCA1 can negatively regulate the cell cycle. Since p21 expression can inhibit cell growth, the authors investigated whether BRCA1 could induce p21. Indeed, using reporter gene assays BRCA1 was shown to activate the p21 promoter 5-20 fold. More importantly, BRCA1 was shown to increase levels of endogenous p21. The NLS, the C-terminal transactivation domain and the Rad51-interacting domain were all necessary for p21 activation. Interestingly, expression of p21 has previously been shown to be required for cell cycle arrest in response to γ irradiation (Waldman et al., 1995). Hence, these results imply that BRCA1 can transactivate the p21 gene and thereby negatively regulate the cell cycle.

Therefore, progress is now being made in elucidating the function of the BRCA1 gene product. The BRCA2 gene shows a number of similarities to BRCA1 (Table 1.4) and it may be that these two breast cancer susceptibility genes function in the same or similar pathways. Whilst the precise function of BRCA1 remains to be
<table>
<thead>
<tr>
<th>Property</th>
<th>BRCA1</th>
<th>BRCA2</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene structure</td>
<td>~80kb genomic DNA, 7.8kb mRNA</td>
<td>~70kb genomic DNA, 10-12kb mRNA</td>
<td>Miki et al., 1994, Wooster et al., 1995</td>
</tr>
<tr>
<td>Conservation in the mouse</td>
<td>58% identity between human and mouse predicted proteins</td>
<td>59% identity between human and mouse predicted proteins</td>
<td>Abel et al., 1995, Connor et al., 1997</td>
</tr>
<tr>
<td>Motifs</td>
<td>RING finger, BRCT domain</td>
<td>BRC repeats</td>
<td>Bork et al., 1996, Koonin et al., 1996, Miki et al., 1994</td>
</tr>
<tr>
<td>Interacting proteins</td>
<td>BARD1, importin α, Rad51</td>
<td>Rad51</td>
<td>Chen et al., 1996, Mizuta et al., 1997, Scully et al., 1997, Wu et al., 1996</td>
</tr>
<tr>
<td>Phenotype of knockout mice</td>
<td>embryonic lethal</td>
<td>embryonic lethal</td>
<td>Gowen et al., 1996, Suzuki et al., 1997</td>
</tr>
<tr>
<td>Transactivation potential</td>
<td>C-terminus</td>
<td>N-terminus</td>
<td>Chapman and Verma, 1996, Milner et al., 1997</td>
</tr>
<tr>
<td>Expression</td>
<td>Cell cycle regulated, associated with cellular proliferation, oestrogen responsive</td>
<td>Cell cycle regulated, associated with cellular proliferation, oestrogen responsive</td>
<td>Spillman and Bowcock, 1996, Vaughn et al., 1996a, Vaughn et al., 1996b</td>
</tr>
</tbody>
</table>

Table 1.4 Similarities between the BRCA1 and BRCA2 genes.
unravelled, evidence is accumulating for a role for BRCA1 in transcription and the cell cycle. It may be possible that BRCA1 acts as a DNA damage checkpoint protein during the cell cycle, by increasing transcription of p21 and/or other genes, and hence arresting cell growth in response to DNA-damaging agents. In this model, BRCA1 would function as a 'caretaker' responsible for maintaining the integrity of the genome, as recently proposed (Kinzler and Vogelstein, 1997).
AIMS OF THIS THESIS

Initially the aim of this thesis was to identify novel transcripts within the BRCA1 region on chromosome 17q, to characterise these candidate genes and to determine their role, if any, in breast carcinogenesis.

During the course of this work, a strong candidate for BRCA1 emerged. Therefore, the second aim was to confirm the isolation of the BRCA1 gene, by performing mutation analysis on breast and breast/ovarian cancer families. The number of mutations identified would contribute to determining the proportion of early-onset breast and ovarian cancer that is due to mutations in the BRCA1 gene. This would also provide information on the spectrum of BRCA1 mutations present in the population and the possibility of screening those at risk.

The BRCA1 protein contains a RING finger domain, but shares no other homology with known proteins. In order to gain insight into the function of this protein, the third aim was to study the conservation of this region, particularly the BRCA1 gene and the neighbouring NBR1 gene, in other species. This could indicate significant regions of the genes and in addition, murine cDNA and genomic clones will act as resources for further studies on these genes, using the mouse as a model.

The final aim was to study the promoter regions of the murine Brca1 and Nbr1 genes, in order to provide insight into the regulation of the expression of these genes. This would highlight similarities and differences in the human and the mouse and would therefore, reveal the feasibility of the mouse as a model for studying the function of these genes in humans.
CHAPTER TWO: MATERIALS AND METHODS

The majority of these protocols have been described previously (Sambrook et al., 1989). Centrifugation, unless otherwise stated, was carried out in a Heraeus Megafuge 2.0R with a 8155 rotor. Details of buffers and solutions are given in section 2.18.

2.1 Preparation of DNA

2.1.1 Preparation of genomic DNA from blood samples and cell lines

DNA for zoo blots was prepared from the following cell lines (ICRF Cell Production Laboratory): human lymphoblastoid, mouse lymphoma, baboon lymphoblastoid, gibbon lymphoma, bovine epithelial, muntjac deer fibroblast, wallaby kidney, and chick primary fibroblasts using the sodium perchlorate method as described (Johns Jr. and Paulus-Thomas, 1989). DNA from blood and lymphoblastoid cell lines from family members of breast and breast/ovarian cancer families was also prepared for mutation analysis by this method.

Approximately 5 x 10⁷ pelleted cells, thawed on ice if necessary, were made up to 10 ml with sterile PBS and vortexed to resuspend. To this was added 35 ml lysis buffer. Alternatively, 35 ml lysis buffer was added directly to 10 ml whole blood. The solution was mixed and incubated on ice for 10 minutes, before centrifuging at 3000 rpm at 4°C for 15 minutes. The supernatant was discarded and the pellet was resuspended in 10 ml resuspension buffer. After vortexing to mix, 100 µl 10 mg/ml RNAse A and 100 µl 20% SDS were added and the solution was incubated at 37°C for 30-60 minutes. A further 750 µl 20% SDS was added, and the lysate was incubated at 60°C for 10 minutes, then allowed to cool to room temperature. Sodium perchlorate was added to a final concentration of 1M, mixed and an equal volume of chloroform/isoamyl alcohol was added. The solution was rotated gently for 30 minutes to mix, before centrifuging at 3000 rpm at 20°C for 15 minutes. The upper, aqueous, layer was transferred to a clean tube and precipitated with 2 volumes ice-cold ethanol. The DNA was hooked out, transferred to an eppendorf tube, washed with 70% ethanol and resuspended in 500 µl TE.
Chapter Two: Materials and Methods

2.1.2 Preparation of genomic DNA from tissue samples

Mouse genomic DNA was prepared from an adult 129J mouse liver (ICRF Biological Resources Laboratory). Tissue was supplied frozen and stored at -70°C. Approximately 0.5 g tissue was chopped into small pieces while still frozen, added to a solution containing 5 ml TE, 63 µl 20% SDS, and proteinase K at a final concentration of 0.2 mg/ml and incubated at 37°C for 12-16 hours. The solution was extracted twice with an equal volume of phenol and the upper, aqueous, layer transferred to a clean tube. RNAse A was added to the aqueous layer to a final concentration of 120 µg/ml and incubated at 37°C for 30 minutes. Proteinase K was then added to a final concentration of 80 µg/ml and the incubation was continued at 37°C for a further 15 minutes, before extracting with an equal volume of phenol. The solution was then extracted twice with an equal volume of chloroform before precipitating with 0.1 volume 4M NaCl and 2 volumes ethanol. DNA was hooked out and resuspended in 200-500 µl TE.

2.1.3 Preparation of plasmid/cosmid DNA by alkaline lysis

Cultures or glycerols were streaked onto LB agar plates supplemented with the appropriate antibiotic (50 µg/ml) to produce single colonies and incubated for 12-16 hours at 37°C. The human cosmids were kanamycin-resistant, all other plasmids and cosmids were ampicillin-resistant. Single colonies were inoculated into 50 ml LB medium with antibiotic and incubated at 37°C, with shaking, for 12-16 hours.

After incubation for 12-16 hours, a 50 ml culture of bacteria was centrifuged at 2800 rpm at 4°C for 20 minutes. The supernatant was discarded and the cell pellet was resuspended in 6 ml solution I. The cells were mixed and then 12 ml solution II was added. After gentle mixing, the solution was incubated on ice for 5 minutes. To this was added 8 ml solution III. The solution was shaken, incubated on ice for 10 minutes and then centrifuged at 2800 rpm at 4°C for 15 minutes. The supernatant was transferred to a fresh tube and incubated with 100 µl 10 mg/ml RNAse A at 37°C for 30 minutes. The solution was then extracted with 5-10 ml chloroform/isoamyl alcohol, followed by precipitation with 15 ml isopropanol at room temperature for 15 minutes. After centrifugation at 2800 rpm at 20°C for 20 minutes, the supernatant was discarded and the pellet washed with 5-10 ml 70% ethanol. The DNA pellet was then air-dried and resuspended in 500 µl TE.
Chapter Two: Materials and Methods

2.1.4 Preparation of high quality plasmid/cosmid DNA

A 50 ml culture was prepared as described above and DNA was prepared from the culture using a QIAGEN column (QIAGEN), following the manufacturer's recommended protocol.

2.1.5 Small-scale preparation of plasmid/cosmid DNA

Single colonies were inoculated into 1.5 ml LB medium with antibiotic and incubated at 37°C, with shaking, for 12-16 hours. The culture was centrifuged in a Heraeus Biofuge centrifuge for 30 seconds at 13000 rpm. Most of the supernatant was discarded, leaving approximately 50-100 μl in the tube in which the pellet was resuspended. To the resuspended pellet was added 300 μl TENS buffer and the solution was vortexed for 2-3 seconds. To this, 150 μl 3M NaOAc was added and centrifuged for 5 minutes at 13000 rpm, after briefly vortexing. The supernatant was transferred to a clean tube and DNA precipitated with 900 μl ethanol, pre-chilled to -20°C. Following centrifugation at 13000 rpm for 2 minutes, the pellet was washed with 70% ethanol and resuspended in 30 μl TE.

2.1.6 Preparation of bacteriophage lambda DNA

Bacteriophage were plated out at sufficient dilutions to obtain large single plaques (see section 2.11). A fresh plaque was then transferred to 50 ml LB medium containing MgSO₄, at a final concentration of 10mM. The culture was then incubated at 37°C, with vigorous shaking, for 12-16 hours until the cells were lysed. Five hundred microlitres of chloroform were added to the culture and the incubation was continued for a further 5 minutes. The culture was then centrifuged for 30 minutes at 2500 rpm at 20°C. The supernatant was transferred to a fresh tube and RNAse and DNAse were each added to a final concentration of 10 μg/ml, followed by incubation at 37°C for 30 minutes. To this, 5 ml 10 X proteinase K buffer was added, followed by proteinase K to a final concentration of 60 μg/ml and the solution was incubated at 37°C for 30-60 minutes. The solution was then precipitated with 0.6 volumes isopropanol and incubated on ice for 10 minutes, followed by centrifugation at 2500 rpm for 30 minutes. The supernatant was discarded and the pellet resuspended in 300-500 μl TE. The solution was then extracted twice with phenol/chloroform/isoamyl alcohol and twice with chloroform/isoamyl alcohol. Precipitation was carried out with 0.6 volumes isopropanol and 0.1 volume NaOAc. The solution was incubated at
room temperature for 10 minutes and then centrifuged at 13000 rpm for 15 minutes in a Heraeus Biofuge. Following washing with 70% ethanol, the pellet was resuspended in 200-400 µl TE.

2.1.7 Preparation of oligonucleotides

Oligonucleotides were supplied fully deprotected and dried down by the ICRF Oligonucleotide Synthesis Laboratory. Before use, the oligonucleotides were resuspended in 0.3M NaOAc, 10mM MgCl₂ and precipitated with 3 volumes cold ethanol. Following centrifugation, the pellets were washed in 80% ethanol and resuspended in TE, at a concentration of 500 ng/µl.

2.2 Preparation of RNA

2.2.1 Preparation of RNA from cell lines

RNA suitable for RT-PCR was prepared from cells grown in culture in a 10 cm dish. The medium was removed from the cell monolayer and 1 ml TRIzol reagent was added (Gibco BRL). The solution was pipetted to lyse the cells and then transferred to a microcentrifuge tube. After incubating at room temperature for 5 minutes, 200 µl chloroform was added. The solution was then centrifuged at 13000 rpm for 5 minutes at 20°C in a Heraeus Biofuge. The upper, aqueous, solution was transferred to a fresh tube, 500 µl isopropanol was added and the solution was incubated at room temperature for 5 minutes. Following centrifugation at 13000 rpm for 10 minutes, 1 ml 75% cold ethanol was added to the pellet. The solution was centrifuged at 4000 rpm for 5 minutes, air-dried briefly and resuspended in 150 µl DEPC-treated water.

2.2.2 Preparation of RNA from whole tissues

RNA was available from mouse tissues which had been prepared as described (Gonda et al., 1982).
2.3 Quantitation of Nucleic Acids

2.3.1 Quantitation using a spectrophotometer

The concentration of DNA was determined by the spectrophotometric measurement of the amount of ultraviolet (UV) light absorbed by the samples at a wavelength of 260 nm, where an optical density of 1.0 corresponds to 50 μg/ml of double-stranded DNA. RNA was quantitated similarly, with an OD_{260} of 1.0 equivalent to 40 μg/ml single stranded RNA.

2.3.2 Quantitation by gel analysis

DNA was also quantitated by gel electrophoresis. Samples were separated on an agarose gel (see section 2.4) adjacent to HindIII/EcoRI digested λ DNA (Northumbria Biologicals Limited) of known concentration. Gels were photographed using a Kodak Digital Science DC40 camera and sample DNA was quantitated relative to the standard using the BioMax ID Image Analysis software (Kodak).

2.4 Size separation of DNA by agarose gel electrophoresis

DNA was separated by electrophoresis through agarose gels in 1 X TAE buffer in a gel electrophoresis tank (Biorad). Routinely, a 0.8% gel was used, but the percentage of agarose was altered depending on the size of the fragments to be separated. Ethidium bromide was added to the gel to a concentration of 0.2 μg/ml. DNA could then be visualised using a UV transilluminator (UVP Dual intensity transilluminator, Genetic Research Instrumentation Limited). Gels were photographed using either a DS 34 Polaroid camera with Polaroid 667 film or a Kodak Digital Science DC40 camera and visualised using the BioMax ID Image Analysis software (Kodak). DNA size markers were separated alongside samples; the choice of standard depended on the predicted sizes of the sample fragments. The standards used were: 100 bp ladder (Pharmacia Biotech), HindIII/EcoRI digested λ DNA (Northumbria Biologicals Limited), PstI digested λ DNA (Northumbria Biologicals Limited) and 0.1-12 kb marker (R&D Systems).

To purify DNA fragments for using as hybridisation probes, DNA was separated on low melting point agarose as described above and visualised using the transilluminator. Fragments were then excised from the gel, added to 100 μl TE
Chapter Two: Materials and Methods

and stored at 4°C. Prior to using as probes for hybridisations, the excised fragments were heated to 65°C for 10 minutes in order to melt the agarose.

2.5 Restriction enzyme digestion of DNA

Genomic and plasmid DNA was digested with restriction enzymes (New England Biolabs and Northumbria Biologicals Limited) under the manufacturers' recommended conditions. Plasmid DNA was digested for 2 hours; genomic DNA was digested for 4-6 hours. Samples were separated on agarose gels to confirm that digestion was complete.

2.6 Amplification of DNA by the Polymerase Chain Reaction (PCR)

2.6.1 PCR reaction

Approximately 50-100 ng DNA was amplified in a 50 μl solution containing each dNTP at a final concentration of 0.125mM, forward and reverse primers each at a final concentration of 2 ng/μl, 1-2 units Taq DNA polymerase (Promega) and reaction buffer (Promega). The reactions were carried out in a PTC-100 Programmable Thermal Controller PCR machine (MJ Research Inc.), using the following conditions: 35 cycles of 94°C for 1 minute (denaturation), 55°C for 1 minute (annealing), 72°C for 1 minute (extension), followed by 1 cycle of 72°C for 5 minutes (final extension). These conditions were suitable for amplifying products of up to 1 kb in length; the extension time was increased by 1 minute for each additional kilobase. An annealing temperature of 55°C was suitable for primers with melting points (T_m) of 58°C, where T_m = 4 x (G+C) + 2 x (A+T). For primers with melting points above or below 55°C, the annealing temperature was adjusted accordingly.

PCR products that were to be subcloned were amplified using a proof reading polymerase such as Pfu polymerase (Stratagene), to minimise the number of sequence errors. The reaction was carried out essentially as above, but substituting the Pfu polymerase and Pfu reaction buffer (Stratagene) for the Taq polymerase and reaction buffer.
2.6.2 Purification of PCR products

PCR products were purified to remove excess primers, dNTPs and enzyme before subcloning or sequencing using one of the following techniques. The gel-based methods were used to separate the desired fragment from non-specific contaminants when a single specific fragment could not be obtained.

**Microspin columns**

Microspin S-400 columns (Pharmacia Biotech) were used to remove excess primers from PCR products prior to subcloning or automated sequencing. The PCR product was added to the column and eluted using the manufacturer's recommended conditions.

**QIAquick columns**

QIAquick spin columns (QIAGEN) were capable of purifying PCR products from primers, nucleotides, polymerases and salts, and hence were particularly useful for purification prior to subcloning. Purification of the PCR product was carried out under the manufacturer's recommended conditions.

**Spin-X columns**

Spin-X columns (Costar) were used to gel-purify PCR products. This purification method was particularly useful for PCR reactions where a single specific product was not easily obtained. The desired product was excised from a gel, added directly to the column and eluted under the manufacturer's recommended conditions. Where necessary, the eluted product was precipitated and resuspended in a smaller volume prior to sequencing.

**GENECLEAN**

Excised gel fragments were also purified using the GENECLEAN kit (Bio 101 Inc.) under the manufacturer's recommended conditions.

2.7 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

One microgram of total RNA or 200 ng poly A+ RNA was denatured at 65°C and then reverse transcription was carried out at 42°C for one hour in 0.125mM
dNTPs, 10 units RNAsin (Promega), 10 units AMV reverse transcriptase (Northumbria Biologicals Limited), 25 ng/μl random hexamers, p(dN)₆ (ICRF Oligonucleotide Synthesis Laboratory) in reverse transcriptase buffer (Northumbria Biologicals Limited). The synthesised cDNA was diluted in 100 μl TE and stored at -20°C. PCR reactions were carried out as described above, using 5 μl of the diluted cDNA.

2.8 Hybridisation of DNA to Southern blots

2.8.1 Southern Transfer of DNA

After electrophoresis, DNA was transferred to Hybond N⁺ membrane (Amersham) by alkali transfer. For DNA fragments larger than 10 kb, gels were depurinated in 0.25M HCl prior to transfer. The gel was placed on a dampened paper wick dipping into 0.4M NaOH. Membrane was placed on top of the gel, followed by a series of paper towels and a weight, so that DNA transferred from the gel to the membrane with the transfer buffer. Gels were blotted for 6-18 hours and then the membranes were neutralised in 2 X SSC for 30 minutes.

For allele specific oligonucleotide hybridisations (see section 2.10), two filters were required from each gel. In this case, a sandwich blot was used to transfer DNA to the membranes. Two sheets of 3MM paper (Whatman), cut to the size of the gel, were placed on a stack of paper towels. A third sheet of 3MM paper, dampened in 0.4M NaOH, was placed on top, followed by membrane pre-wetted in 0.4M NaOH and then the gel. A second sheet of dampened membrane was placed on top of the gel, followed by a dampened sheet of 3MM paper, two additional sheets of 3MM paper and a stack of paper towels. A small weight was placed on top and the gel blotted for 2-4 hours. Membranes were then neutralised in 2 X SSC for 30 minutes.

2.8.2 Radioactive labelling of DNA probes

DNA probes for hybridisation were labelled by one of two techniques. In general, oligo-labelling was the chosen technique. However, if short oligonucleotides were to be labelled, for example for allele specific oligonucleotide hybridisations, then end-labelling was more appropriate.
Oligolabelling

Fifty nanograms of DNA, to be used as a probe, was denatured for 10 minutes and labelled by the randomly primed oligo-labelling method (Feinberg and Vogelstein, 1984) with 2.5 μl [α-32P] dCTP (Amersham) in 1 X oligo-labelling buffer with 1 unit Klenow fragment DNA polymerase and BSA, at a final concentration of 0.4 mg/ml. The reaction was incubated at 37°C for 1 hour. The radioactively-labelled probe was separated from the unincorporated [α-32P] dCTP through a Sephadex G-50 spin column. The probe was denatured by boiling and, if necessary, preannealed at 65°C for 1-3 hours with a 2000-fold excess (w/w) of sonicated human placental DNA (Sigma) or human COT-1 DNA (Gibco BRL).

End labelling

Between 10-50 pmoles oligonucleotide probe was labelled with 3 μl [γ-32P] dATP (Amersham) and 10 units T4 polynucleotide kinase (Northumbria Biologicals Limited) in 50mM Tris-HCl, pH7.6, 10mM MgCl2, 5mM dithiothreitol (DTT), 0.1mM spermidine, and 0.1mM EDTA. The reaction was incubated at 37°C for one hour and purified through a Sephadex G-25 spin column.

2.8.3 Prehybridisation and hybridisation of DNA probes to filters

Filters were prehybridised at 65°C for 3-16 hours in 10-20 ml of: 5 X SSPE, 5 X Denhardt’s solution, 0.5% SDS, 0.01% yeast RNA. To compete out repetitive sequences, zoo blots and in some cases cDNA library filters were prehybridised for a further 2-3 hours, in the presence of denatured sonicated human placental DNA (Sigma) or human COT-1 DNA (Gibco BRL).

Filters were then hybridised in 10 ml prehybridisation solution, containing the labelled probe, with a specific activity of approximately 1 x 10^8 cpm/μg, at 65°C for 12-24 hours. After hybridisation, filters were washed in 1 X SSC, 0.1% SDS at 65°C for 30 minutes, followed by 0.2 X SSC, 0.1% SDS at 65°C for 30 minutes. Zoo blots were washed, first at room temperature in 2 X SSC, 0.1% SDS for 30 minutes, then 1 X SSC, 0.1% SDS for 30 minutes and then at 65°C in 1 X SSC, 0.1% SDS for 15 minutes. Filters were exposed to Kodak XAR5 or Fuji film in cassettes with intensifier screens, at -70°C for 1-7 days, before developing in an IGP film processor (IGP).
Probes were removed from filters in 0.5% SDS, preheated to 90°C, before reprobing filters.

2.8.4 Prehybridisation and hybridisation of oligonucleotide probes to filters

Prehybridisation and hybridisation were carried out essentially as described above, but incubated at 50°C. Filters were washed in 5 X SSC, 0.1% SDS at room temperature for 10 minutes and then in 5 X SSC, 0.1% SDS at 50°C for 10 minutes. Filters were exposed to film as described above.

2.9 Hybridisation of DNA probes to Northern blots

Multiple tissue Northern blots (Clontech) were prehybridised according to manufacturer's conditions in 5 X SSPE, 10 X Denhardt's, 2% SDS, 50% formamide and 100 μg/ml freshly denatured, sheared salmon sperm DNA, at 42°C for 3-6 hours. Blots were then hybridised in prehybridisation solution, with the addition of labelled probe with a specific activity of 5 x 10^8 cpm/μg at 42°C for 18-24 hours. Blots were washed in 2 X SSC, 0.05% SDS at room temperature for 40 minutes, with one change of solution, followed by 0.1 X SSC, 0.1% SDS at 50°C for 40 minutes, with one change of solution. For hybridisations of human probes onto mouse filters, the second wash was carried out in 2 X SSC, 0.05% SDS at 50°C for 40 minutes. Filters were exposed to film as described above. Probes were either removed from blots in 0.5% SDS, preheated to 90°C, or blots were left to decay before reprobing.

2.10 Allele specific oligonucleotide (ASO) hybridisation

Human genomic DNA from breast/ovarian cancer families was PCR amplified, as described in section 2.6 and one fifth of the reaction separated on an agarose gel. DNA was transferred to Hybond N+ using a sandwich blot (see section 2.8.1) and two filters obtained from each gel. One filter was hybridised to the wild-type oligonucleotide and the other to the mutant oligonucleotide.

2.10.1 Design of primers

Oligonucleotides were designed to be 15-17 nucleotides in length. Wherever possible, wild-type and mutant oligonucleotides were designed to be the same length and to have the same T_m. Oligonucleotides were designed such that the
mutation was in the central position of the oligonucleotide, to create maximal instability on hybridisation with normal DNA.

2.10.2 Hybridisation

Filters were prehybridised in 5 X SSPE, 5 X Denhardt's, 0.5% SDS, and 0.01% yeast RNA, for 30 minutes at 4-5°C below the T_m of the oligo. Filters were then hybridised in 5 X SSPE, 5 X Denhardt's, 0.5% SDS, 0.01% yeast RNA and 1 x 10^6 cpm/ml end labelled probe, for 1 hour at 4-5°C below the T_m. Filters were washed for 2-10 minutes at the T_m in 5 X SSC, 0.1% SDS. They were exposed to Kodak XAR5 or Fuji film in cassettes with intensifying screens, at -70°C for 2-24 hours.

2.11 Isolation of human and mouse cDNA clones

Commercial human placenta, foetal brain, HeLa, heart, foetal retina and ovary cDNA libraries were available from Stratagene and human foetal liver and testis cDNA libraries from Clontech. Human skeletal muscle cDNA libraries were donated by Dr. Yvonne Edwards.

Commercial mouse brain, testis and teratocarcinoma cDNA libraries were obtained from Stratagene.

2.11.1 Plating of cDNA libraries

A culture of the appropriate bacterial host strain was set up in LB medium with 0.2% maltose and incubated for 12-16 hours at 37°C, with shaking. The library bacteriophage stock was titred by making a series of 10-fold dilutions of the bacteriophage in phage storage medium (PSM). For each dilution, 300 µl of the prepared host culture was combined with 10 µl of the diluted bacteriophage and preabsorbed at 37°C for 20 minutes. Approximately 4-5 ml LM top agarose was then added to each tube and the contents of the tube poured onto 9 cm LM agar plates, swirling to spread the top agarose evenly. Plates were incubated at 37°C for 8-18 hours, until plaques were clearly visible. The number of plaques was counted and the titre calculated.

Libraries were plated out, either at 1 x 10^5 (low density) or 5 x 10^5 (high density) plaque forming units (pfu) per 22 cm x 22 cm plate, as described above. Four
plates were prepared for each library, allowing at least $4 \times 10^5$ pfu to be screened for each library. Plates were incubated at 37°C for 6 hours, or until the plaques were clearly visible but not confluent. Plates were stored at 4°C.

Lifts were taken from plates onto Hybond N+ membrane squares, 20 cm x 20 cm (Amersham). Filters were placed on plates and 5-10 registration marks made using a needle; these were also marked on the base of the plates. Filters were then placed with the DNA side facing up, in denaturing solution for 7 minutes. Next, filters were transferred to neutralising solution for 5 minutes and then to fresh neutralising solution for 3 minutes. Finally, filters were rinsed in 2X SSC for 3 minutes and then allowed to air-dry. Filters were baked for 2 hours at 80°C in a vacuum oven.

2.11.2 Screening of Libraries

Probes were hybridised to library filters as described in section 2.8. Conditions were essentially the same for the screening of human and mouse libraries. However, whenever human DNA probes were hybridised to mouse DNA filters, the stringency of washing was reduced. Any positives obtained were picked for further analysis: registration marks were transferred from the filters to the autoradiographs. Autoradiographs could then be aligned with the plate and a plug, corresponding to the positive clone, was picked into 1 ml PSM and 20 μl chloroform using a sterile Pasteur pipette. The tube was vortexed briefly to release the bacteriophage particles and incubated at room temperature for 1-2 hours. Dilutions were made from this stock, in the range of $10^{-3}$ to $10^{-5}$, and replated for a second round of screening. Lifts were taken as described above and the probe hybridised to these filters. When single positive plaques were obtained, the plaque was picked into PSM and chloroform, for isolation and purification of lambda DNA.

2.11.3 Isolation of purified clones

For the libraries obtained from Stratagene, inserts were in the Uni-ZAP XR or lambda ZAPII vector. These vectors have been designed to allow in vivo excision and recircularisation of any cloned insert contained within the lambda vector, to form phagemid containing the cloned insert. In vivo excision was carried out according to the manufacturer's recommended protocol (Stratagene). Briefly, 200 μl of XL1-Blue MRF' cells were combined with 100 μl bacteriophage stock plus 1
µl R408 helper phage (Stratagene) and incubated at 37°C for 15 minutes. Next, 5 
ml LB medium was added and the cultures were incubated for 3 hours at 37°C, 
with shaking. Cultures were then heated at 70°C for 20 minutes and centrifuged 
for 15 minutes at 3000 rpm. The supernatant, containing the pBluescript 
phagemid packaged as a filamentous bacteriophage particle, was decanted into a 
stere tube and stored at 4°C. The phagemid was rescued by combining 100 µl 
bacteriophage stock with 100 µl XL1-Blue MRF' cells. After incubating at 37°C for 
15 minutes, 2-50 µl of the culture was plated onto LB agar plates supplemented 
with ampicillin and incubated at 37°C for 12-16 hours. Colonies appearing on the 
plate contain the pBluescript double-stranded phagemid with the cloned DNA 
insert.

For other vectors (including most libraries from Clontech), bacteriophage DNA 
was prepared as described in section 2.1.6 and the inserts subcloned if necessary 
(see section 2.16)

2.11.4 Construction and screening of direct selection mini-libraries

Enriched cDNA libraries were made by direct selection in collaboration with B. 
Korn et al (see chapter one, Figure 1.6). Three libraries were constructed from 
cDNA libraries of adult skeletal muscle, foetal liver, foetal brain, and breast 
tissue, using pools of cosmids:
72G mini-library from cosmids C1198, D05717, B09127, C05123, F0829, G1151, 
H11167, B07165, G01152;
GAP2 mini-library from cosmids A1028, B0576, G01152, C05123, B07165, B0277;
GAP3 mini-library from cosmids F0520, F0345, G124, C02179.
Gridded library filters were screened as described in section 2.8.

2.12 Isolation of mouse genomic DNA clones

2.12.1 Plating of lambda bacteriophage libraries

A high-density mouse 129Sv genomic DNA library in λ2001 phage was plated 
out as described in section 2.11.1 (Donated by Drs F. Otto and M. Owen).
2.12.2 Plating of cosmid libraries

A mouse 129Sv genomic DNA cosmid library was donated by Dr A-M. Frischau. Hybond N membrane (Amersham) was laid on top of a 22 cm x 22 cm LB agar plate with ampicillin and 2.5 x 10^5 DH1 bacterial colonies were streaked out per plate, using a glass spreader. Membranes and plates were incubated together at 37°C, for approximately 14 hours. The membranes (primary filters) were then removed from the agar plates to make replicas. Fresh filters were laid on top of the primary filters and registration marks made. The primary and replica filters were separated and each was laid on a fresh agar plate. The primary filter was incubated for 1.5 hours at 37°C and then a second replica was made. The primary filter was then incubated for a further 2 hours. Replicas were incubated for 5 hours at 37°C. After incubation, the replica filters were denatured and neutralised as described in section 2.11.1, followed by baking at 80°C for 2 hours, in a vacuum oven. Primary filters were stored on a sheet of 3MM paper (Whatman), dampened in 1 X freezing solution at -80°C.

2.12.3 Screening of libraries

Filters were hybridised to probes as described in section 2.8. The stringency of the wash solutions was reduced when hybridising human DNA probes to mouse DNA filters. Positive bacteriophage plaques were subjected to second and third rounds of screening, where necessary, to obtain single positive clones. Positive cosmid clones were picked by aligning the autoradiograph with the primary filter and excising a membrane fragment into 500 μl LB medium. Dilutions were streaked onto LB agar plates containing ampicillin, lifts taken as above and filters hybridised, to obtain single positive colonies.

2.12.4 Isolation of purified clones

Lambda bacteriophage DNA was isolated and purified as described in section 2.1.6. DNA from positive cosmid clones was prepared as described in section 2.1.3.

2.13 Rapid amplification of cDNA ends (RACE)

The RACE technique was used to obtain the 5' ends of cDNAs when these were not isolated by screening cDNA libraries. This was carried out using the 5'
ampliFINDER RACE kit (Clontech), or the Marathon cDNA Amplification Kit (Clontech), according to the manufacturer's recommended conditions. Briefly, cDNA was synthesised from poly A+ RNA, using either a gene-specific primer or a modified oligo dT primer (Clontech). An adaptor was ligated to the cDNA and then the desired product was amplified by PCR, using an adaptor primer and a gene-specific primer. PCR primers were designed with linkers, to facilitate subcloning into the pDIRECT vector (Clontech). To increase specificity in the PCR reaction, a TaqStart antibody (Clontech) was used under the manufacturer's recommended conditions.

2.14 DNA sequencing

The sequencing of cDNA and genomic DNA clones described in chapter three was carried out manually; all other sequencing, described in chapters four to six, was performed using an automated sequencer.

2.14.1 Manual sequencing

DNA was prepared using QIAGEN columns (see section 2.1.4). For each reaction, 3 µg DNA was sequenced using 0.5-1.0 pmoles primer. Sequencing reactions were carried out using the dideoxynucleotide chain termination method, using a Sequenase version 2.0 T7 kit (United States Biochemical), following the manufacturer's recommended protocol. Briefly, templates were denatured, annealed to the primer and then labelled with [α-35S] dATP. The reactions were terminated and the samples were heated to 90°C for 5 minutes, before loading on a 5% polyacrylamide gel. The samples were separated on the gel by electrophoresing in 1 X TBE for 2.5 to 5 hours at 75 Watts. Gels were fixed in 10% methanol/10% glacial acetic acid, dried under vacuum and exposed to Kodak XAR5 or Fuji film for one to three days at room temperature.

2.14.2 Automated sequencing

Plasmid and cosmid DNA, prepared either by alkaline lysis or by QIAGEN columns, could generally be sequenced directly without the need for PCR. Bacteriophage DNA prepared as above could also be sequenced directly. PCR products were purified as described above. The amount of DNA required depended on both the quality of the preparation and the source of DNA.
Generally, 200 ng purified PCR product, 0.5 μg plasmid, 2 μg cosmid, or 2 μg bacteriophage DNA was sufficient in a 20 μl reaction.

**Dye terminator sequencing reactions**

Cycle-sequencing reactions were carried out using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems), according to the manufacturer's conditions. Briefly, DNA was added to a reaction containing 3.2 pmoles primer and premix (Applied Biosystems), consisting of dNTPs, fluorescently-labelled dyedeoxy terminators, and Taq polymerase in buffer. Parameters for cycle sequencing were 25 cycles of 96°C for 30 seconds, 50°C for 30 seconds, and 60°C for 4 minutes, in a PTC-100 PCR machine. Excess dye-labelled terminators were removed from the products by ethanol precipitation, using the manufacturer's recommended protocols (Applied Biosystems). Pellets were air-dried before resuspending in loading buffer (5 parts deionised formamide : 1 part 50mM EDTA, pH8.0). Loading buffer was supplemented with dextran blue to aid visualisation of the samples.

**Sequenase dye terminator sequencing reactions**

Exon 9 of the *BRCA1* gene could not be sequenced sufficiently accurately for mutation analysis by the above method. Therefore, it was sequenced using the PRISM Solid Phase Sequenase Dye Terminator DNA Sequencing Kit (Applied Biosystems). Prior to sequencing, PCR products were generated as described above, replacing one primer in each reaction with 4 pmole of the corresponding HPLC-purified, biotinylated primer. PCR products were immobilised on Dynabeads M-280 Streptavidin (Dynal), and the strands separated according to the method described by Dynal. Sequencing reactions were then carried out according to the manufacturer's recommended conditions (Applied Biosystems). Samples were ethanol precipitated, air-dried and resuspended in loading buffer.

**Preparation and running of gels**

Samples were heat-denatured and separated on a 4.75% polyacrylamide gel for 14 hours at 2500 Volts, 40 mAmps, 30 Watts on a 373A automated DNA sequencer (Applied Biosystems), in 1 X TBE buffer. Alternatively, samples were heat-denatured and separated on a 4.25% polyacrylamide gel for 7 hours on a 377 automated sequencer (Applied Biosystems), using the 2X sequencing program.
Data were collected and analysed using the Data Collection and Sequencing Analysis programs (Applied Biosystems).

2.15 Computer analysis of DNA sequences

Sequences were initially edited using the Sequencing analysis or EditView programs (Applied Biosystems). Sequence contigs were assembled using the Editseq and Seqman programs of Lasergene for Macintosh (DNASTAR Inc.) and edited further if required. Mutations were detected using either the Seqman program (Lasergene) or the Sequence Navigator program (Applied Biosystems).

Further analysis was carried out using the Wisconsin Package version 9.0, Genetics Computer Group (GCG), Madison, Wisconsin. Alignments were carried out using the BESTFIT and GAP programs, protein motif searches using the MOTIFS program, and FASTA and TFASTA programs were used for homology searches.

2.16 Subcloning of DNA fragments

2.16.1 Preparation of vector

Cosmid and bacteriophage fragments were subcloned into pBluescript KS (Stratagene). Fragments for reporter gene analysis were subcloned into the pGL3 basic vector (Promega). Vector DNA was digested with the appropriate restriction enzyme or enzymes, under manufacturer's conditions. If a single enzyme was used, then the vector DNA was treated with alkaline phosphatase (Boehringer Mannheim) to prevent re-ligation. Phosphatase was inactivated by phenol/chloroform extraction, followed by ethanol precipitation.

2.16.2 Preparation of insert DNA

Insert DNA was digested with the appropriate restriction enzyme or enzymes, under the manufacturer's conditions. Where necessary, fragments were gel-purified, using the GENECLEAN technique, or Spin-X columns, as described in section 2.6.2. PCR products to be cloned were purified using QIAquick columns, as described in section 2.6.2.
2.16.3 Preparation of competent cells

Plasmids were transformed into either DH5α or XL1 Blue MRF’ cells. Competent cells were prepared from single colonies.

Chemically competent cells

The method used was adapted from that previously described (Chung and Miller, 1988). Cells were inoculated into 20 ml LB medium and incubated at 37°C, with shaking, for 12-16 hours. Then 0.5 ml of this culture was used to seed 100 ml LB medium and incubated until the cells were in early log phase, with an OD$_{600}$ of 0.3-0.6. Cells were harvested at 2800 rpm for 10 minutes at 4°C and resuspended in 0.1 volume transformation and storage buffer. Cells were incubated on ice for 10 minutes and then aliquots were transferred to chilled tubes and stored at -70°C.

Electroporation competent cells

Cells were inoculated into 50 ml LB medium and incubated at 37°C for 12-16 hours. This culture was then diluted 1:100 in LB medium and grown to an OD$_{600}$= 0.5. The culture was chilled on ice, centrifuged at 3000 rpm at 4°C for 15 minutes and the cell pellet was gently resuspended in the original volume in cold 10% glycerol, 1mM Hepes, pH7.0. This procedure was repeated twice and then the pellet was resuspended in 1/50 the original volume in cold 10% glycerol. Cells were centrifuged and resuspended in 1/500 the original volume in 10% glycerol, and aliquots were stored at -70°C.

2.16.4 Ligations

Vector DNA (10 ng) was ligated with 50 ng insert DNA, in a 20 μl reaction volume containing 1 μl 10mM ATP, 1 μl 0.1M DTT, ligation buffer (Northumbria Biologicals Limited) and 1 μl (4 units) T4 DNA ligase (Northumbria Biologicals Limited). The reaction was incubated at 15°C for 12-24 hours.
2.16.5 Transformation procedure protocol

Chemical transformation

Competent cells were mixed with 5-10 µl ligation mix and incubated on ice for 10 minutes. Reactions were then heat-shocked at 37°C for 1 minute, and then chilled on ice for 1 minute. Reactions were spread onto LB agar plates, supplemented with the appropriate antibiotic and incubated for 12-16 hours at 37°C.

Transformation by electroporation

Ligations were ethanol precipitated to remove salts and resuspended in 20 µl sterile water. Competent cells (150 µl) were mixed with 5-10 µl ligation mix in a chilled electroporation cuvette (2 mm). The cells were electroporated using a Biorad gene pulser, with 2.5 kV, 200 Ohms and 25 µF. Next, 500 µl SOB medium was added and samples were plated onto LB agar plates containing the appropriate antibiotic.

2.16.6 Analysis of transformants

Colonies were picked into 1.5 ml LB medium containing antibiotic and incubated for 12-16 hours at 37°C. DNA was then prepared according to the method described in section 2.1.5. Clones were then analysed for the correct insert by restriction enzyme digestion, PCR or sequencing. For transformations with large numbers of colonies and a high vector background, colony lifts were taken, as described in section 2.11.1. The filters were hybridised to an insert-specific probe and positive clones picked and analysed as above.

2.17 Growing cell lines in culture for reporter gene assays

2.17.1 Passaging of cells

A mouse fibroblast cell line, 3T3K, was supplied by the ICRF Cell Production Laboratory. Cells were grown in DMEM medium (Gibco BRL), supplemented with 10% foetal calf serum (Gibco BRL), L-glutamine at a final concentration of 4 mM, 50 IU/ml penicillin (Gibco BRL) and 50 µg/ml streptomycin (Gibco BRL) and incubated at 37°C with 5% carbon dioxide. Prior to passaging, cells were rinsed with versene and then incubated with 0.025% trypsin in versene containing 1% phenol red, at room temperature, for 5 minutes. Detached cells
were neutralised with media and then centrifuged at 1100 rpm for 5 minutes in a MSE Centaur centrifuge. Cell pellets were washed in PBS and centrifuged again at 1100 rpm for 5 minutes. Cells were resuspended in media and plated at the desired dilution.

2.17.2 Transfection of cells in culture

Cells were transfected with 10 μg of the pGL3 basic vector (Promega), containing the test insert and 1 μg of a reference plasmid, pRL-TK (Promega). The pGL3 basic vector with no insert was used as a negative control and the pGL3 control vector (Promega) was used as a positive control.

Cells were prepared for transfection as described above and resuspended in PBS. Cells were counted using a haemocytometer and resuspended in PBS at a concentration of 1 x 10^6 cells/ml. DNA was added to a 4 mm electroporation cuvette and 800 μl cells (8 x 10^5) added. Samples were incubated at room temperature for 5-10 minutes and then electroporated using a Biorad Gene pulser at 200 volts and 950 μF. Samples were incubated at room temperature for a further 10 minutes and then each sample was carefully layered onto 2 ml foetal calf serum and centrifuged at 1500 rpm for 5 minutes in a MSE Centaur centrifuge, to remove dead cells. The supernatant was removed, the cell pellet gently resuspended in 1 ml DMEM and transferred to a well in a 24-well culture plate. Transfected cell cultures were incubated at 37°C for 40-48 hours before harvesting. The media was changed once after 24 hours to remove any dead cells.

2.17.3 Reporter gene assays

Luciferase assays were carried out using the Dual Luciferase Reporter Assay system (Promega) and luminescence was measured on a Turner Designs 20/20 Luminometer, under the manufacturer's recommended conditions. The use of dual reporters allows the simultaneous expression and measurement of two reporter enzymes within a single system. The pRL-TK reference vector contains the Renilla luciferase gene and the pGL3 vectors contain the firefly luciferase gene. These luciferases have different enzyme structures and substrate requirements, and hence it is possible to discriminate between their respective bioluminescent reactions.
Medium was removed from culture plates and the cells rinsed with PBS. Then 100 μl passive lysis buffer (Promega) was added to each well and the plates were incubated, with gentle shaking, at room temperature for 15 minutes. Cell lysates were transferred to fresh tubes and stored on ice, to be assayed for luciferase activity. To perform the assay, 20 μl cell lysate was transferred to a luminometer tube and placed in the luminometer. Into this 100 μl Luciferase Assay Reagent II was injected and the luminescence due to the firefly luciferase was measured. Then 100 μl Stop and Glo solution was injected and the luminescence due to the Renilla luciferase was measured. Promoter activity was measured as the firefly luminescence divided by the Renilla luminescence and compared to either the positive or the negative control.

### 2.18 Buffers and solutions

<table>
<thead>
<tr>
<th>Buffer/ Solution</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>chloroform/isoamyl alcohol</td>
<td>24 parts chloroform to 1 part isoamyl alcohol</td>
</tr>
<tr>
<td>denaturing solution</td>
<td>0.5M NaOH, 1.5M NaCl</td>
</tr>
<tr>
<td>50 x Denhardt's solution</td>
<td>5g Ficoll, 5g polyvinylpyrrolidone, 5g BSA in 500 ml dH₂O</td>
</tr>
<tr>
<td>10 X freezing solution</td>
<td>6.3g K₂HPO₄, 0.45g sodium citrate, 0.09g (NH₄)₂SO₄, 1.8g KH₂PO₄, 44g glycerol, in 100 ml H₂O; dilute to 1 X in LB medium</td>
</tr>
<tr>
<td>6 X gel loading buffer</td>
<td>0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in dH₂O</td>
</tr>
<tr>
<td>LB medium</td>
<td>10g bacto-tryptone, 5g bacto-yeast extract, 10g NaCl in 1 litre dH₂O, pH7.0</td>
</tr>
<tr>
<td>Material/Buffer Description</td>
<td>Composition</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>LB agar</td>
<td>15g agar in 1 litre LB medium</td>
</tr>
<tr>
<td>LM agar</td>
<td>0.2g MgSO4.7H2O in 100 ml LB agar</td>
</tr>
<tr>
<td>LM top agarose</td>
<td>0.2g MgSO4.7H2O, 0.7g agarose in 100 ml LB medium</td>
</tr>
<tr>
<td>lysis buffer (section 2.1.1)</td>
<td>320mM sucrose, 10mM Tris base, 5mM MgCl2, 1% Triton X-100, pH8.0</td>
</tr>
<tr>
<td>neutralising solution</td>
<td>0.5M Tris pH7.5, 1.5M NaCl</td>
</tr>
<tr>
<td>5 X oligo-labelling buffer</td>
<td>0.2M Hepes, pH6.6, 100nM/ ml each of dGTP, dATP, dTTP, 0.25M Tris, pH8.0,</td>
</tr>
<tr>
<td></td>
<td>25mM MgCl2, 0.35% 2-mercaptoethanol, 25 units/ml random hexamers (Pharmacia)</td>
</tr>
<tr>
<td>PBS</td>
<td>8g NaCl, 0.25g KCl, 1.43g Na2HPO4, 0.25g KH2PO4 in 1 litre dH2O, pH7.4</td>
</tr>
<tr>
<td>10 X proteinase K buffer</td>
<td>175.25g NaCl, 186g EDTA, 121.1g Tris base in 1 litre dH2O, pH7.6</td>
</tr>
<tr>
<td><strong>phage storage medium (PSM)</strong></td>
<td>5.8g NaCl, 2g MgSO₄·7H₂O, 50 ml 1M Tris, pH 7.5, 10 ml 2% gelatin in 1 litre dH₂O</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>phenol/chloroform/isoamyl alcohol</strong></td>
<td>25 parts phenol to 24 parts chloroform to 1 part isoamyl alcohol</td>
</tr>
<tr>
<td><strong>resuspension buffer (section 2.1.1)</strong></td>
<td>50mM Tris base, 150mM NaCl, 100mM EDTA, pH8.0</td>
</tr>
<tr>
<td><strong>SOB medium</strong></td>
<td>20g bacto-tryptone, 5g bacto-yeast extract, 0.6g NaCl, 0.2g KCl, 2g MgCl₂, 2.5g MgSO₄</td>
</tr>
<tr>
<td><strong>solution I</strong></td>
<td>50mM glucose, 25mM Tris.Cl, pH8.0, 10mM EDTA, pH8.0</td>
</tr>
<tr>
<td><strong>solution II</strong></td>
<td>0.2M NaOH, 1% SDS</td>
</tr>
<tr>
<td><strong>solution III</strong></td>
<td>600 ml 5M KOAc, 115 ml glacial acetic acid in 1 litre dH₂O</td>
</tr>
<tr>
<td><strong>20 X SSC</strong></td>
<td>175.3g NaCl, 88.2g NaOAc in 1 litre dH₂O, pH7.0</td>
</tr>
<tr>
<td><strong>20 X SSPE</strong></td>
<td>175.3g NaCl, 27.6g NaH₂PO₄, 7.4g EDTA in 1 litre dH₂O, pH7.4</td>
</tr>
<tr>
<td>Buffer Type</td>
<td>Composition</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------------</td>
</tr>
</tbody>
</table>
| 1 X TAE                   | 40mM Tris acetate   
1mM EDTA               |
| 1 X TBE                   | 50mM Tris base    
50mM Boric acid  
2mM EDTA               |
| TE                        | 10mM Tris.HCl, pH8.0   
1mM EDTA               |
| TENS buffer               | 0.1M NaOH,   
0.5% SDS in TE          |
| transformation and storage buffer | 100g PEG      
50 ml DMSO    
10 ml 1M MgSO₄   
10 ml 1M MgCl₂ in 1 litre LB medium |
CHAPTER THREE: ISOLATION OF CANDIDATE GENES WITHIN THE 
BRCA1 REGION ON CHROMOSOME 17q

The BRCA1 gene was localised to chromosome 17q and this was subsequently 
narrowed down to a 1-1.5 Mb region between the markers D17S776 and D17S78 
(Goldgar et al., 1994, Hall et al., 1990, Simard et al., 1993). A physical map 
covering this region was constructed in our laboratory, to facilitate the positional 
cloning of the gene. At the time of commencing the work described in this thesis, 
a cosmid contig was being constructed between the RNU2 locus and the PPY 
gene (see chapter one, Figure 1.5) (Jones et al., 1994), and therefore the task of 
isolating novel transcripts within this region began.

In order to identify novel transcripts and to produce the most comprehensive 
transcript map of this region, several techniques were used in our laboratory, 
including exon amplification (Brown et al., 1995) and the identification of CpG 
islands (Jones et al., 1994). To complement these methods, I chose to study the 
cosmids between RNU2 and PPY for cross-species homology. In addition, three 
mini-libraries, 72G, GAP2 and GAP3, were generated from pools of cosmids 
within this region, by direct selection (in collaboration with B. Korn), (see chapter 
one, section 1.4.4). The mini-libraries were constructed from adult skeletal 
muscle, foetal liver, foetal brain and breast tissue; the cosmids used in the 
construction of each library are illustrated in Figure 3.1.

3.1 Identification of transcripts by searching for cross-species homology

The principle behind this technique is that coding regions are more often 
conserved between species than are non-coding regions. This method has proved 
to be effective in the isolation of a number of genes for human disorders, 
including Duchenne muscular dystrophy (Monaco et al., 1986), Wilms’ tumour 
(Call et al., 1990), neurofibromatosis type I (Wallace et al., 1990) and 
adrenoleukodystrophy (Mosser et al., 1993). One advantage of this method is that 
genomic DNA is used and it is therefore not dependent on the expression of the 
gene; hence, this technique complements expression-dependent methods such as 
direct selection.

The minimal cosmid contig spanning RNU2 to PPY comprises 14 cosmids and 
covers a genomic distance of approximately 400 kb (Figure 3.1) (Jones et al., 1994).
Figure 3.1 Diagram showing the cosmid contig between the RNU2 locus and the PPY gene on chromosome 17q21. Cosmids are illustrated by black horizontal lines with their names written above. Cosmids boxed in blue indicate those used in constructing the 72G mini-library, green indicates those used for the GAP2 mini-library and red indicates those used for the GAP3 mini-library.
Chapter Three: Isolation of candidate genes

Each of these cosmids was digested with EcoRI, gel-purified and the resulting 82 fragments were individually hybridised to zoo blots, containing DNA from a variety of species, including human, mouse and chick. The fragments were named such that G1151-a was the smallest fragment of cosmid G1151; G1151-b the next smallest, and so on. The majority of these fragments did not show any evidence for conservation, or were highly repetitive (that is, they did not show specific hybridisation and may contain a repeat sequence such as an Alu element). However, cross-species homology was observed for 10 fragments isolated from 5 cosmids: C1198, A1028, G01152, B07165 and G1151 (Table 3.1, Figure 3.2). The extent of conservation varied: some fragments were conserved only in primates (C1198-e, C1198-f, C1198-g, A1028-c, G01152-ci and cii, G01152-ciii and B01765-d), while others were conserved in mammals and birds (A1028-h, B07165-c and G1151-a).

3.1.1 Cosmid C1198 transcripts

Three fragments from cosmid C1198 exhibited similar conservation in primates: C1198-e, C1198-f and C1198-g. Given the similar hybridisation patterns, it was likely that these fragments represent different regions of the same gene. The 72G mini-library was screened with each of these fragments, to obtain cDNA clones. The fragments hybridised to, amongst others, a clone, P9g1, already being characterised in our laboratory (Jones et al., 1994). Partial sequencing analysis showed that cosmid C1198 and P9g1 had homology to the env and pol genes of the human endogenous retrovirus type K locus (HERV-K). There are about 25-50 full-length copies of the HERV-K family in the human genome and the integration of these elements can disrupt the function of normal genes if they are inserted into the exons or the promoter regions (reviewed in Lower et al., 1996). Hence, this HERV-K related gene is unlikely to be the BRCA1 tumour suppressor gene and was not pursued.

3.1.2 Cosmid G01152 transcripts

Three genomic fragments from cosmid G01152 were conserved in primates, although they were slightly repetitive. A CpG island had been identified on this cosmid and had already been shown to be linked to a transcript on cosmids G01152 and B07165. This transcript was later identified as the VHR phosphatase (Jones et al., 1994) and it is possible that the conserved cosmid G01152 fragments comprise a portion of this gene. The VHR dual-specificity phosphatase is a
<table>
<thead>
<tr>
<th>Probe</th>
<th>Size of probe (kb)</th>
<th>Species in which conservation was observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1198-e</td>
<td>2.5</td>
<td>human and gibbon</td>
</tr>
<tr>
<td>C1198-f</td>
<td>3</td>
<td>human and baboon</td>
</tr>
<tr>
<td>C1198-g</td>
<td>6</td>
<td>human and gibbon</td>
</tr>
<tr>
<td>A1028-c</td>
<td>0.75</td>
<td>human and gibbon</td>
</tr>
<tr>
<td>A1028-h</td>
<td>5.8</td>
<td>human, baboon, gibbon, wallaby, deer, bovine, mouse and chick</td>
</tr>
<tr>
<td>G01152-ci and cii</td>
<td>2.9</td>
<td>human and gibbon</td>
</tr>
<tr>
<td>G01152-ciii</td>
<td>3.2</td>
<td>human and gibbon</td>
</tr>
<tr>
<td>B07165-c</td>
<td>1.8</td>
<td>human, gibbon, deer, mouse and chick</td>
</tr>
<tr>
<td>B07165-d</td>
<td>2.1</td>
<td>human, baboon and gibbon</td>
</tr>
<tr>
<td>G1151-a</td>
<td>0.5</td>
<td>human, baboon, gibbon, bovine, deer, mouse and chick</td>
</tr>
</tbody>
</table>

Table 3.1 Summary of results of hybridisations to zoo blots.
Figure 3.2 Hybridisation of cosmid fragments to zoo blots. The top panel shows EcoRI digested cosmids separated on an agarose gel, stained with ethidium bromide. The panels below indicate the hybridisation results with the fragments boxed in the panel above.
member of the protein tyrosine phosphatases which are implicated in cell signalling, cell growth and proliferation and oncogenic transformation (Ishibashi et al., 1992). However, analysis of this gene in breast and breast/ovarian cancer families did not reveal any mutations (Jones et al., 1994, Kamb et al., 1994a).

3.1.3 Cosmid B07165 transcripts

Two fragments from cosmid B07165 were conserved between human and other species, but differed in the extent of their conservation. Fragment B07165-d, which was conserved in primates, was used to screen human placenta, foetal brain and HeLa cDNA libraries (4 x 10^5 clones from each library). No positive clones were obtained. The fragment was hybridised to a Northern blot containing poly A+ RNA from a range of human adult tissues (Clontech), but no signal was detected. This suggests that this fragment may not encode a gene, or that it is not well represented either in the tissues on the Northern blot or in the cDNA libraries screened.

Fragment B07165-c, which was conserved in human, gibbon, mouse, deer and chick (Figure 3.2), hybridised to several clones in the GAP2 and 72G mini-libraries. However, no positive clones were obtained from screening human placenta, foetal brain, and HeLa cDNA libraries (4 x 10^5 clones from each). Northern blot analysis, using this fragment as a probe, detected transcripts of 1.8 kb and 4.8 kb in heart, skeletal muscle and pancreas. The 1.8 kb transcript was also detected in thymus, prostate, testis, ovary, small intestine, colon and peripheral blood leukocytes (Figure 3.3). These results suggested that B07165-c does contain coding sequence of a gene; however, the transcript encoded by this gene is not well represented in the cDNA libraries screened.

3.1.4 Cosmid G1151 transcripts

Cosmid fragment G1151-a was conserved in all species tested (Figure 3.2). However, no positive clones were obtained on screening the 72G mini-library with this fragment. To determine in which tissues this gene might be expressed, the fragment was hybridised to a Northern blot (Figure 3.3) and a transcript of 1.3 kb was observed in adult heart and skeletal muscle. No signal was detected on screening cDNA libraries from 14 additional tissues, indicating that the expression of this gene is highly tissue-specific. Since expression was detected in heart and skeletal muscle, cDNA libraries from these tissues were screened.
Figure 3. Hybridization of cosmids fragments to multiple tissue Northern blots.

Probe: C1151-a

Blots

Probe: A1028-h

Probe: B07165-c
However, no positive clones were obtained on screening $4.7 \times 10^6$ clones from skeletal muscle libraries and $4 \times 10^5$ from a heart library.

Seven additional cDNA libraries were subsequently screened, using the G1151-a fragment as a probe. One positive clone, GR1, was obtained from a human foetal retina library. GR1 was 1.1 kb in size and hybridised to chromosome 17. This clone and the G1151-a genomic fragment were entirely sequenced. Database searches between the G1151-a fragment and known sequences revealed no homology, suggesting that it is a novel gene. The GR1 clone had an \textit{Alu} repeat at one end and overlapped with G1151-a by 100 bp at the other end, but no other homology to any known genes was identified. In addition, several other repeat elements, including LINES and SINES, were identified within GR1. Northern blot analysis, using GR1 as a probe, revealed no detectable signal, and therefore it is possible that this clone is not derived from the G1151-a gene.

Because of the difficulties in isolating cDNAs from this fragment of cosmid G1151, 72 mini-library clones hybridising to the whole cosmid were investigated, since they could be 3' or 5' ends of this gene. Seven clones were picked and grown for analysis. Of these, three contained no insert. The remainder were hybridised to cosmid DNA and genomic DNA from chromosome 17, to confirm that they were from this region and not from a related gene elsewhere in the genome. This would indicate which clones would be suitable probes for Northern blot analysis, to show whether or not their tissue expression and transcript size were the same as the G1151-a gene.

One clone was very repetitive, and therefore an unsuitable probe for Northern blot analysis. Three clones mapped to chromosome 17 and two of these have been used as probes on Northern blots, but no signal was detected for either probe in any of the tissues tested, including heart and skeletal muscle. The remaining clone has not yet been tested on Northern blots.

Therefore, only one cDNA clone, GR1, has been identified to date, for this gene. However, the lack of signal on Northern blots and the repetitive elements within this clone, suggest that it may be non-specific. The Northern blot result with the G1151-a probe indicates that this fragment does contain expressed sequence, but it appears to be expressed at low levels and in a narrow range of tissues, which has made isolation of cDNAs difficult.
3.1.5 Cosmid A1028 transcripts

Two conserved fragments were identified on cosmid A1028: fragment A1028-c, which is conserved in human and gibbon and fragment A1028-h, which is conserved in a number of species including human, gibbon, wallaby, mouse and chick (Figure 3.2). When these fragments were hybridised to the GAP2 mini-library, positives clones were obtained with each and Northern blot analysis using fragment A1028-h as a probe showed expression of a 4.5 kb transcript in a range of tissues (Figure 3.3), suggesting that this cosmid contains expressed sequences. This is supported by the detection of a CpG island on this cosmid (K. A. Jones, personal communication), which is a marker for the promoter region of a gene.

Four clones from the GAP2 mini-library, that hybridised to fragment A1028-h, were picked and grown for analysis. Three clones clearly mapped to chromosome 17. These clones were sequenced, but only one clone, GAP2-K10, gave clear sequence and database searches with this sequence revealed no homology with other genes.

Cosmid fragment A1028-h was used to screen a HeLa cDNA library (4 x 10^5 clones). Three positive clones were obtained, of which two had fragments that did not hybridise to cosmids from this region and thus were discarded. The remaining clone, HA1, was sequenced in its entirety (1.3 kb) and found to overlap with GAP2-K10. An additional 1 x 10^6 clones from the HeLa cDNA library were screened with A1028-h and five positive clones were obtained. These clones were sequenced from either end, using vector primers, and database searches performed to identify homology with other genes. One clone had high homology to mitochondrial DNA and one had homology to a ribosomal protein and so both were discarded. The third clone was the same as HA1 and the two remaining clones, hah7 and hah2, did not have any homology to known genes, although hah2 was contained within hah7. Sequence homology was found between the hah7 clone and genomic sequence of a PAC from this region (accession number: AC003098).

Four clones from the GAP2 mini-library, that hybridised to fragment A1028-c, were also picked for analysis. All four mapped to chromosome 17 and all were sequenced. Three clones did not overlap with HA1 or GAP2-K10. One clone, GAP2-L2, overlapped by 90 bp with HA1, but then the sequences diverged. This
suggested that either these cDNAs were alternatively spliced forms of the same gene, or they were from related genes, or that some clones contained genomic sequence.

Subsequent database homology searches have shown that HA1, GAP2-K10 and GAP2-L2 form part of a new gene, DLG3, which was identified by Smith et al. during their isolation of candidate genes within the BRCA1 region (Figure 3.4) (Smith et al., 1996a). The DLG3 gene has a coding region of 1755 bp, consisting of 15 exons. The sequence from HA1, GAP2-K10 and GAP2-L2 is identical to nucleotides 325-1170 of DLG3. HA1 diverges from GAP2-L2 and DLG3 at the boundary of exons 6 and 7 (nucleotide 946). This, together with sequence comparison between HA1 and a PAG from this region (accession number: AC003098), suggests that the divergent sequence in the HA1 cDNA clone may be genomic sequence resulting from partial splicing. The hah7 clone appears to be intronic sequence, as compared to the PAC and DLG3 sequences.

The DLG3 gene shares homology with the Drosophila discs-large tumour suppressor gene, dlg-A, in which, mutations cause overgrowth of the imaginal discs (Woods and Bryant, 1991). The discs-large family is thought to be involved in signal transduction and in mediating protein-protein interactions at the cytoplasmic surface of the cell membrane. The DLG3 gene product has three regions conserved with other members of the discs-large family of proteins: a Drosophila discs-large homology region (DHR), an SH3 (src homology region 3) motif and a guanylate kinase (GK) domain (Smith et al., 1996a). DLG3 lies near DLG2, which is another discs-large gene within the BRCA1 region (Mazoyer et al., 1995). The DLG2 gene was analysed in sporadic breast cancers showing LOH at chromosome 17q. However, no mutations were detected (Mazoyer et al., 1995). Therefore, DLG2 does not appear to be involved in breast tumorigenesis. Nevertheless, given the tumour suppressor activity of dlg-A, it is possible that mutations in DLG2 and DLG3 may be involved in the development of other types of tumours.

3.2 Analysis of clones from mini-libraries created by direct selection

Three mini-libraries, 72G, GAP2 and GAP3, had been constructed by direct selection, using cosmids from the RNU2 to PYY region (Figure 3.1). Whole cosmids were hybridised to each mini-library, in order to identify positive clones, and hence transcripts across the region. Screening of the 72G and GAP2 mini-
Figure 3.4 Schematic diagram illustrating the isolated cDNA clones from the *DLG3* gene. Thin horizontal lines indicate cDNA clones and thick horizontal lines indicate homology domains: DHR = *Drosophila* homology domain, SH3 = *src* oncogene homology motif 3, GK = guanylate kinase domain. The dashed line indicates the portion of cDNA clone HA1 that does not match the *DLG3* cDNA. The translation start and stop sites of *DLG3* are shown.
libraries with the appropriate cosmids revealed positive clones scattered throughout this region, many of which corresponded to transcripts identified by other methods. When whole cosmids were hybridised to the GAP3 mini-library, no positive clones were obtained. This may have been due to these cosmids being more repetitive than the cosmids used to generate the 72G and GAP2 libraries, or due to the proportion of coding DNA to non-coding DNA being too low to obtain clear results. Using cosmid fragments could overcome this problem. To determine which fragments to use, cosmids were digested with EcoRI, separated on agarose gels, blotted onto filters and probed with the pool of PCR products of the cDNAs from the GAP3 mini-library. Hybridisation was observed to 12 cosmid fragments. Four of these fragments were known to contain genes already being isolated in our laboratory, now known to be the human dual-specificity phosphatase and the \textit{HAL64} gene (Jones \textit{et al.}, 1994). The remaining eight cosmid fragments were hybridised to the mini-library filters, to identify any unknown genes. Each of these fragments identified positive clones in the mini-library. Since these cosmids represented a region for which few cDNAs had been isolated, 8 clones were picked for analysis (Table 3.2).

3.2.1 Transcripts from cosmids F0520 and F0345

Three clones, GAP3-O3, GAP3-J6 and GAP3-P21, detected by screening with fragments from cosmids F0520 and F0345, were analysed. Each clone was hybridised to the cosmids and to genomic DNA from chromosome 17, to confirm that they were derived from this region. GAP3-P21 was highly repetitive and GAP3-J6 did not hybridise to chromosome 17 and so these clones were not pursued.

The GAP3-O3 cDNA clone, 0.76 kb, was sequenced. Database searches revealed no homology with any known genes at either the protein or the nucleotide levels. Hybridisation of this clone to a multiple tissue Northern blot revealed a 0.8 kb band of very low intensity in prostate and colon, suggesting that this may be a novel gene with high tissue-specificity (Figure 3.5).

3.2.2 Cosmid G124 transcripts

GAP3-L10, GAP3-E1 and GAP3-N10 were detected by hybridisation with cosmid G124. Further analysis revealed that GAP3-L10 was highly repetitive and that GAP3-N10 did not map to chromosome 17. GAP3-E1 hybridised to both the
<table>
<thead>
<tr>
<th>Name of clone</th>
<th>Cosmid fragments with which the clone was detected</th>
<th>Size of clone (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAP3-O3</td>
<td>F0520-d, F0345-b, F0345-c</td>
<td>0.76</td>
</tr>
<tr>
<td>GAP3-J6</td>
<td>F0520-b, F0345-c</td>
<td>0.4</td>
</tr>
<tr>
<td>GAP3-P21</td>
<td>F0520-d, F0345-b</td>
<td>0.65</td>
</tr>
<tr>
<td>GAP3-L10</td>
<td>G124-d</td>
<td>0.9</td>
</tr>
<tr>
<td>GAP3-E1</td>
<td>G124-b</td>
<td>0.6</td>
</tr>
<tr>
<td>GAP3-N10</td>
<td>G124-b</td>
<td>0.7</td>
</tr>
<tr>
<td>GAP3-L1</td>
<td>C02179-f</td>
<td>0.7</td>
</tr>
<tr>
<td>GAP3-O12</td>
<td>C02179-f</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Table 3.2 Clones from the GAP3 mini-library that were picked for analysis.
Figure 3.5 Hybridisation of GAP3 cDNA clones to multiple tissue Northern blots.

- Probe: GAP3-L1
  - heart
  - brain
  - placenta
  - lung
  - liver
  - skeletal muscle
  - kidney
  - pancreas
  - 1.3 kb
  - 2.4 kb
  - 4.4 kb
  - 7.5 kb

- Probe: GAP3-O3
  - spleen
  - thymus
  - prostate
  - testis
  - ovary
  - small intestine
  - colon
  - peripheral blood leukocytes
  - 1.3 kb
  - 2.4 kb
  - 4.4 kb
  - 7.5 kb
cosmid and genomic DNA from chromosome 17. Northern blot analysis with this clone as a probe showed low-intensity transcripts of approximately 5 kb and 8 kb in brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (Figure 3.5). GAP3-E1 was sequenced and database searches have revealed that this clone shares homology with HRH1, a putative RNA helicase and homologue of the yeast PRP22 mRNA splicing protein (Ono et al., 1994). The product of the yeast PRP22 gene acts late in the splicing of yeast pre-mRNA, mediating the release of the spliced mRNA from the spliceosome (Company et al., 1991). The HRH1 gene was also detected by other groups (Brody et al., 1995, Friedman et al., 1995a, Osborne-Lawrence et al., 1995).

3.3.3 Cosmid C02179 transcripts

GAP3-L1 and GAP3-O12 were identified by the same cosmid fragment and may represent the same gene. GAP3-L1 hybridised to several transcripts, including transcripts of 3.5 kb and 7.8 kb in various tissues on Northern blots (Figure 3.5). GAP3-L1 was sequenced and database searches revealed identity to an expressed sequence tag from a human skeletal muscle cDNA library (accession number: Z25191).

Hence, adult skeletal muscle cDNA libraries (4.3 x 10^6 clones) were screened using GAP3-L1 as a probe, but no positive clones were identified. Subsequent screening of ovary, placenta, foetal liver, testis (4 x 10^5 clones each) and foetal retina cDNA libraries (2 x 10^6 clones) revealed no further cDNAs for this gene.

3.3 Summary of transcripts identified

The techniques of searching for cross-species homology and direct selection have been successfully used to isolate and localise at least eight genes within the BRCA1 region (Figure 3.6). Whilst several genes were detected by both methods (for example, the DLG3 gene), some were detected by only one method (for example, G1151-a gene was only detected by cross-species homology). A number of other genes in this region were isolated in our laboratory by exon amplification, or further analysis of the mini-libraries (ET6B, 8Eg2, HAL64, and MOX1) (Figure 3.6). This reinforces the importance of using more than one technique to achieve a comprehensive transcript map of a region.
Figure 3.6 Transcript map of the region between the RNU2 locus and the PPY gene on chromosome 17. Cosmids are illustrated by horizontal lines with their names written above. Transcripts are illustrated by filled rectangles.
The BRCA1 gene was subsequently identified (Miki et al., 1994). It maps centromeric to the RNU2 locus, and hence was not detected in the region described in this chapter. However, several genes of potential biological significance were isolated in this region and are described below.

- **VHR phosphatase**, a vaccinia-related dual-specificity tyrosine/serine phosphatase (Friedman et al., 1995a, Ishibashi et al., 1992, Jones et al., 1994, Kamb et al., 1994a). Protein tyrosine phosphatases have been implicated in cell growth, proliferation, signalling, cell cycle regulation and oncogenic transformation.

- **MOX1**, the human homologue of the mouse mox1 gene, which contains a homeobox domain (Brody et al., 1995, Futreal et al., 1994a, Jones et al., 1994). Homeobox genes may regulate growth and differentiation.

- **HAL64**, a human ADP ribosylation factor-like gene (Harshman et al., 1995, Jones et al., 1994, Smith et al., 1995). This gene has homology to GTP binding proteins and in particular to ADP ribosylation factors, suggesting a potential role in signal transduction.

- **DLG2** and **DLG3**, human homologues of the *Drosophila* tumour suppressor *dlg-A* (Mazoyer et al., 1995, Smith et al., 1996a). The DLG2 and DLG3 genes encode proteins with three conserved regions: a 90 amino acid repeat domain, an SH3 (src homology region 3) motif and a guanylate kinase domain. These motifs are shared among the discs-large family of proteins, which may be involved in signalling pathways through interactions with G-protein binding proteins.

- **HRH1**, a human homologue of the yeast PRP22 mRNA splicing factor (Brody et al., 1995, Friedman et al., 1995a, Osborne-Lawrence et al., 1995).
CHAPTER FOUR: MUTATION ANALYSIS OF THE BRCA1 GENE IN
BREAST AND BREAST/OVARIAN CANCER FAMILIES

A strong candidate for the BRCA1 gene was isolated in 1994 (Miki et al., 1994). To
confirm that the BRCA1 gene had been found, many groups across the world
focused on mutation analysis of this gene in breast and breast/ovarian cancer
families. Epidemiological data have suggested that mutations in this gene may
account for 45% of inherited breast cancer and almost all of inherited
breast/ovarian cancer (Easton et al., 1993). Hence, mutation analyses would
confirm the proportion of cancers due to BRCA1 mutations. In addition, such
analyses would facilitate studies on genotype/phenotype correlations and would
also open up the possibility of genetic screening for individuals at risk of
developing breast or ovarian cancer.

The Imperial Cancer Research Fund has a large collection of breast and
breast/ovarian cancer families from London, Oxford and Leeds. A total of 53
eyear-onset breast and breast/ovarian cancer families were selected for the
mutation analysis described in this chapter.

4.1 Mutation analysis in breast and breast/ovarian cancer families with strong
evidence for linkage to BRCA1

Linkage analysis to both BRCA1 and BRCA2 had been performed on 33 of the
selected families, of which 19 had positive lod scores for BRCA1 (Spurr et al.,
1993; DT Bishop, unpublished data). Four families were identified as being
particularly likely to harbour BRCA1 mutations, due to a positive lod score for
BRCA1, a low or negative lod score for BRCA2 and the presence of at least one
case of ovarian cancer in the family (Table 4.1).

A highly-sensitive mutation detection technique was required to analyse DNA
from these families. Direct sequencing is the most sensitive method although it is
labour-intensive. However, the advent of automated sequencing has greatly
enhanced the efficiency of this technique and so it was the technique chosen for
analysing the four families (BC3, BOV2, BOV3, and PRU7B5).

One affected member from each family was selected for analysis. Primers were
designed within the introns of the BRCA1 gene, such that each exon and flanking
<table>
<thead>
<tr>
<th>Family ID</th>
<th>Number of breast cancers</th>
<th>Number of ovarian cancers</th>
<th>Number of other cancers</th>
<th>Lod score for linkage to BRCA1</th>
<th>Lod score for linkage to BRCA2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC3</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>1.41</td>
<td>0</td>
</tr>
<tr>
<td>BOV2</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>0.33</td>
<td>-0.11</td>
</tr>
<tr>
<td>BOV3</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>7.31</td>
<td>0</td>
</tr>
<tr>
<td>PRU7B5</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>0.28</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 4.1 Breast and breast/ovarian cancer families analysed by sequencing for BRCA1 mutations.
splice sites could be amplified by PCR and sequenced (Appendix A). Exon 11 is approximately 3.5 kb in length and so it was amplified in four segments. PCR products were sequenced in both directions, using the forward and reverse primers. Due to the extreme size of the \textit{BRCA1} coding region (approximately 5.7 kb), this work was undertaken in collaboration with C-F Xu in our laboratory.

Sequencing revealed mutations in two families (BOV3 and BC3). A 2 bp (AG) deletion was detected in exon 2 in the BOV3 family (Figure 4.1). This mutation is predicted to cause a frameshift leading to premature termination 16 amino acids downstream, resulting in a severely truncated protein. Additional BOV3 family members were analysed for the presence of this mutation. Five affected individuals carried the 185delAG mutation and five unaffected individuals, assessed as non-carriers following linkage analysis, did not have the mutation. This indicated that the mutation was segregating with the breast and ovarian cancers and that it was likely to be the disease-causing mutation (Figure 4.2). It was important to confirm that this was not a polymorphism present in the general population. Whilst this may seem unlikely for an alteration resulting in premature termination, a polymorphic stop codon has been reported in the \textit{BRCA1} gene (Mazoyer et al., 1996). Forty random DNA samples, collected in London, were analysed by allele specific oligonucleotide hybridisation for the presence of the 185delAG mutation. The mutation was not detected in any of these samples.

The BC3 family had a 1 bp deletion in exon 11 (3896delT), leading to a frameshift and premature truncation. This mutation was also shown to segregate with the disease and was not found in the general population.

In the remaining two families (BOV2 and PRU7B5) no mutations were detected although the entire coding region was sequenced in both directions. Since a regulatory mutation (resulting in the expression of only one allele) had been reported in a family (Miki et al., 1994), the BOV2 and PRU7B5 families were investigated for the presence of regulatory mutations. This was performed by looking for loss of heterozygosity at the RNA level, where heterozygosity had been observed in the genomic DNA. Sequencing analysis had revealed that the affected individual from PRU7B5 was heterozygous for a number of polymorphisms at the genomic level. A polymorphism in exon 13 was selected, since the 'T' to 'C' sequence change abolishes an EarI restriction enzyme site. Using primers from exons 13 and 15, RNA from this individual was analysed by
Figure 4.1 Identification of the 185delAG mutation in the BOV3 breast/ovarian cancer family by direct sequencing. The top panel shows sequence from an unaffected family member; the bottom panel shows sequence from an affected family member, carrying the mutation. Arrows indicate the position of the mutation.
Figure 4.2 Pedigree of the BOV3 family indicating the individuals known to have the 185delAG mutant allele and those having the wild-type allele.
RT-PCR. Digestion of the PCR product with EarI would distinguish between the allele containing 'T' and the allele containing 'C'. Upon digestion of the RT-PCR product with the EarI restriction enzyme, a single band was obtained, indicating that the PCR product had not been digested and that only the 'C' allele was present in the cDNA. This suggests that a regulatory mutation may be present in this family. To determine if a mutation might be present in the regulatory region of the BRCA1 gene, 1.3 kb of genomic DNA upstream of the 5' end of the gene was sequenced. No mutations were found, and hence the precise mutation in this family is not yet known.

The proband from the BOV2 family was homozygous for all polymorphisms, and therefore could not be analysed for regulatory mutations as described above. Homozygosity for all the polymorphisms suggested that this individual could have a deletion of BRCA1 on one chromosome. Using a cosmid containing the BRCA1 gene, FISH was performed on cells from the proband in this family (carried out by the ICRF Cytogenetics laboratory). The results indicated that both copies were present. Hence, no mutation has been found in this family.

4.2 Screening for common BRCA1 mutations in breast and breast/ovarian cancer families

Whilst no mutational hotspots have been found in the BRCA1 gene and most mutations are unique, two mutations are relatively common: the 185delAG mutation in exon 2 and the exon 20 5382insC mutation (Shattuck-Eidens et al., 1995). In addition, a mutation in exon 11, 2800delAA, has been reported to be relatively common in the Scottish population (Boyd et al., 1995). Since one of these mutations, 185delAG, had already been observed in one family in our collection, the remaining families were screened for these common mutations by allele specific oligonucleotide hybridisation.

Oligonucleotides were designed containing either the wild-type sequence or the mutant sequence for each mutation. These were hybridised to filters containing PCR products of the relevant exon from individuals from breast and breast/ovarian cancer families. For each mutation, 45-50 samples were screened. Neither the exon 20 nor the exon 11 mutation was detected in any samples. However, two additional individuals were found to harbour the 185delAG mutation (Figure 4.3). One family, ICRF178, contained two sisters with breast and ovarian cancer, a further sister with ovarian cancer and their mother with
Figure 4.3 Results from screening for the 185delAG mutation by ASO. An oligonucleotide containing the mutant sequence was hybridised to PCR products of BRCA1 exon 2 from individuals from breast and breast/ovarian cancer families. Positive controls are shown in lanes 20 and 49; negative controls in lanes 7 and 31. The asterisks above lanes 16 and 41 show the ICRF546 and ICRF178 families identified as carrying this mutation.
breast cancer. The other family, ICRF546, contained several cases of breast cancer but no reported cases of ovarian cancer.

The 185delAG mutation occurs at a relatively high frequency in the Ashkenazi Jewish population and individuals carrying the mutation all share the same haplotype, suggestive of a founder effect (Friedman et al., 1995b, Neuhausen et al., 1996, Simard et al., 1994). However, haplotype analysis of the BOV3, ICRF546 and ICRF178 families showed that although these three families share the same haplotype, it differs from that found in the Ashkenazi Jewish population, suggesting that these mutations may have arisen independently.

4.3 Summary of mutations identified in breast and breast/ovarian cancer families

Five germline mutations in the BRCA1 gene have been detected in our breast and breast/ovarian cancer families by direct sequencing, ASO and RT-PCR (Xu et al., 1997b) (Table 4.2). All these mutations result in a truncated protein, or loss of a BRCA1 transcript. This is in agreement with other studies, where the majority of mutations reported are nonsense or frameshift mutations (Couch et al., 1996), (see chapter one, section 1.5.2).

A regulatory mutation has been found in one family, but the precise mutation has yet to be identified. Sequencing of 1.3 kb of sequence upstream of the 5' end of the BRCA1 gene did not reveal any mutations. It is possible that the mutation could lie in another regulatory region, such as an intron. Recently, several large deletions have been reported in the BRCA1 region, one of which was found in a family with a regulatory mutation (Petrij-Bosch et al., 1997, Puget et al., 1997, Swensen et al., 1997). Affected individuals from this family have a 14 kb deletion at the 5' end of the BRCA1 gene, removing the promoter region, and hence preventing transcription of this allele. Therefore, studies are under way to examine family PRU7B5 for large genomic deletions in the BRCA1 gene that would result in an absent or unstable transcript.

No mutation was detected in the BOV2 family, but given the recent reports of large genomic deletions in breast and breast/ovarian cancer families, this family should perhaps also be analysed for deletions, as it may yet contain a BRCA1 mutation. Alternatively, the disease in this family may be due to a mutation in another predisposing gene, such as BRCA2, or an as yet unknown gene. The
Table 4.2 BRCA1 mutations detected in ICRF families.

<table>
<thead>
<tr>
<th>Family ID</th>
<th>Mutation</th>
<th>Technique by which mutation was detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC3</td>
<td>3896delT</td>
<td>sequencing</td>
</tr>
<tr>
<td>BOV3</td>
<td>185delAG</td>
<td>sequencing</td>
</tr>
<tr>
<td>ICRF178</td>
<td>185delAG</td>
<td>ASO</td>
</tr>
<tr>
<td>ICRF546</td>
<td>185delAG</td>
<td>ASO</td>
</tr>
<tr>
<td>PRU7B5</td>
<td>loss of transcript</td>
<td>RT-PCR</td>
</tr>
</tbody>
</table>
family is small and has only two affected sisters, one with breast cancer and one with breast and ovarian cancer. Therefore, it is possible that this is a chance clustering of sporadic cancers and not due to an inherited susceptibility.
CHAPTER FIVE: ISOLATION OF THE MURINE HOMOLOGUES OF THE 
BRCA1 AND NBR1 GENES

The NBR1 gene was initially identified as a candidate for the tumour antigen CA125, used in the management of ovarian cancer (Campbell et al., 1994b). It was subsequently found to map to the BRCA1 region on chromosome 17q and became an attractive candidate for the BRCA1 gene, since families predicted to carry a BRCA1 mutation often have an increased risk of ovarian cancer (Ford et al., 1995). However, no mutations in the NBR1 gene were found in breast or breast/ovarian cancer families (Campbell et al., 1994b). Following the isolation of the BRCA1 gene, initial physical mapping of the region suggested that the NBR1 gene lay adjacent to the BRCA1 gene (Brown et al., 1994). Refinement of this map indicated that this region is in fact duplicated and that the NBR1 gene is adjacent to pseudo-copies of the 5' exons of the BRCA1 gene (Brown et al., 1996). This chapter and chapter six describe the conservation of these genes and the duplicated region in other species.

5.1 Conservation of the BRCA1 and NBR1 genes in other species

Whilst valuable information about transformation can be gained from in vitro studies, human cancer occurs in the context of a complex interaction with its environment - that is, neighbouring tissues, the immune system, hormones and dietary factors. Isolation of mouse homologues facilitates the generation of an animal model in which to study these interactions. In addition, conservation of predicted proteins between two species can indicate potentially significant domains, and hence clues to the functions of the proteins.

Previous results (see chapter three) have shown that many genes within the human BRCA1 region are conserved in other species. The conservation of the BRCA1 and NBR1 genes was therefore also studied. Hybridisation of cDNA clones from these genes to zoo blots revealed that both genes are conserved in a range of species, including mouse and chick (Figure 5.1). To study further the conservation of these genes and to provide an animal model of breast tumorigenesis, the murine homologues of the BRCA1 and NBR1 genes were isolated.
Figure 5.1 Conservation of the *BRCA1* and *NBR1* genes. The left panel shows the results from hybridisation of a human *BRCA1* cDNA clone of exons 11-16 to a zoo blot. The right panel shows hybridisation of a human *NBR1* PCR product of exons 1B and 2 to a zoo blot.
5.2 Isolation of the murine homologue of the NBR1 gene

For technical reasons, such as availability of cDNA clones, initial studies focused on the NBR1 gene.

5.2.1 Isolation of murine Nbr1 cDNA clones

The full-length human NBR1 cDNA clone was used to screen a mouse testis cDNA library (2 x 10⁶ clones), at low stringency, and fifteen positive clones were obtained. PCR and sequencing analysis using vector primers indicated that these clones fell into two categories: those of 500 bp, homologous to human NBR1 exons 16-18 and those of 950 bp, homologous to exons 3-9 of human NBR1. Since the inserts of these cDNA clones were relatively small and full coverage of the mouse Nbr1 cDNA was not obtained, a mouse brain cDNA library (2 x 10⁶ clones) was screened. A further fifteen positive clones were obtained. Each of these clones was grown for analysis and the ends sequenced using vector primers. Two clones were non-specific, with homology to other proteins. The remaining thirteen clones all shared homology with the human NBR1 gene. The cDNA clones ranged in size from 1 kb to 2.8 kb and between them covered exon 2 to the 3’ untranslated region, as compared to the human NBR1 cDNA sequence. This region was entirely contained within two single clones, mb10 and mb3, of 1.8 kb and 2.7 kb respectively (Figure 5.2). Two clones, mb1 and mb12, may be alternatively spliced transcripts, since they lack the equivalent of human exon 14. Comparison with the human sequence revealed that neither the ATG for translation initiation, nor a polyadenylation signal, was present in the mb10 mb3 sequence, suggesting that the full-length mouse cDNA had not been obtained.

In order to obtain the 3’ end of the cDNA, a 200 bp PCR product, UTF3-UTR1, was generated from the 3’ untranslated region present in mb3 (Figure 5.2). This product was used as a probe to screen the mouse brain cDNA library (4 x 10⁶ clones) and four positive clones were obtained. One clone, mb25, extended the 3’ untranslated sequence by 25 bp and contained a polyadenylation signal, AAATAA.

Since few 5’ cDNA clones had been obtained during the initial screen of the mouse brain cDNA library and difficulties can be encountered in obtaining 5’ sequences from a library generated by priming with oligo dT, the 5’ end of the
Figure 5.2 Schematic diagram of the mouse Nbr1 cDNA illustrating the coverage of the isolated cDNA clones (adapted from Chambers and Solomon, 1996). The Nbr1 cDNA is illustrated by a thick horizontal line. Vertical lines indicate the exon/intron boundaries and the exons are numbered. R3 denotes the clone obtained from RACE; mb denotes clones obtained from the mouse brain cDNA library. Clones mb12 and mb1 represent alternatively spliced products that do not contain exon 14. The position of the PCR product UTF3-UTR1 is indicated.
mouse Nbr1 gene was sought by RACE. Using mouse brain poly A+ RNA and a primer from Nbr1 exon 3, cDNA was synthesised. A primer was designed in exon 2 of the Nbr1 cDNA, to amplify the 5' end. The resulting products were cloned into the pDIRECT vector (Clontech) and colonies analysed by PCR and sequencing. Three types of clones of differing sizes were obtained: R1 (a 348 bp product), R2 (a 359 bp product) and R3 (a 656 bp product), giving 264 bp, 275 bp and 571 bp, respectively, of new 5' sequence. The ATG for translation initiation (as compared with the human sequence) was present in all these clones. Sequence comparison of the RACE clones with a mouse cosmid clone from the region (see following section) and hybridisation studies, suggest that R3 may be chimaeric and that 278 bp at the extreme 5' end of the clone may not be derived from this region. R1 and R2 were contained within the 3' sequence of R3. If the non-chimaeric sequence from R3 is a true 5' cDNA clone, then the full-length cDNA isolated is 4444 bp (Appendix B).

Northern blot analysis, using mb25 as a probe, identified a transcript of approximately 4.5 kb in all tissues examined, in accordance with the size of the isolated cDNAs (Figure 5.3). An additional transcript of 4 kb was detected in testis. The nature of this transcript is not known, although it may be a product of alternative splicing. Several genes have altered transcript sizes in testis, many of which are non-functional due to the absence or truncation of the untranslated regions involved in regulating translation or mRNA stability (reviewed in Ivell, 1992).

5.2.2 Isolation of murine Nbr1 genomic clones

In order to analyse the gene structure of the mouse Nbr1 gene, mouse genomic clones were isolated. Cosmid and bacteriophage lambda libraries made from 129Sv mice were available (donated by Drs. A-M. Frischauf, F. Otto and M. Owen). The cosmid library was selected since cosmids may contain inserts of 40 kb, whereas bacteriophage have inserts of about 20 kb, and hence fewer cosmids should be required to obtain the entire Nbr1 gene.

The mouse cosmid library had been prepared as six fractions (1-6). Fraction 7 was a pool of fractions 1-6 and this pool was plated for screening. A PCR product amplified from exons 1B and 2 of the human NBR1 cDNA had shown good conservation when used to probe a zoo blot (Figure 5.1) and so this PCR product was used to screen fraction 7 of the cosmid library (5 x 10^5 clones). Three positive
Figure 5.3 Hybridisation of mouse *Nbr1* to a Northern blot. A 1.2 kb *EcoRI* fragment from cDNA clone mb25 was hybridised to a mouse multiple tissue Northern blot.
clones were identified which, by restriction enzyme digests, appeared to be the same. Southern blots were prepared from digested DNA of this clone, MCHCA1, and hybridised to a human PCR product of exons 1B and 2 and oligonucleotides from exons 1, 4 and 5 of the mouse Nbr1 cDNA clones. Specific hybridisation was observed with all probes except the exon 5 oligonucleotide, suggesting that this cosmid contains the 5' exons but may not extend beyond intron 4 of the Nbr1 gene.

In order to isolate 3' clones, fraction 7 (5 x 10^5 clones) was screened with a mouse testis cDNA clone, mt4, which covers exons 3-9, but no positive clones were identified. To determine appropriate fractions to plate and screen, LB medium was inoculated with each of the fractions and these cultures grown for extraction of DNA. Southern blots were prepared from 10 µg digested DNA from each of these fractions and mt4 was used as a probe. Hybridisation was observed to fraction 4 and this fraction was therefore plated and screened with mt4. However, no positive clones were detected.

In an attempt to isolate further cosmids, primers from the murine 3' untranslated region of Nbr1 were used to amplify each of the cosmid fractions. These primers had previously been used to amplify a 200 bp product, UTF3-UTR1, from the mouse cDNA clone, mb25 (see section 5.2.1, Figure 5.2). A 200 bp fragment was successfully amplified from fractions 1, 2, 4 and 6, and so fraction 1 was plated and screened with the mouse mb25 3' cDNA clone, but no positive clones were detected. A 1 kb EcoRI fragment from mb10, covering exons 2-9, was then hybridised to fraction 1 and a single positive clone, cos10-1, was identified. However, this clone proved to be highly unstable: upon culturing large proportions of the insert DNA were lost. Sufficient culturing to obtain enough DNA for thorough analysis was not possible, although adequate quantities for PCR could be obtained after only a short incubation of the culture. By designing primers from the isolated mouse cDNA clones, PCR reactions were carried out which indicated that exons 5-18 are present on this cosmid. Transformation of the cosmid into a different strain of E. Coli (XL1 Blue MRF') did not improve the stability of the clone.

Given the problems encountered when using the cosmid library, the bacteriophage lambda library (2 x 10^6 clones) was screened with a 1 kb EcoRI fragment from mb10. Two positive clones, λ2001-1 and λ200-3, were identified.
Hybridisations with PCR products amplified from the cosmids revealed that \( \lambda 2001-3 \) contains exons 3-12 and \( \lambda 2001-1 \) contains exons 6-18 (Figure 5.4).

To identify clones containing the full 3' end of the gene, the lambda library was screened with the UTF3-UTR1 PCR product (Figure 5.2). Two positive clones were identified, \( \lambda 2001-7 \) and \( \lambda 2001-8 \). PCR analysis showed that, like \( \lambda 2001-1 \), these clones contain exons 17 and 18 (Figure 5.4). However, PCR reactions with additional primers from exon 18 and the 3' untranslated region, including primers UTF3 and UTR1, all failed. This suggested that these sequences were not present on the lambda clones. Previous PCR analysis had shown that the UTF3-UTR1 PCR product was 200 bp in genomic DNA, and therefore the failure of this PCR product could not be due to a large intron. Using this product as a probe, hybridisation to fragments containing the lambda arms of \( \lambda 2001-7 \) and \( \lambda 2001-8 \) was detected, but it was not clear whether or not this was specific binding. Further analysis was carried out by sequencing and revealed that an additional intron was present within exon 18. Using a primer from this intron, sequence that matched the vector polylinker was obtained from \( \lambda 2001-1 \) and \( \lambda 2001-7 \). Clearly, these lambda clones do not contain the 3' untranslated region of the \( Nbr1 \) gene and the hybridisation results obtained with UTF3-UTR1 must therefore reflect non-specific binding. Sequencing of \( \lambda 2001-8 \) with the same intron 18a primer revealed that this clone extends further, since no match with the vector sequence was observed. The absence of 3' untranslated region sequence on \( \lambda 2001-7 \) was surprising, since a PCR product from the 3' untranslated region had been used to isolate these clones. However, this PCR product was amplified from a cDNA clone and it is possible that sufficient contaminating DNA from the plasmid was present to hybridise to these clones that contained no 3' untranslated region. Therefore, although the precise extent of 3' sequence on \( \lambda 2001-8 \) has not been determined, together the isolated cosmid and lambda clones cover at least exons 1b-18a of the \( Nbr1 \) gene (Figure 5.4).

The genomic clones were used to study the exon/intron structure of the \( Nbr1 \) gene. The authenticity of the previously isolated 5' RACE clones could be analysed by comparison with cosmid clone MCHCA1. PCR and sequencing analysis of the 348 bp sequence common to all three RACE clones showed that this sequence is derived from genomic sequence present on MCHCA1. Sequences present in the cosmid and not the RACE clones indicate that three introns exist before exon 2. Sequence of a further 29 bp present in RACE clone R3, upstream of this region, matches the cosmid and subsequently diverges from it. To
Figure 5.4 Schematic diagram illustrating the genomic sequence covered by Nbr1 cosmid and bacteriophage lambda clones. Nbr1 exons are denoted by rectangles. The exact locations of exons beyond exon 18a have not been identified. E = EcoRI restriction enzyme sites, indicated by vertical lines. Cosmid and lambda clones are represented by horizontal lines and the vector ends, where known, are indicated by hatched boxes.
investigate whether or not this divergence was due to an intron, PCR primers were designed and the non-homologous region amplified, to use as a probe on Southern blots of MCHCA1 DNA and an overlapping lambda clone, BL1 (see section 5.3.2). No hybridisation was observed, indicating that this sequence is not present within approximately 18 kb of Nbr1 exon 1b and therefore, it may not be part of the Nbr1 gene. The most 5' exon has 85% homology to human NBR1 exon 1B and so the mouse exon was called exon 1b. The two intervening exons between exons 1b and 2 in the mouse sequence have been designated 1c and 1d (see Figure 5.6).

To study the exon/intron structure over the remainder of the gene, primers were designed throughout the cDNA (Appendix C). These primers were used to amplify adjacent exons from the cosmid or lambda bacteriophage DNA, to determine the size of the intervening intron. Each PCR product was then sequenced in both directions, to determine the precise location of the exon/intron boundary. Additional sequencing primers in the exons or introns were designed where necessary. In this manner the structure of the mouse Nbr1 gene was determined as far as exon 18 (Table 5.1).

5.2.3 Comparison of the human and mouse NBR1 genes

The human and mouse NBR1 genes are very similar, encoding proteins of 966 and 988 amino acids, respectively. Northern blot analysis showed that a 4.5 kb NBR1 mRNA is widely expressed in human and mouse tissues (Figure 5.3). Alternative splicing of murine exon 14 was observed in brain cDNA clones: five clones possessed the sequence, while two lacked it. The predicted protein sequences are 89% similar and 84% identical (Figure 5.5). High conservation was observed throughout the sequence and suggests that the entire protein is functionally significant. The only recognisable domain within the protein, a B-box domain, was also highly homologous, being 92% similar between the two species.

The gene structure is largely the same in the two species. Two additional introns have been observed within exons 4 and 18 in the mouse sequence. An extra 75 bp with no homology to the human sequence were observed at the end of murine exon 15. PCR and sequencing analysis revealed that this sequence is an additional exon, present in all six clones covering this region. This may be a new exon that is not present in the human, or it may be alternatively spliced. In
<table>
<thead>
<tr>
<th>Exon</th>
<th>Position of first nucleotide of exon</th>
<th>Size of exon (bp)</th>
<th>3' splice junction</th>
<th>5' splice junction</th>
<th>intron size (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>1</td>
<td>97</td>
<td>TG/gtaagg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>1c</td>
<td>98</td>
<td>111</td>
<td>aag/A</td>
<td>AG/gtaaat</td>
<td>1.6</td>
</tr>
<tr>
<td>1d</td>
<td>210</td>
<td>66</td>
<td>tag/G</td>
<td>AG/gtgtgt</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>276</td>
<td>117</td>
<td>cag/C</td>
<td>TG/gtaagt</td>
<td>3.0</td>
</tr>
<tr>
<td>3</td>
<td>393</td>
<td>63</td>
<td>cag/G</td>
<td>AG/gtaag</td>
<td>0.9</td>
</tr>
<tr>
<td>4a</td>
<td>456</td>
<td>19</td>
<td>tag/A</td>
<td>AG/gtaagt</td>
<td>1.5</td>
</tr>
<tr>
<td>4b</td>
<td>475</td>
<td>23</td>
<td>tag/G</td>
<td>AG/gtaagt</td>
<td>1.0</td>
</tr>
<tr>
<td>5</td>
<td>498</td>
<td>198</td>
<td>tag/A</td>
<td>AG/gtatga</td>
<td>0.8</td>
</tr>
<tr>
<td>6</td>
<td>696</td>
<td>78</td>
<td>cag/C</td>
<td>TG/gtgagt</td>
<td>0.4</td>
</tr>
<tr>
<td>7</td>
<td>774</td>
<td>215</td>
<td>tag/T</td>
<td>AG/gtaagc</td>
<td>0.7</td>
</tr>
<tr>
<td>8</td>
<td>989</td>
<td>168</td>
<td>cag/C</td>
<td>AG/gtaagag</td>
<td>1.1</td>
</tr>
<tr>
<td>9</td>
<td>1157</td>
<td>210</td>
<td>cag/G</td>
<td>AT/gtaagt</td>
<td>0.6</td>
</tr>
<tr>
<td>10</td>
<td>1367</td>
<td>160</td>
<td>cag/G</td>
<td>AG/gtacgt</td>
<td>0.1</td>
</tr>
<tr>
<td>11</td>
<td>1527</td>
<td>291</td>
<td>cag/C</td>
<td>AG/gtgagc</td>
<td>1.3</td>
</tr>
<tr>
<td>12</td>
<td>1818</td>
<td>150</td>
<td>tag/G</td>
<td>AG/gtaagc</td>
<td>0.5</td>
</tr>
<tr>
<td>13</td>
<td>1968</td>
<td>76</td>
<td>tag/G</td>
<td>GG/gtaaga</td>
<td>0.5</td>
</tr>
<tr>
<td>14</td>
<td>2044</td>
<td>111</td>
<td>cag/A</td>
<td>TG/gtaag</td>
<td>0.2</td>
</tr>
<tr>
<td>15a</td>
<td>2155</td>
<td>165</td>
<td>cag/A</td>
<td>GA/gtgagt</td>
<td>0.1</td>
</tr>
<tr>
<td>15b</td>
<td>2320</td>
<td>75</td>
<td>cag/G</td>
<td>CT/gtaagt</td>
<td>1.7</td>
</tr>
<tr>
<td>16</td>
<td>2395</td>
<td>423</td>
<td>tag/T</td>
<td>AG/gtacgt</td>
<td>0.9</td>
</tr>
<tr>
<td>17</td>
<td>2813</td>
<td>107</td>
<td>aag/G</td>
<td>AG/gtaxgc</td>
<td>0.8</td>
</tr>
<tr>
<td>18a</td>
<td>2920</td>
<td>57</td>
<td>tag/A</td>
<td>AG/gtaacg</td>
<td>nd&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>18b</td>
<td>2977</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>19</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
</tbody>
</table>

<sup>a</sup> Exonic sequence is shown in capitals, intronic sequence in lowercase.

<sup>b</sup> nd = not determined

Table 5.1 Exon/intron structure of the mouse *Nbr1* gene.
Figure 5.5 Comparison of the human and mouse NBR1 predicted amino acid sequences. The human sequence is shown on the top line, the mouse on the bottom. The B-box is double-underlined.
contrast, the human sequence contains an extra 27 bp, in the middle of exon 16, which are not found in the mouse sequence.

The major difference between the mouse and human NBR1 genes was observed at the 5' end. In humans alternative exons, 1A and 1B, are present, either of which may be spliced to exon 2, containing the translation initiation codon. In the mouse an exon 85% homologous to human exon 1B was isolated by RACE and a further two exons, 1c and 1d, are present between exons 1b and 2 (Figure 5.6). Exons 1c and 1d have no homology to the human cDNA sequence, nor to 2.8 kb of genomic sequence downstream of human exon 1B. No homologue of human exon 1A was isolated by RACE or by screening the cDNA library, although there is some evidence to suggest that this exon may exist in the mouse (see chapter six).

5.3 Isolation of the murine homologue of the BRCA1 gene

5.3.1 Isolation of murine Brca1 cDNA clones

A human cDNA clone containing exons 11-16 was used to screen a mouse brain cDNA library (4 x 10^6 clones), but no positive clones were obtained. This could reflect poor homology between the human and mouse sequences over this region. Therefore, to obtain a more suitable probe for isolating the murine homologue of BRCA1, primers were designed from partial sequence of mouse Brca1 exon 11 that had become available (accession number: U27542), (Marquis et al., 1995). These primers were used to amplify a 814 bp fragment from mouse total genomic DNA. Since human BRCA1 is expressed at relatively high levels in testis, this PCR product was hybridised to a mouse teratocarcinoma cDNA library (4 x 10^6 clones). Twelve positive clones were obtained. These clones were fully characterised by BL Griffiths in our laboratory. By comparison with the human and the recently published mouse Brca1 mRNA sequence (Bennett et al., 1995) (accession number: U32446), two clones, MT1 and MT8, together cover nucleotides 26-4476, exons 1a to 15. To obtain the 3' end of the cDNA, the mouse teratocarcinoma cDNA library was screened with a 476 bp 3' EcoRI/HindIII fragment from MT8. Fifteen positive clones were obtained, one of which, MT25, covers nucleotides 3098-6055 of the mouse sequence, including 500 bp of the 3' untranslated region.
Figure 5.6 Comparison of the 5' ends of the human and mouse \textit{NBR1} genes (adapted from Chambers and Solomon, 1996). a) Illustration of the 5' end of the human \textit{NBR1} gene, showing the alternative transcription start sites. b) Illustration of the 5' end of the murine \textit{Nbr1} gene. Boxes represent exons, asterisks denote the translation initiation sites. Transcripts are represented by thick horizontal lines.
5.3.2 Isolation of genomic clones of the murine \textit{Brca1} gene

Genomic clones of the murine \textit{Brca1} gene were also isolated. In the first instance, PCR products from the human \textit{BRCA1} gene were hybridised to cosmid MCHCA1 (previously isolated by a human \textit{NBR1} PCR product). Specific signal was detected with a product containing exon 2 of the human \textit{BRCA1} gene. This suggested that the 5' ends of both the \textit{Nbr1} and the \textit{Brca1} genes might lie on this cosmid. PCR products 3' of exon 2 did not hybridise to MCHCA1.

Exon 2 hybridised to a 0.6 kb \textit{EcoRI/BamHI} fragment (Figure 5.7). To extend the contig, this fragment was gel-purified and used to screen the lambda bacteriophage library (2 x 10^6 clones). One positive clone, BL1, was obtained. PCR and sequencing analysis of BL1 indicated that exons 2-6 of \textit{Brca1} are present on this clone, but not exons 1 or 7 (Figure 5.7). A PCR product from exons 5 and 6 hybridised to \textit{EcoRI} fragments of 1.2 and 1.9 kb in size. In order to prioritise these probes, the fragments were tested as probes on blots containing genomic DNA. The 1.2 kb fragment was not repetitive and was selected for screening the lambda library (2 x 10^6 clones). Four positive clones were identified. PCR analysis with primers from exons 7 and 8 indicated that two clones did not contain these sequences and they were discarded. The two remaining clones were analysed by PCR with primers from exon 11; a fragment was amplified from one clone, BL5/6-4. These results were confirmed by hybridisation. In addition, PCR and sequencing indicated that exon 12 is present on BL5/6-4. Southern blot analysis suggested that exons 13 and 14 are not contained within this clone (Figure 5.7).

A 1.4 kb \textit{BamHI} fragment was detected when lambda blots were hybridised to a PCR product from exons 11 and 12 (Figure 5.7). This fragment was not repetitive and was used to isolate five positive clones from the bacteriophage lambda library (2 x 10^6 clones). Hybridisation of a PCR product from exons 13 and 14 was observed to three clones: BL11-1A, BL11-1B and BL11-2A. Hybridisation on BL11-1B was to a fragment containing the bacteriophage arm, suggesting that this clone may not extend as far as the other two isolated clones, BL11-1A and BL11-2A. Hence, in total, more than 35 kb of genomic DNA encompassing the \textit{Brca1} gene has been cloned (Figure 5.7).
Figure 5.7 Schematic diagram illustrating the genomic sequence covered by Brca1 cosmid and bacteriophage lambda clones. Approximate positions of the Brca1 exons are denoted by rectangles. E = EcoRI and B = BamHI restriction enzyme sites, indicated by vertical lines; not all BamHI sites are illustrated. Cosmid and lambda clones are represented by horizontal lines and the vector ends indicated by hatched boxes.
5.3.3 Comparison of the human and mouse BRCA1 genes

The structure of the human and mouse BRCA1 genes appears to be very similar. The Brca1 cosmid and lambda genomic clones were used to partially characterise the exon/intron structure of exons 1a to 8 and exon 11, by PCR and sequencing analyses; no differences between the human and mouse exon/intron boundaries were detected.

During the course of this work, the murine Brca1 cDNA sequence was published by several other groups (Abel et al., 1995, Bennett et al., 1995, Sharan et al., 1995). This sequence was used to confirm the validity of our current and subsequently isolated cDNA clones. From analysis of the isolated cDNA clones, the predicted amino acid sequences of the BRCA1 human and mouse proteins are 58% identical (Figure 5.8) (Abel et al., 1995, Bennett et al., 1995, Sharan et al., 1995). In addition, two regions of the protein are particularly highly conserved, indicative of functional significance. The RING finger is identical between the two species and a domain towards the carboxy terminus of the proteins is 91% similar between the two species (Figure 5.8). There is now accumulating evidence to suggest that both these domains may be critical for the normal function of the BRCA1 protein; the RING finger interacts with a novel protein, BARD1 (Wu et al., 1996) and the conserved C-terminal domain, the BRCT domain, has transactivating properties (Chapman and Verma, 1996, Monteiro et al., 1996).

5.4 Summary of the conservation observed between the human and mouse BRCA1 and NBR1 genes

Both cDNA and genomic clones of the murine homologues of the human NBR1 and BRCA1 genes have been isolated. The human and mouse NBR1 predicted proteins are similar in size and highly conserved (87% similar) (Chambers and Solomon, 1996). Like the human gene, murine Nbr1 is expressed in a wide range of tissues. Northern blot analysis reveals a transcript of 4.5 kb in all tissues examined.

Comparison of the NBR1 gene structure in human and mouse identified three additional exons and five additional introns in the mouse gene. The 5' end of the gene was the most divergent region between the species. In humans two alternate first exons, 1A and 1B, are present, either of which may be spliced to exon 2. In the mouse, only a homologue of exon 1B was cloned and two
Figure 5.8 Comparison of the human and mouse BRCA1 predicted amino acid sequences. The human sequence is shown on the top line, and the mouse sequence on the bottom. The RING finger domain and the C-terminal conserved domain are shown in boxes.
intervening exons, 1c and 1d, are present upstream of exon 2. Two additional upstream translation initiation codons are present within exons 1c and 1d. If used, these would code for peptides of 56 and 12 amino acids, respectively. However, neither is present in a favourable context compared with the Kozak consensus sequence, GCCA/GCCATGG (Kozak, 1987) and they are therefore unlikely to be used. Hence, the translation initiation site is likely to be the same in both species, despite the difference in length of the 5' untranslated region.

The structure of the human and mouse BRCA1 genes is also very similar. Partial characterisation of the gene structure revealed few differences between the two species. However, the human gene has two alternate 5' exons, exons 1A and 1B, but only exon 1a appears to be present in the mouse (see chapter six).

From the sequencing analysis of the cDNA clones described in this chapter, the predicted proteins in the human and mouse can be compared. High conservation throughout the NBR1 protein is observed in the human and mouse, although this has not identified any functionally significant domains. However, the BRCA1 predicted amino acid sequence, whilst well conserved between human and mouse (58% homologous), contains two highly conserved, and hence potentially important regions: the RING finger and a region near the carboxy terminus. RING fingers are thought to be involved in protein-protein interactions and indeed, a novel protein, BARD1, has been found to interact with BRCA1 via the RING finger (Wu et al., 1996). Whilst the function of BARD1 is not yet known, it is co-expressed with BRCA1, and therefore may be crucial for the normal function of BRCA1. The C-terminal domain, BRCT, has been identified as a domain found in several proteins, including those involved in cell cycle regulation and DNA repair (Callebaut and Mornon, 1997, Koonin et al., 1996). In addition, the BRCA1 BRCT domain is capable of activating transcription of a reporter gene and may provide a role for BRCA1 in transcriptional control (Chapman and Verma, 1996, Monteiro et al., 1996).

Hence, a full-length cDNA of the murine Nbr1 gene, and cDNA clones covering the entire coding region of the murine Brca1 gene, have been isolated. Genomic clones from the murine Brca1/Nbr1 region, covering approximately 60 kb, have also been isolated. Cloning of the murine homologues of these genes has identified important domains and will provide resources for using the mouse as an animal model to study the functions of these genes.
CHAPTER SIX: ANALYSIS OF THE PROMOTER REGIONS OF THE MURINE Brca1 AND Nbr1 GENES

Gene transcription is regulated by the interaction of proteins with regulatory elements located upstream of a gene. Analysis of this promoter region of a gene can identify potentially important regulatory regions and the factors responsible for controlling gene expression. These factors may regulate the expression level of the gene under different conditions and in response to certain stimuli. Comparison of the promoter regions in human and mouse may indicate similarities or differences in gene regulation between the two species. This is particularly important if the mouse is to be used as a model of the human disease. In order to begin to understand the regulation of the Brca1 and Nbr1 genes, their promoter regions have been analysed.

6.1 Localisation of the murine Nbr1 gene adjacent to the Brca1 gene

The human BRCA1 and NBR1 genes are located in a duplicated region such that the NBR1 gene lies adjacent to pseudo-copies of 5' exons of the BRCA1 gene and the BRCA1 gene lies adjacent to the NBR2 gene, which shares homology with the 5' exons of the NBR1 gene (see chapter one, section 1.5.5). Hence, the human BRCA1 and NBR1 genes have distinct promoters, located more than 40 kb apart. To determine if the duplication is present in other species, this region was examined in detail in the mouse.

Results described in chapter five indicate that 5' exons from both the Brca1 and the Nbr1 genes lie on the same cosmid, MCHCA1. The region was further analysed by the hybridisation of exon-specific probes, which suggested that exons 1a and 2 of the Brca1 gene and exons 1b, 1c, 1d and 2 of the Nbr1 gene lie on the same 14.5 kb PstI fragment (Figure 6.1). To determine the precise distance between these genes and their promoter regions, PCR primers were designed from exon 1a of the Brca1 gene and exon 1b of the Nbr1 gene and a product of approximately 1 kb was obtained from cosmid MCHCA1 DNA. Sequencing analysis confirmed the validity of this PCR product and revealed that the distance between exon 1a of Brca1 and exon 1b of Nbr1 is only 991 bp.

In humans, this region is duplicated and pseudo-copies of exons 1A, 1B and 3 of the NBR1 gene and exons 1A, 1B and 2 of the BRCA1 gene are present. The
Figure 6.1 Partial map of cosmid MCHCA1, indicating the relative positions of the Brca1 and Nbr1 genes. The approximate positions of the exons are indicated by rectangles. E = EcoRI restriction enzyme site, B = BamHI restriction enzyme site and P = PstI restriction enzyme site. The 14.5 kb PstI fragment and the 5.5 kb EcoRI/PstI fragment are shown.
results from mouse cosmid MCHCA1 suggested that the duplication may not be present in the mouse. To confirm this, 5' exons of the Brca1 and Nbr1 genes were sequenced from the cosmid and compared with the cDNA sequences (accession numbers: U32446 and U73039). Sequences of genes and pseudogenes are rarely identical and the human BRCA1 and NBR1 real and duplicated exons range between 90-94% identity (Brown et al., 1996). However, the sequences of the exons from the mouse cosmid and cDNA clones were identical, indicating that the exons present on the cosmid were unlikely to be pseudo-copies. In addition, sequences of these exons on overlapping lambda clones also matched the sequence of these exons on the cosmid and cDNA clones. Hybridisation of a PCR product from exons 1B and 2 of the human NBR1 gene was observed to a single EcoRI fragment in mouse genomic DNA and to a single EcoRI fragment of the same size in cosmid MCHCA1 (see chapter five, Figure 5.1). The same probe hybridises to two fragments in human genomic DNA (see chapter five, Figure 5.1), corresponding to the real and pseudo-copies of these NBR1 exons. Furthermore, the NBR2 gene, which is adjacent to the human BRCA1 gene and contains pseudo-copies of NBR1 5' exons, is not found in this region of the mouse and hybridisations to zoo blots suggest that the NBR2 gene may only be present in primates (Xu et al., 1997c). Hence no duplication of this region is present in the mouse, and unlike the human, the Brca1 and Nbr1 genes lie less than 1 kb apart (Figure 6.2).

6.2 Analysis of the intergenic region between the Brca1 and Nbr1 genes

The discovery of an intergenic region of less than 1 kb suggested that the Brca1 and Nbr1 genes could have overlapping promoters, or could share a bidirectional promoter. In addition, these two genes could overlap if further 5' exons are present. In the human, both the BRCA1 and NBR1 genes have alternate first exons, 1A and 1B (Brown et al., 1994, Xu et al., 1995). The BRCA1 alternate exons are controlled by separate promoters, one of which is bidirectional with the NBR2 gene (Xu et al., 1997c). Homologues of exon 1A of BRCA1 and exon 1B of NBR1 have been isolated from the mouse (Bennett et al., 1995, Chambers and Solomon, 1996), but no homologues of BRCA1 exon 1B or NBR1 exon 1A have been identified.

To ascertain whether either of these exons is expressed in the mouse, to determine if the Brca1 and Nbr1 genes are closer than 991 bp, and to see if multiple promoters are present, the region was examined further in the mouse.
Figure 6.2 Comparison of the BRCA1 genomic region in the human (top panel) and the mouse (bottom panel). The first three exons of BRCA1 and exons 1A, 1B and 3 of NBR1 are duplicated in the human. The first two exons of human NBR2 are homologous to exons 1A and 3 of human NBR1. The region is not duplicated in the mouse and no homologue of the NBR2 gene is present in this region. The boxes represent exons and the arrows indicate the direction of transcription. This diagram is not drawn to scale.
Sequence comparison of the mouse sequence up to 1 kb downstream of Brca1 exon 1a revealed no homology to human BRCA1 exon 1B. This sequence would be predicted to have homology to human exon 1B if it is present in the mouse, since exons 1A and 1B are separated by an intron of only 158 bp in the human. The absence of an exon 1B homologue in the mouse is not surprising, since the human exon 1B sequence contains an Alu repeat element, a repetitive sequence found only in primates.

The intervening sequence between the murine Brca1 and Nbr1 genes was compared with human NBR1 exon 1A. A region of 484 bp of the mouse sequence has 76% identity with human NBR1 exon 1A, suggesting that a homologue of exon 1A may be expressed in the mouse. However, no cDNA clones containing this exon have been isolated either by screening cDNA libraries, or by RACE (see chapter five). Therefore, to determine if the sequence homologous to human NBR1 exon 1A is expressed, a primer was designed within this sequence, to use for RT-PCR. Using this primer and a primer from exon 2, a fragment of approximately 630 bp was amplified from mouse testis and thymus RNA (Figure 6.3). To determine if bands were amplified from other tissues, RT-PCR products were analysed by Southern hybridisation, using a PCR product from exons 1d and 2 as a probe. This revealed that the 630 bp fragment was also present in liver and bladder, but no signal was detected in heart, salivary gland, muscle, kidney, uterus or ovary. This was in contrast to the expression of the exon 1b transcript, which was detected in most of the tissues examined (Figure 6.3).

Sequencing analysis of the 630 bp RT-PCR product from testis revealed that 275 bp corresponds to exons 1c, 1d and 2 of Nbr1. The remaining 355 bp has homology to exon 1A of the human NBR1 gene. This strongly suggests that a homologue of exon 1A is expressed in the mouse. This places the Brca1 and Nbr1 genes less than 328 bp apart, depending on the precise location of the 5' end of Nbr1 exon 1a, which has not yet been determined. These results suggest that Nbr1 gene expression may be controlled by two promoters, one of which could be shared with the Brca1 gene.

A 5.5 kb EcoRI/PstI fragment, encompassing Brca1 exon 1a to Nbr1 exon 2, was subcloned into pBluescript (Stratagene) and sequenced in its entirety. This sequence was compared with the transcription factor sites database, to identify potential factors involved in the regulation of the Brca1 and Nbr1 genes. Several CCAAT boxes, Sp1 binding sites, AP-1 and AP-2 sites are present (Figure 6.4). No
Figure 6.3 Analysis of the expression of Nbr1 exon 1a and exon 1b transcripts by RT-PCR. RT-PCR products were separated on agarose gels, blotted onto membrane and hybridised to a PCR product from exons 1d and 2 of Nbr1. The top panel shows the results from an RT-PCR reaction using primers from exon 1a and exon 2; the bottom panel indicates results from using primers from exons 1b and 2.
Figure 6.4 Diagram illustrating the sequence between the Brca1 and Nbr1 genes. Brca1 exon 1a is shown as a hatched pink box; Nbr1 exons are shown as hatched yellow boxes. The unhashed yellow box indicates the sequence, upstream of the RT-PCR product, with homology to human NBR1 exon 1A. Transcription factor binding sites are illustrated by dashed horizontal lines. Sites conserved with the human are shown as solid black lines. Green horizontal lines indicate potential initiator elements.
TATA boxes, which direct accurate transcription initiation, were observed. Many TATA-less genes are now known, including housekeeping genes, oncogenes, growth factors and transcription factors. The promoters of these genes contain an initiator sequence which is sufficient to direct accurate transcription initiation (reviewed in Azizkhan et al., 1993, Smale, 1997). A loose consensus for the initiator sequence has been described: PyPyA+1NT/APyPy (Javahery et al., 1994). Analysis of the sequence of the promoter regions of the mouse Brca1 and Nbr1 genes reveals potential initiators for exons 1a and 1b of the Nbr1 gene (CTAGTCT and TTAGATT, respectively) (Figure 6.4). A less well conserved initiator is also present upstream of Brca1 exon 1a (GTAATTG) (Figure 6.4).

6.3 Analysis of the activities of the murine Brca1 and Nbr1 gene promoters using a luciferase reporter gene assay system

The close proximity of the Brca1 and Nbr1 genes suggests that a bidirectional promoter may regulate expression of both genes. Many adjacent genes have been shown to be under the regulation of a single bidirectional promoter: for example, the Surf-1 and Surf-2 genes (Lennard et al., 1994, Lennard and Fried, 1991), the DHFR and hMSH3 genes (Crouse et al., 1985, Shinya and Shimada, 1994) and the TAP1 and LMP2 genes (Wright et al., 1995). Other divergent genes may have distinct promoters but share some regulatory elements: for example, the collagen type IV genes, COL4A1 and COL4A2 (Heikkila et al., 1993, Kaytes et al., 1988, Pollner et al., 1997) and the Xenopus H2A and H2B genes (El-Hodiri and Perry, 1995). Many of these, like the Brca1 and Nbr1 gene promoters, also lack TATA boxes. Therefore, the promoter regions of the Brca1 and Nbr1 genes were studied to determine if they also might share some regulatory elements.

The promoter analysis experiments were carried out using a dual luciferase reporter gene system (Promega). The promoter sequences were cloned into a pGL3 basic vector, upstream of the firefly luciferase gene. These constructs were transfected into cell lines along with a vector containing the Renilla luciferase gene and its promoter, as a control for transfection efficiency. These two luciferases have different substrate requirements and so it is possible to measure the luminescence from each separately and compare the experimental luminescence (firefly) to the reference (Renilla).

A 5.5 kb EcoRI/PstI fragment, covering exon 1a of the Brca1 gene and exons 1a, 1b, 1c, 1d and 2 of the Nbr1 gene, had been cloned into pBluescript (see section
This plasmid was used as a template to amplify promoter regions by PCR. Primers were designed that contained suitable restriction enzyme sites for cloning the resultant PCR products into the pGL3 basic vector. A series of fragments was amplified and cloned into the pGL3 basic vector, some in both orientations to test for bidirectional promoter activity. Each clone was fully sequenced in both directions, to confirm the accuracy of the sequence. This yielded ten constructs (Figure 6.5). A mouse 3T3K fibroblast cell line was selected for analysis of these promoter constructs, because RT-PCR performed on RNA from these cells confirmed that Brca1 exon 1a-, Nbr1 exon 1a- and Nbr1 exon 1b-containing transcripts were expressed. Hence, the necessary transcription factors would be present in these cells for regulation of the experimental promoters.

The promoter constructs and the reference plasmid were transiently transfected into the 3T3K cells and harvested after two days. Cell lysates were assayed for luminescence and compared to either a negative control, the pGL3 basic vector, or to a positive control, the pGL3 control vector, which contains the firefly promoter upstream of the gene. The activity of pGLMP0.6N was arbitrarily given the value of 100 and used as the standard for subsequent experiments. The Nbr1 β promoter (transcribing exon 1b) had the strongest activity, approximately 3-fold more than the Brca1 promoter and 4-fold stronger than the Nbr1 α promoter (transcribing exon 1a) (Figure 6.6). Expression of Brca1 and Nbr1 exon 1a transcripts is likely to be controlled by a bidirectional promoter, since activity was observed in both directions with the pGLMP1.0, pGLMP0.7 and pGLMP0.4 constructs. Little variation in activity was seen between the different constructs representing the Nbr1 β promoter. A 2-4-fold reduction in activity was observed with construct pGLMP0.4B in the Brca1 direction, compared with pGLMP1.0B and pGLMP0.7B, suggesting the presence of a positive factor in the sequence lacking in pGLMP0.4B. A conserved Sp1 binding site is present in this region and the absence of this could be responsible for the observed reduction in promoter activity. The pGLMP1.0N construct showed greatest Nbr1 α activity. This construct contains a PEA3 binding site, conserved in humans, that is absent in the pGLMP0.7N and pGLMP0.4N constructs.

Promoter analysis has also been carried out on the human BRCA1 gene. BRCA1 promoter α is bidirectional and also controls expression of the divergent NBR2 gene (Xu et al., 1997c). A 270 bp fragment was found to have maximal activity in both directions and this was approximately 2.5-fold stronger in the NBR2 orientation. These two constructs were transfected into mouse fibroblasts.
Figure 6.5 Diagram illustrating the constructs used in the luciferase reporter gene assays. *Brcal* exon 1a is represented by a box and *Nbrl* exons 1a and 1b by hatched boxes. The constructs, and the direction in which they were cloned, are illustrated by arrows. The coverage of the constructs is shown in nucleotides (nts), where the numbering refers to that shown in Figure 6.4.
Figure 6.6 Analysis of the Brca1 and Nbr1 promoter regions using a luciferase reporter gene assay system. The results from three experiments are illustrated. The name of each construct, and the promoter region from which it is derived, is indicated. The activity of the pGLMP0.6N construct was arbitrarily defined as 100 and all other luciferase activities are measured relative to this construct.
Activity was observed in both directions (Figure 6.6), indicating that at least some of the factors involved in regulating this human promoter are also present in the mouse.

6.4 Summary of the results from studying the Brca1/ Nbr1 intergenic region

Promoters consist of cis-regulatory elements that are responsible for accurate and efficient gene expression. Control of expression through the use of multiple promoters and alternative transcription start sites is a frequently used mechanism. Both the human BRCA1 and NBR1 genes have alternate first exons and the BRCA1 gene is under the control of two promoters (Brown et al., 1994, Xu et al., 1995, Xu et al., 1997c). The murine Nbr1 gene appears to have two alternative 5' exons, exons 1a and 1b. Exon 1b was cloned by RACE; exon 1a has yet to be cloned, but results from RT-PCR studies indicate that it is expressed. This suggested that the Nbr1 gene may be under the control of two promoters. Promoter constructs were designed and analysed, using a luciferase reporter gene assay system and two separate promoters do indeed seem to be controlling Nbr1 expression.

Multiple promoters provide additional flexibility in the control of gene expression (reviewed in Ayoubi and Van de Ven, 1996). They can regulate the expression level through different strengths, or different responsiveness to cellular or metabolic stimuli. For example, the glucokinase gene has two promoters, only one of which is responsive to insulin (Iynedjian, 1993) and the mdm2 gene has a promoter strongly activated in the presence of p53 and a second promoter active in its absence (Barak et al., 1994). In addition, transcripts may have different stabilities and translation efficiencies, due to the different 5' leader exons: for example, the HOX-5.1 gene has transcripts of different sizes, differing in their stability (Cianetti et al., 1990). Alternative promoters are also often responsible for tissue-specific expression or developmental stage-specific expression: one promoter may have ubiquitous expression, whilst the other may have a more restricted pattern. For example, the porphobilinogen deaminase gene has a housekeeping promoter and a second promoter that is only active in erythroid cells (Chretien et al., 1988). This may also be true in the case of Nbr1. Transcripts containing exon 1b were detected in almost all tissues examined, whereas the 1a transcript had a much more restricted pattern. This could be investigated further by performing the reporter gene assays in a variety of cell lines. Alternatively, the observed tissue-specific expression may reflect
differences in the strength of the promoters, such that the 1b transcript is more readily detected than the 1a transcript. The reporter gene assays suggest that promoter β may indeed be stronger than promoter α.

Analysis of the intergenic region between the Nbr1 and Brca1 genes indicates that the 5' ends of these genes are separated by less than 328 bp. This head to head arrangement has been observed for several other genes, many of which share a bidirectional promoter. Therefore, constructs were designed to test whether or not the Nbr1 α promoter was bidirectional. Activity was observed in both directions, suggesting that the Brca1 and Nbr1 genes may indeed share a bidirectional promoter, or at least some regulatory elements.

Adjacent genes may often be part of the same family: for example, the Surf-1 and Surf-2 genes (Lennard and Fried, 1991), and the Xenopus histone genes, H2A and H2B (El-Hodiri and Perry, 1995). They may interact with each other: for example, the collagen type IV genes (Kaytes et al., 1988), or they may be involved in the same pathway: for example, the chicken GPAT and AIRC genes are both involved in the purine biosynthetic pathway (Gavalas and Zalkin, 1995). The Brca1 and Nbr1 genes have no sequence homology and are therefore unlikely to be members of the same family. However, it may be worth investigating whether or not they interact with each other, or are involved in the same pathway.

The human BRCA1 gene also has a bidirectional promoter, promoter α, that is shared with the NBR2 gene. However, there are differences between the human and mouse genes. Whilst two alternative first exons, with separate promoters, exist in the human, only one of these exons is present in the mouse. Therefore, the control of gene expression in the human and the mouse is likely to be different and a single promoter in the mouse may restrict the diversity of expression. In addition, the BRCA1 bidirectional promoter is shared with NBR2 in the human, but with NBR1 in the mouse. Several transcription factor binding sites present in the mouse are not conserved in the human, which may also permit differences in the regulation of gene expression in the two species. For these reasons, caution must be exercised in using the mouse as a model for human BRCA1-induced carcinogenesis.
CHAPTER SEVEN: DISCUSSION

7.1 Findings of this thesis

7.1.1 Isolation of candidate cDNAs within the BRCA1 gene region

At the time of commencing this thesis, the BRCA1 gene had been mapped to chromosome 17q and physical maps were being constructed. The next step was to identify genes within the region. The techniques I chose to use were to look for evolutionarily conserved sequences and to analyse clones from mini-libraries created by direct selection, as described in chapter three.

Cross-species homology has been used successfully to identify several disease genes: for example, the Duchenne muscular dystrophy gene (Monaco et al., 1986). However, since this technique will not detect genes that have diverged from other species, clones from libraries created by direct selection were also analysed. This method can detect transcripts that have diverged from other species and should also enable detection of rare transcripts (Korn et al., 1992). One disadvantage of this method is that it will not detect transcripts that are not expressed in the tissues used to construct the library. Our libraries were therefore generated from pooled cDNA libraries. Since each technique has some disadvantages, it is usual to employ a variety of methods. Other methods being used in our laboratory included exon trapping and identification of CpG islands (Brown et al., 1995, Jones et al., 1994).

Using these techniques, at least seven genes were identified in the 400 kb region between the RNU2 locus and the PPy gene. The importance of using more than one technique was illustrated by the observation that not all transcripts were detected by all methods: for example, the G1151-a gene was only detected by cross-species homology and the HRH1 gene was only detected by analysis of the mini-libraries. In addition, several of the genes identified in our laboratory were not isolated by other groups, and Harshman et al report a similar lack of overlap of their transcripts with others, although the true extent of overlap will not be known until full-length clones have been isolated for all the transcripts (Harshman et al., 1995). These results may suggest that the transcript map is not complete.
In addition to the difficulty of constructing a complete transcript map, problems in obtaining full-length cDNA clones were also encountered. For one gene, G1151-a, no satisfactory cDNA clones were isolated despite the screening of numerous libraries. This may reflect the problem of identifying appropriate libraries to screen when a gene has a tissue-specific pattern of expression, or is developmentally regulated (information which may not be available at the time). In the absence of cDNA clones, Northern blot analysis was used to verify the presence of expressed sequence within the genomic fragment.

Characterisation of the cDNAs isolated in these studies revealed that four of the transcripts identified have no homology to known genes and may therefore be novel genes. One transcript, HRH1, has homology to the yeast pre-mRNA processing factor, PRP22, that is thought to be involved in the release of spliced mRNAs from the spliceosome (Company et al., 1991, Ono et al., 1994). Another transcript has homology to the human endogenous retrovirus type K family (HERV-K). These retroviruses are found integrated into the human genome. If integration occurs within an exon of a gene, it could interrupt the open reading frame and lead to loss of function. Alternatively, integration into promoter regions of proto-oncogenes could lead to overexpression and tumour development (reviewed in Lower et al., 1996). A transcript, identified as the DLG3 gene, was also isolated. This DLG3 gene product has homology to the Drosophila discs-large family of proteins (Smith et al., 1996a). One family member, dlg-A, is a tumour suppressor gene, thought to be involved in signal transduction. Mutations in this gene lead to loss of apical-basolateral polarity, disruption of normal cell-cell adhesion and neoplastic overgrowth of the imaginal disc epithelium (Woods and Bryant, 1991).

7.1.2 Mutation analysis of the BRCA1 gene

A strong candidate for the BRCA1 gene was isolated in 1994 (Miki et al., 1994). To confirm the isolation of the susceptibility gene, worldwide efforts focused on the identification of mutations in breast and breast/ovarian cancer families. Our laboratory focused on 53 families with early-onset cancer. Four of these families had good evidence for linkage to BRCA1 and were considered very likely to carry a BRCA1 mutation. The method of direct sequencing was selected, as this is currently the most accurate and sensitive method of detecting mutations. Truncating mutations, that segregate with the disease, were found in two families; a third family is believed to carry a regulatory mutation, since the
affected family members had lost expression of one allele (Xu et al., 1997b). No mutation was identified in the fourth family and it is possible that the disease in this family is due to a germline mutation in a different gene, for example, BRCA2.

The remaining families were screened for three common mutations, the 185delAG mutation in exon 2 (which had already been detected in one of our families), the 2800delAA mutation in exon 11 and the 5382insC mutation in exon 20 (Boyd et al., 1995, Shattuck-Eidens et al., 1995). The method of ASO hybridisation was used because this is a simple and rapid screening technique. Two families carrying the exon 2 mutation were identified, but no families were found to carry the exon 11 or exon 20 mutations.

The 185delAG mutation has been reported in several families, most of which share the same haplotype and are of Ashkenazi Jewish origin (Friedman et al., 1995b, Neuhausen et al., 1996, Simard et al., 1994). The three families in our study carrying the 185delAG mutation all share the same haplotype, although this is different from that seen in the Ashkenazi Jewish population, suggesting that these mutations may have arisen independently. Approximately 1% of the Ashkenazi Jewish population is estimated to carry the 185delAG mutation, and statistical analyses predict that this mutation may account for 16% of breast cancer and 39% of ovarian cancer in Ashkenazi Jewish women under the age of 50 (Roa et al., 1996, Struewing et al., 1995). Two other mutations have been found to occur frequently in this population: 5382insC in BRCA1 and 6174delT in BRCA2. These three mutations together may account for up to 60% of breast cancer and 30% of ovarian cancer in Ashkenazi Jewish women under the age of 40 (Abeliovich et al., 1997). Hence, screening may now be feasible for women in this population.

Of the five mutations identified in this thesis, four were small deletions or insertions predicted to cause a truncated protein. This is in agreement with other observations and it is currently estimated that 87% of BRCA1 mutations result in truncated or absent BRCA1 protein (Couch et al., 1996).

7.1.3 Conservation of the BRCA1 region in the mouse

The mouse homologues of several human tumour suppressor genes have been isolated and many of these are highly conserved: for example, the human and mouse NF1 predicted proteins are 98% identical and the WT1 predicted protein
is 95% homologous between the two species (Bernards et al., 1993, Buckler et al., 1991b). From the results of hybridisations to zoo blots presented in chapter five, the BRCA1 gene and the neighbouring NBR1 gene are conserved in a range of species, including mice. Isolation of the murine homologues reveals that the NBR1 predicted protein is highly conserved between the human and the mouse (84% identity). This high homology suggests that the NBR1 protein has a significant, although as yet unknown, function in both species.

The conservation of BRCA1 is lower (58%) than that observed for other tumour suppressor genes, although it is similar to that for BRCA2 (Connor et al., 1997, McAllister et al., 1997, Sharan and Bradley, 1997). High homology was observed in two domains, the RING finger and the BRCT domain, both of which are likely to be important for normal function: the RING finger interacts with a novel protein BARD1 (Wu et al., 1996) and the BRCT domain has transactivation potential (Chapman and Verma, 1996, Monteiro et al., 1996). Hence, the identification of homologues can contribute to identifying functional domains.

Analysis of the isolated Nbr1 and Brca1 murine cDNA and genomic clones also indicates that the structure of these genes is well conserved, with only a few differences observed in the exon/intron boundaries of the human and mouse NBR1 genes and none in the human and mouse BRCA1 genes.

7.1.4 Analysis of the promoter regions of the Brca1 and Nbr1 genes

Analysis of cDNA clones and RT-PCR reactions suggests that the mouse Nbr1 gene, like the human, may have two alternate 5' exons. Whilst screening cDNA libraries and RACE have failed to clone one of these murine exons (exon 1a), the presence of this exon in RT-PCR products suggests that this sequence may be expressed in the mouse. This is supported by the results from the reporter gene assays, where activity was observed with constructs upstream of each exon, suggesting that the expression of these transcripts may be controlled by two separate promoters.

The presence of multiple promoters and transcription start sites is a common phenomenon: for example, the human AMP deaminase gene has three alternative 5' exons, under the control of three promoters (Mahnke-Zizelman et al., 1996) and the rat pyruvate carboxylase gene uses two promoters to generate five transcripts (Jitrapakee et al., 1997). The use of multiple promoters and
transcription start sites provides diversity and flexibility in gene expression. Transcripts with different 5' leader exons may have different stabilities or translation efficiencies and alternative promoters may differ in their level of expression, their tissue specificity and their response to stimuli such as hormones. It was therefore interesting to observe that, whilst the transcript containing Nbr1 exon 1b was readily detectable in almost all tissues examined, the transcript containing exon 1a was only easily detected in testis and thymus. The presence in the mouse of two additional 5' exons, exons 1c and 1d, increases the length of the 5' untranslated region compared with the human. This could create differences between the two species in the stability or translation efficiencies of the transcripts.

Like NBR1, the human BRCA1 gene also has two alternative 5' exons, exons 1A and 1B, under the control of two promoters (Xu et al., 1997c). The mouse homologue of exon 1A has been identified and cloned. However, no sequence with homology to exon 1B has been identified in either the murine cDNA or genomic clones. This, together with the observation that human exon 1B contains an Alu element (unique to primates), suggests that only exon 1a is present in the mouse. Therefore, due to the lack of exon 1B in the mouse, the control of BRCA1 expression may differ between human and mouse.

Studies of the isolated murine cDNA and genomic clones indicate that exon 1a of the Brca1 gene and exon 1b of the Nbr1 gene lie head to head, less than 1 kb apart. The intergenic region is probably much smaller than this, since RT-PCR results suggest that Nbr1 exon 1a, exists upstream of exon 1b. The 5' primer used to amplify exon 1a in the RT-PCR reaction is located 328 bp from Brca1 exon 1a, but the two genes may be closer, since the precise 5' end of Nbr1 exon 1a is not yet known.

The close proximity of the 5' ends of these two genes indicates that they might share a bidirectional promoter, or at least some regulatory elements. Indeed, the reporter gene assays showed that the sequence between Brca1 exon 1a and Nbr1 exon 1a does have promoter activity when cloned in either direction. There are two models for bidirectional transcription: it can be due to two independent promoters arranged back to back - for example the human and mouse COL4A1 and COL4A2 genes (Heikkila et al., 1993, Kaytes et al., 1988, Schmidt et al., 1993) - or to a single bidirectional promoter driving expression of both genes - for example the human and mouse dihydrofolate reductase and hMSH3 genes.
(Crouse et al., 1985, Farnham et al., 1985, Shinya and Shimada, 1994). The Wilms' tumour gene, WT1, is divergently expressed from the Wit1 gene (Campbell et al., 1994a). In this case, exon 1 of the WT1 gene is also found in the Wit1 transcripts and this antisense WT1 exon 1 transcript is capable of downregulating WT1 expression (Eccles et al., 1994, Malik et al., 1995). Whether or not Brcal is regulated by antisense transcripts will not be known until the precise 5' end of Nbr1 exon 1a is determined.

The human and mouse BRCA1 promoters lack a TATA box, as is often observed for bidirectional promoters. In the absence of the TATA box, an initiator element controls the accurate initiation of transcription (reviewed in Azizkhan et al., 1993, Smale, 1997). Potential initiator elements for the Brcal and the Nbr1 genes have been identified (see chapter six, Figure 6.4). The presence of multiple Sp1 binding sites is common in TATA-less promoters and these sites are probably involved in determining both the strength and accuracy of the promoter. There is an Sp1 binding site present in Brcal intron 1a that is conserved in the human and may therefore be important in controlling transcription.

Some divergent transcripts have similar functions, or interact together: for example, the collagen type IV genes (Kaytes et al., 1988) and the Xenopus histone H2A and H2B genes (El-Hodiri and Perry, 1995). Others are involved in the same pathway: for example, the chicken GPAT and AIRC genes are both involved in the de novo purine biosynthetic pathway (Gavalas and Zalkin, 1995). Whether or not Brcal and Nbr1 are related in any way remains to be determined.

7.2 Implications for future work

During the positional cloning of the BRCA1 gene, several novel transcripts were identified. In addition, transcripts with homology to tumour suppressor genes and human endogenous retroviruses were also identified. Full-length clones of the novel genes should be isolated and the role in tumorigenesis of these, the DLG3 gene and the HERV-K related gene investigated.

The mutation analysis of the BRCA1 gene in breast and breast/ovarian cancer families detected a regulatory mutation in family PRU7B5. Studies of RNA from an affected individual revealed that only one allele is expressed. The precise mutation in this family has not been identified, although it may be present in the regulatory regions of the gene, such as introns, untranslated regions, or the
promoter. Recently, several groups have reported large genomic deletions, some of which include the promoter region and may account for the presumed regulatory mutations (Petrij-Bosch et al., 1997, Puget et al., 1997, Swensen et al., 1997). Hence, family PRU7B5 is now being examined for the presence of large genomic deletions that might result in the loss of transcript. Sporadic breast tumours, in which BRCA1 mutations have not yet been found, could also be investigated for large genomic deletions, or other possibilities, as discussed in chapter one (see section 1.5.4).

Isolation of the murine homologue of the NBR1 gene showed that this gene is highly conserved between the two species. However, the function of NBR1 is still unknown. A B-box, a type of zinc finger, is the only motif identified in NBR1. Several B-box-containing genes have transformation potential when fused to other genes, for example, ret and pml (Freemont, 1993). Further experiments are required to determine the functional significance of this motif in NBR1. In addition, the relationship between NBR1 and the tumour antigen, CA125 and hence the role of NBR1 in ovarian tumorigenesis, needs to be addressed.

The isolated murine cDNA and genomic clones will act as resources for studying the function of the BRCA1 and NBR1 genes in the mouse. A variety of studies will now be possible, including the study of the expression of these genes throughout development and in response to stimuli. The murine genome can be manipulated to generate transgenic and knockout mice. This could provide insight into mechanisms of tumour development, and hence the generation and assessment of potential therapeutic strategies. Such studies may also aid the elucidation of the function of NBR1. Many studies are already under way with the Brca1 gene and have shown that the expression of this gene is critical for normal embryonic development in the mouse (Gowen et al., 1996, Hakem et al., 1996, Liu et al., 1996).

Characterisation of the promoter regions of the Brca1 and Nbr1 genes suggests that these genes share some regulatory elements. Further analysis will indicate whether or not they share a truly bidirectional promoter and if the expression of these genes is co-regulated. The generation of more constructs and site-directed mutagenesis of transcription factor binding sites will further reveal the significance of the identified binding sites. The promoter activities of these constructs in other cell lines could be determined to assay for tissue-specific expression.
Transfection of human constructs into a mouse cell line revealed that at least some regulatory elements are shared between the two species. This suggests that although a duplication is present in the human and \textit{BRCA1} is adjacent to the \textit{NBR2} gene not the \textit{NBR1} gene as in the mouse, sufficient similarity may remain to confer similar control of \textit{BRCA1} expression in both the human and the mouse. A murine homologue of the \textit{NBR2} gene is not present in this region of the mouse, although it is possible that such a gene may exist elsewhere in the murine genome. The function of \textit{NBR2} is unknown and studies are under way to determine the significance of this gene and its role, if any, in tumorigenesis.

In conclusion, several genes from the \textit{BRCA1} region have been isolated and characterised in the human and the mouse. Many of these genes are novel and their functions are as yet unknown. At least one newly identified gene, \textit{DLG3}, has homology to a tumour suppressor gene and therefore, its role in tumorigenesis should be investigated. Since at least three regions of LOH on chromosome 17q have been reported in breast and ovarian tumours (see chapter one, section 1.5.4), it is possible that one or more of these novel genes may also be involved in tumorigenesis. Two genes, \textit{NBR1} and \textit{NBR2}, are adjacent to \textit{BRCA1} in the mouse and the human respectively and hence may be directly involved in \textit{BRCA1}-induced carcinogenesis. The isolation of the murine homologues of the \textit{BRCA1} and \textit{NBR1} genes will be valuable tools in unravelling the function and complexity of expression of these genes.
APPENDIX A: Primers used for PCR and sequencing during the mutation analysis of the *BRCA1* gene in breast and breast/ovarian cancer families.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>exon 2</td>
<td>GAAGTTGTCAATT TATAAAACCTTT</td>
<td>TGTCTTTCTCCCTAGTATGT</td>
</tr>
<tr>
<td>exon 3</td>
<td>TCCTGACACAGCAGACATTTA</td>
<td>TTGGATTTTCTTCTACCTTA</td>
</tr>
<tr>
<td>exon 4</td>
<td>GTCAAAGAGATAGAATGTGAGC</td>
<td>CCCGTCTTACAGAAAACAC</td>
</tr>
<tr>
<td>exon 5</td>
<td>CTCTTAAGGGCCAGTTGTGAG</td>
<td>TTCCCTACTGTTGTTGCTTCC</td>
</tr>
<tr>
<td>exons 6 and 7</td>
<td>CTTATTTTAGTGTCCTTTAAAAGG</td>
<td>TCGGTTTCACTCTGTAGAAG</td>
</tr>
<tr>
<td>exon 8</td>
<td>TGTTAGCTGACTGATGATGGT</td>
<td>ATCCAGCAATTATTATTAATAC</td>
</tr>
<tr>
<td>exon 9</td>
<td>CCACAGTAGATGCTCAGTAATA</td>
<td>TAGGAAAAATACCAGCTTCATAG</td>
</tr>
<tr>
<td>exon 10</td>
<td>TGGTCAGCTTTCTGTAAATCG</td>
<td>GTATCTACCACCTCTCTTCAG</td>
</tr>
<tr>
<td>exon 11</td>
<td>GGAATTTAATGAAAGAGATGAGC</td>
<td>TGTGAGGGACGCTCTTG</td>
</tr>
<tr>
<td></td>
<td>TACGGGAAACCTATCGGAA</td>
<td>ATCTTTGGGGTCTTCAG</td>
</tr>
<tr>
<td></td>
<td>GTGTTCAAAATACCAGCTGACTTA</td>
<td>TAGGAAAAATACCAGCTTCATAG</td>
</tr>
<tr>
<td></td>
<td>GGAGCCCGTTCATTAGTAC</td>
<td>GTGCTCCCAAGCACATAA</td>
</tr>
<tr>
<td></td>
<td>CCAAGTACAGTGAGCACAAATTA</td>
<td>GTGCTCCCAAGCACATAA</td>
</tr>
</tbody>
</table>

152
<table>
<thead>
<tr>
<th>Exon 12</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GTCCTGCCAATGAGAAGAAA</td>
<td>TGCAGCAACCTAAGAATGT</td>
</tr>
<tr>
<td>Exon 13</td>
<td>AATGGAAAGCTTCTCAAAGTA</td>
<td>ATGTTGGGCTAGGTCCTTAC</td>
</tr>
<tr>
<td>Exon 14</td>
<td>CTAACCTGAATTATCCTATCA</td>
<td>GTGATATAATGCCTGTATGCA</td>
</tr>
<tr>
<td>Exon 15</td>
<td>TGGCTGCCAGGAAGTATG</td>
<td>AACCAGAATATCTTTATGTAGGA</td>
</tr>
<tr>
<td>Exon 16</td>
<td>AATTCTTAACAGAGACCAGAAC</td>
<td>AAAACTCTTTCCAGAATGTTGT</td>
</tr>
<tr>
<td>Exon 17</td>
<td>GTGTAAGACGTGCAGGATTG</td>
<td>TCGCCTCATGTGGTTTA</td>
</tr>
<tr>
<td>Exon 18</td>
<td>GGCTCTTTAGCTTCTTAGGAC</td>
<td>GAGACCATTITCCCAGCATC</td>
</tr>
<tr>
<td>Exon 19</td>
<td>CTGTCATTTCTCTGTGCTC</td>
<td>CATTTAAGGAAAGTGTGC</td>
</tr>
<tr>
<td>Exon 20</td>
<td>ATATGACGTGCTGCTCCAC</td>
<td>GGGAAATCCAAATTACACAGC</td>
</tr>
<tr>
<td>Exon 21</td>
<td>AAGCTTCTCTTTTTGAAAGTC</td>
<td>GTAGAGAAATAGAATAGCCTCT</td>
</tr>
<tr>
<td>Exon 22</td>
<td>TCCCATGGAGAGGCTTGGCT</td>
<td>GAGAAGACTTTCTGAGGCTAC</td>
</tr>
<tr>
<td>Exon 23</td>
<td>CAGAGCAAGACCCCTGTCTC</td>
<td>ACTGTGCTACTAAGCACCA</td>
</tr>
</tbody>
</table>
exon 24

<table>
<thead>
<tr>
<th>Type</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>forward primer</td>
<td>ATGAATTGACACTAATCTCTGC</td>
</tr>
<tr>
<td>reverse primer</td>
<td>GTAGCCAGGACAGTAGAAGGA</td>
</tr>
</tbody>
</table>
APPENDIX B: Nucleotide sequence of the murine Nbr1 mRNA (accession number: U73039)

1  CGCCATGTTA GATTCACCCC GCAGGGATAG CAGGGAGCT GGCCGCGA
51  GGTCTTTGCA CTGGCCTACTG GCAAGCGCCC CCCGTGGGGC GAAAGCTGAA
101  AAAACAAACA AACCAACCAA CTGTGAAGCT GCCTGGAGTT GCTGGGGAGT
151  TGATGAGCCA GCGCTGCTCC AAATTTTACA GCTCTTGAAG AGGAAAGGATT
201  GCATTTCAAG CTGACCTAGA ACTCTTGAGG TTTCCATTAC ATCCCTGACTC
251  AACTCCCTAG CGAATGGGAC TACACTCCCA GCCAGAGCCT ATGAAACCAC
301  AGGTTACTCT AAAATGTGACTTTTAAAAATGAAACTCAAAGCTTTCTGGTT
351  TTGAGGATTG CAGACATGAAGACCATGGGGGAGTTCATGGGAA AGGAGAAGCC
401  TTCTGTTTTT CAGAGAAGACTTCTCAGT CAGTCAAGC TCCAGAGCCC
451  CTGCCAGAGA GAGGAACTTTTCTGACATCCA GATATTGAGC TATGAAACCAC
501  GTCTTGGGAT CAGACATGAAGACCATGGGGGAGTTCATGGGAA AGGAGAAGCC
551  CCCTGTTTTT CAGAGAAGACTTCTCAGT CAGTCAAGC TCCAGAGCCC
601  CTGCCAGAGA GAGGAACTTTTCTGACATCCA GATATTGAGC TATGAAACCAC
651  GTCTTGGGAT CAGACATGAAGACCATGGGGGAGTTCATGGGAA AGGAGAAGCC
701  TTCTGTTTTT CAGAGAAGACTTCTCAGT CAGTCAAGC TCCAGAGCCC
751  CTGCCAGAGA GAGGAACTTTTCTGACATCCA GATATTGAGC TATGAAACCAC
801  GTCTTGGGAT CAGACATGAAGACCATGGGGGAGTTCATGGGAA AGGAGAAGCC
851  TTCTGTTTTT CAGAGAAGACTTCTCAGT CAGTCAAGC TCCAGAGCCC
901  CTGCCAGAGA GAGGAACTTTTCTGACATCCA GATATTGAGC TATGAAACCAC
951  CTACAACATT GTGAAGAGTT GTGAAGCTGG ACCATACACCATGACACTA
1001  ACCATGTGTTT GTTAAGCTTG GGAGACCTTG TTTGTGACAT CTCTGAGC
1051  TTTTCTCTTTT CCAATGGTAC CCGCTCTCCTCGTCTCTGTCCT CCCTTTGGG
1101  AGTCATTAG CAAAACAGGG TGGACAAGAA CTTTGTGAAA GCAGAAAGCC
1151  AAAGGTTGGG GGCTGAGCGA AACGAGCAGT ACATATGCCA AAGGCTTGGT
1201  AAAAGGCAGC TTAAGTCTCA CAGGAAGATT ATTTATGGA ATTCAGACCA
1251  CAGAACTTCTT CAGAGAAGACTTCTCAGT CAGTCAAGC TCCAGAGCCC
1301  CAGAACTTCTT CAGAGAAGACTTCTCAGT CAGTCAAGC TCCAGAGCCC
1351  AGTCTACAC ACCAGGTCG CTCTTGAGG CCCCCGAGAAGGATGAA GCCAGAAAG
1401  ACTCTCCAGG TTTGGGTTT GTGCCTTCTGG TGGCAGGTTT GGGATCAGCT
1451  TCAAGGAAAG GTGCTCGGAG CGGAGAGCCT GGAGATAAGT TTATGGAAGG
1501  ATGTGAAGTTT GACAGAGACT TGTTCAAGG CAGAGACTTCT CTCTGAGC
1551  ACTCTCGCTCTT CACAGGACG AAGATACATG TGTGACATAGT TGGGTAAGG
1601  TTGAGGATTG CAGACATGAAGACCATGGGGGAGTTCATGGGAA AGGAGAAGCC
1651  TTGAGGATTG CAGACATGAAGACCATGGGGGAGTTCATGGGAA AGGAGAAGCC
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1701</td>
<td>CCTCGGGTTT GGTGCAGTAT CATAGTGAGT CTTTTTCTTT CTCCTGAGAG</td>
</tr>
<tr>
<td>1751</td>
<td>CTCCTGATAAT GTTGAAGGGA ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>1801</td>
<td>TCTCCTGTGA GCAAGAGGAA GCTTTTCTTC TAGCTGAGAG AGAGATTCCCA</td>
</tr>
<tr>
<td>1851</td>
<td>TTGGGTGAGGG ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>1901</td>
<td>CTCCTGACTG TCTCCTCTGC CATAGTGAGG AAGATCCTAA TGAGAGAGCC</td>
</tr>
<tr>
<td>1951</td>
<td>GCAGCCACTT GTGCTGCTTG GATGGATCTG AAGAGGACAG CTCAAGGACA</td>
</tr>
<tr>
<td>2001</td>
<td>TTGGGTGAGGG ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>2051</td>
<td>GCAAGAACAC GAAAGGACAG TTAGGGGAGG GAGGGATTT TAAAGCACCT</td>
</tr>
<tr>
<td>2101</td>
<td>CTCAGATGACTG CTTTACCGAG TCTGAGAAGG AGGAGGAGAC GAAAACTTCC</td>
</tr>
<tr>
<td>2151</td>
<td>AGGTAGGGTCT CCCCAGAGG GAAATGACC GATGTATTCC CCGAGGGACA</td>
</tr>
<tr>
<td>2201</td>
<td>GCAGCCACTT GTGCTGCTTG GATGGATCTG AAGAGGACAG CTCAAGGACA</td>
</tr>
<tr>
<td>2251</td>
<td>TTGGGTGAGGG ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>2301</td>
<td>GCAAGAACAC GAAAGGACAG TTAGGGGAGG GAGGGATTT TAAAGCACCT</td>
</tr>
<tr>
<td>2351</td>
<td>CTCAGATGACTG CTTTACCGAG TCTGAGAAGG AGGAGGAGAC GAAAACTTCC</td>
</tr>
<tr>
<td>2401</td>
<td>AGGTAGGGTCT CCCCAGAGG GAAATGACC GATGTATTCC CCGAGGGACA</td>
</tr>
<tr>
<td>2451</td>
<td>GCAGCCACTT GTGCTGCTTG GATGGATCTG AAGAGGACAG CTCAAGGACA</td>
</tr>
<tr>
<td>2501</td>
<td>TTGGGTGAGGG ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>2551</td>
<td>GCAAGAACAC GAAAGGACAG TTAGGGGAGG GAGGGATTT TAAAGCACCT</td>
</tr>
<tr>
<td>2601</td>
<td>CTCAGATGACTG CTTTACCGAG TCTGAGAAGG AGGAGGAGAC GAAAACTTCC</td>
</tr>
<tr>
<td>2651</td>
<td>AGGTAGGGTCT CCCCAGAGG GAAATGACC GATGTATTCC CCGAGGGACA</td>
</tr>
<tr>
<td>2701</td>
<td>GCAGCCACTT GTGCTGCTTG GATGGATCTG AAGAGGACAG CTCAAGGACA</td>
</tr>
<tr>
<td>2751</td>
<td>TTGGGTGAGGG ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>2801</td>
<td>GCAAGAACAC GAAAGGACAG TTAGGGGAGG GAGGGATTT TAAAGCACCT</td>
</tr>
<tr>
<td>2851</td>
<td>CTCAGATGACTG CTTTACCGAG TCTGAGAAGG AGGAGGAGAC GAAAACTTCC</td>
</tr>
<tr>
<td>2901</td>
<td>AGGTAGGGTCT CCCCAGAGG GAAATGACC GATGTATTCC CCGAGGGACA</td>
</tr>
<tr>
<td>2951</td>
<td>GCAGCCACTT GTGCTGCTTG GATGGATCTG AAGAGGACAG CTCAAGGACA</td>
</tr>
<tr>
<td>3001</td>
<td>TTGGGTGAGGG ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>3051</td>
<td>GCAAGAACAC GAAAGGACAG TTAGGGGAGG GAGGGATTT TAAAGCACCT</td>
</tr>
<tr>
<td>3101</td>
<td>CTCAGATGACTG CTTTACCGAG TCTGAGAAGG AGGAGGAGAC GAAAACTTCC</td>
</tr>
<tr>
<td>3151</td>
<td>AGGTAGGGTCT CCCCAGAGG GAAATGACC GATGTATTCC CCGAGGGACA</td>
</tr>
<tr>
<td>3201</td>
<td>GCAGCCACTT GTGCTGCTTG GATGGATCTG AAGAGGACAG CTCAAGGACA</td>
</tr>
<tr>
<td>3251</td>
<td>TTGGGTGAGGG ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>3301</td>
<td>GCAAGAACAC GAAAGGACAG TTAGGGGAGG GAGGGATTT TAAAGCACCT</td>
</tr>
<tr>
<td>3351</td>
<td>CTCAGATGACTG CTTTACCGAG TCTGAGAAGG AGGAGGAGAC GAAAACTTCC</td>
</tr>
<tr>
<td>3401</td>
<td>GCAGCCACTT GTGCTGCTTG GATGGATCTG AAGAGGACAG CTCAAGGACA</td>
</tr>
<tr>
<td>3451</td>
<td>TTGGGTGAGGG ACAGGATACG CTCAAGCCAA GCTGACGATT</td>
</tr>
<tr>
<td>3501</td>
<td>GCAAGAACAC GAAAGGACAG TTAGGGGAGG GAGGGATTT TAAAGCACCT</td>
</tr>
<tr>
<td>3551</td>
<td>CTCAGATGACTG CTTTACCGAG TCTGAGAAGG AGGAGGAGAC GAAAACTTCC</td>
</tr>
<tr>
<td>3601</td>
<td>GCAAGAACAC GAAAGGACAG TTAGGGGAGG GAGGGATTT TAAAGCACCT</td>
</tr>
</tbody>
</table>
3651  TGAATGTTTC CAAGAATTAT TCACCTAGTT TCTATAACCA ATCGCAAGCC
3701  TCATAGCTAC ATAAAGACAG CAGATCAAGA AAGTCATGTT CCTCAAGAC
3751  TTGAAATGAA GAACACTTTT GCTTTTCAGAA GCCCTAGCAG GAACCTGGAA
3801  GACAGGAGG CATTCCGGTT AACCCACAGC ACCACCTGAG CAGTGTCTGT
3851  CAAAGTAAATT TTGCATAAAT TTTATTTAAG GAGAATCCTG GGTAGGTGTA
3901  AATATTCTTC CTAATCTATG GGGGGGAC CAC CTGTTACAGA AAGGGAAAGT
3951  TAGTAGTCGT CCTTTTACTG ACTGCACAG TGGATGGGAA GGACAAGCAT
4001  GTGCAATAC TCATTCAGT AAGGTGGGTT GTGGGTGAAA GGACACAGAA
4051  CTGCTGCTGA CTTTCCGAAT GTGTGAATTC CCCTTGAGAG TATCCAGACC
4101  CTAGAAAAGCC AGAACAGGAA AGTAAATGCTT TTGGGCTAAC CCACAGAAAG
4151  GGTGCTTTTC CATTAGAATA GATGGGAAGA AGGCTCTGTC AAGAGTTCTA
4201  GGCCTCATTG AGTTTATAAA GGGGAAGCTT GGGCCAGAT GTGGAAAGGAA
4251  AAGGCAACAC AACAATAAAG ATTTATTCA TGTGGAGAAA AAGGTATCTG
4301  AACTCCAGCT CAAAGGGGAT AGTTAAAATA GTATTATTTA ACTTAGTTGG
4351  TATTTATAAT AAATGCGGAT CTTAATTACT CATATATGT TTATTTACAA
4401  TCTAATCTCTT GAAGGCTTCT AAATAACCA CAAATCAAAC TGCA
### APPENDIX C: Primers used for determining the mouse *Nbr1* exon/intron structure

<table>
<thead>
<tr>
<th>Introns</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>1F1 AACGGTCTTTGCACACTGGCCA</td>
<td>1R3 TTCTTCCTTCTTAGGAGCTG</td>
<td></td>
</tr>
<tr>
<td>1c</td>
<td>1F2 ACTTGAACGTGCTGGGGAGT</td>
<td>1R1 TAGTCCAGCTACTCAGAAG</td>
<td></td>
</tr>
<tr>
<td>1d</td>
<td>1F3 GCTGACCTAGAACTCTTGGGA</td>
<td>2R1 CAGGTTGTATTTTGTGGATC</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2F1 CTCAAAGCTTTCTGGTCTCG</td>
<td>3R1 CCTCACTCCAGGTATTTTATCTGGATAG</td>
<td></td>
</tr>
<tr>
<td>3 and 4a</td>
<td>3F1 GGTAAAGCTTTCTCTTGAATC</td>
<td>4R1 AAGTGCTTCTTATTCAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4IR1 TTGCTCACTCCACTCTCAGTA</td>
<td>4IF2 CACTGGAACCTATAGGATTAG</td>
<td></td>
</tr>
<tr>
<td>4b and 5</td>
<td>5F2 GTGAATATGAAGAAGCAGC</td>
<td>6R2 CATCTCCAGGTAGCTTGAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5IR1 CGGGAGTCATTCTATTCTTA</td>
<td>6R1 GCAGGCTCCTCTTGTTGC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6R1 GCAGGCTCCTCTTGTTGC</td>
<td>5F1 CAAGTCCACGAAAGGTACCA</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6F2 CTGCTCCTTGTGACACAGAC</td>
<td>7R1 GGCATTGAGACTTCTGTGGG</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7F1 AGAACCAGTTCACTGCGCAT</td>
<td>8R1 AATGGCTCAGAGGATATCAC</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8F1 CCCATCCTACAACATTTGTTG</td>
<td>9R1 TAGACTGCAGCAAGCTCTCA</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9F1 AATCCATGGACTTCAAGATC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>intron</td>
<td>forward primer</td>
<td>reverse primer</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10F1 GTTTATTAAACACTGGAGGA</td>
<td>10R1 CTGAGTGGTGCGATAACTGG</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11F1 CAGGAGCCCTGATAATGTT</td>
<td>11R1 GTACAGCTCCCTTACCTGG</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12F3 ATCTGTGACCTCGACTGG</td>
<td>12R1 GTACAGCTCCCTTACCTGG</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>13F2 CCTGACTTCTTTAGCTGC</td>
<td>13R2 GTGTTGTGTGGGACCTCCT</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14F1 TCTGCCACATGACGCTCCTT</td>
<td>14R1 CAGGAGGTCTTTAAATCC</td>
<td></td>
</tr>
<tr>
<td>15a</td>
<td>15F1 TTGCAGGCAGGTCCCCAC</td>
<td>15R1 CAGGGGCTCTCAGCTTCCTC</td>
<td></td>
</tr>
<tr>
<td>15b</td>
<td>15F2 TTACAGCTGTATCCACCGA</td>
<td>15R2 TTCTGCAGAATCCAGGAGC</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>16F1 CTGGGCTCGAACCAGTCCACA</td>
<td>16R1 TTCTCAACAATGTGGACAA</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>17F1 TCATCAGGAGTTGATTCAACA</td>
<td>17R1 CTCCCTGATGTCATCAGGG</td>
<td></td>
</tr>
<tr>
<td>18a</td>
<td>18R1 CTTGTCGCTTGGGCGAGTC</td>
<td>17F1 CCTCTGTGACTAAATCAGC</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


Neuhausen, S.L., Mazoyer, S., Friedman, L., Stratton, M., Offit, K., Caligo, A.,
Tomlinson, G., Cannon-Albright, L., Bishop, T., Kelsell, D., Solomon, E., Weber,
B., Couch, F., Struwing, J., Tonin, P., Durocher, F., Narod, S., Skolnick, M.H.,
Lenoir, G., Serova, O., Ponder, B., Stoppa-Lyonnet, D., Easton, D., King, M.C. and
mutations in 61 families: results of an international study. *Am J Hum Genet* 58:
271-280.

Nicolaides, N.C., Papadopoulos, N., Liu, B., Wei, Y.-F., Carter, K.C., Ruben, S.M.,
Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D.,
Venter, J.C., Dunlop, M.G., Hamilton, S.R., Petersen, G.M., de la Chapelle, A.,

Niederacher, D., Picard, F., van Roeyen, C., An, H.X., Bender, H.G. and
breast carcinomas detected by fluorescent-labeled microsatellite analysis. *Genes
Chromosomes Cancer* 18: 181-192.

Norris, J., Fan, D.J., Aleman, C., Marks, J.R., Futreal, P.A., Wiseman, R.W.,
subclass of Alu DNA repeats which can function as estrogen receptor-dependent

Offit, K., Gilewski, T., McGuire, P., Schluger, A., Hampel, H., Brown, K.,
Germline BRCA1 185delAG mutations in Jewish women with breast cancer. *Lancet*
347: 1643-1645.

Ohtani-Pujita, N., Pujita, T., Aoike, A., Osifchin, N.E., Robbins, P.D. and Sakai, T.
(1993). CpG methylation inactivates the promoter activity of the human


protein that can interact in vivo with the BRCA1 gene product. *Nature Genet* **14**: 430-440.


