UNIVERSITY OF SOUTHAMPTON

SCHOOL OF MEDICINE

Division of Cancer Sciences

An Investigation of DNA Sequence Variants of Unknown Significance in Hereditary Breast Cancer

by

Ellen Roxane Copson

Thesis for the degree of Doctor of Philosophy

December 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

DIVISION OF CANCER SCIENCES

SCHOOL OF MEDICINE

Doctor of Philosophy

AN INVESTIGATION OF DNA SEQUENCE VARIANTS OF UNKNOWN SIGNIFICANCE IN HEREDITARY BREAST CANCER

by Ellen Roxane Copson

Inheritance of a protein truncating mutation of the tumour suppressor gene BRCA1 causes approximately 5% of breast tumours. Over 450 distinct BRCA1 missense mutations have also been found in patients with a family history of breast cancer and the functional significance of most of these is unclear. Increasing evidence suggests that DNA missense mutations can affect RNA stability or sequence by disrupting splicing regulators. I have used a variety of techniques to investigate the effect of genomic BRCA1 missense mutations on transcript expression.

BRCA1 monoallelic expression has previously been reported in association with missense mutations. I therefore initially used the technique of Pyrosequencing™ to identify imbalances in BRCA1 cDNA allele expression. Analysis of patients with known BRCA1 truncating mutations, missense mutations and controls identified no examples of monoallelic expression and indicated that a previous report of monoallelic BRCA1 expression was erroneous.

I subsequently developed a series of multiplex RT-PCR reactions using overlapping primer pairs to identify alternative BRCA1 transcripts in the same groups of subjects. This technique effectively demonstrated the common BRCA1 isoforms and confirmed that the c4185A>G silent mutation is associated with deletion of exon 12.

To provide a systematic analysis of the effect of BRCA1 mutations on splicing regulation I designed minigenes centred on 4 different BRCA1 exons; 5,6,10 and 18. All mutations within these exons reported to the Breast Cancer Information Core database were introduced into the appropriate minigenes, and wildtype and mutant minigenes were transfected into HEK 293 cells. The resulting transcripts were examined to identify aberrant splicing. Only one of twenty-one missense mutations investigated resulted in alternative transcripts, suggesting that only a small proportion of BRCA1 missense mutations do affect splicing.

Additional work was also performed to investigate whether the MDM2 SNP 309 acts as a disease modifying gene in BRCA1 associated hereditary cancer.

Acknowledgements

First and foremost I must thank my principal supervisors, Professor Diana Eccles and Professor Peter Johnson for their continuing support, guidance, enthusiasm and encouragement throughout the last three years. During the eighteen months I spent working in the Wessex Regional Genetics Laboratories, Salisbury District Hospital, I was constantly advised by David Robinson, Helen White, Julie Silibourne and Francis Grand –a big thank you to you all and also to Professor Nick Cross for allowing me to use the facilities there. The second half of my research period was spent in the Department of Human Genetics, Southampton University and I am very grateful to Dr. Igor Vorechovsky for supervising me throughout this period. I must also thank Professor David Wilson for his support and for allowing me to use the facilities of the Duthie building. In particular, I wish to express my immense gratitude to Jana Kralovicova for all her practical help and support during my battles with minigenes. Mr. Paul Strike has also been extremely helpful in advising me on statistical analyses.

Lastly, I wish to thank Charlie and Emily for all their understanding and patience, and my parents, Pam and Terry, and Philly for all their help and encouragement, and in particular Peter, who taught me so much.

This work was funded by Cancer Research UK.

Abbreviations

ATP	adenosine triphosphate
BARD-1	BRCA-1 associated ring domain protein
BASC	BRCA-1 associated genome-surveillance complex
BCS	breast conserving surgery
BPS	branch point sequence
CFTR	cystic fibrosis transmembrane conductance regulator
Chk2	Checkpoint kinase 2
CSCE	conformation sensitive capillary electrophoresis
CTP	cytosine triphosphate
cDNA	complementary DNA
DNA	deoxyribonucleic acid
DSB	double strand break
EJC	exon junction complex
ER	oestrogen receptor
ESI	exonic splicing inhibitor
ESS	exonic splicing enhancer
GTP	guanine triposphate
hnRNP	heterogeneous nuclear ribonucleoprotein
HR	homologous recombination
ISE	intronic splicing enhancer
ISI	intronic splicing inhibitor
MRNA	messenger ribonucleic acid
MAN Sc1	Manchester Score 1
MAN Sc2	Manchester Score 2
MLPA	multiplex ligation probe analysis
MRI	magnetic resonance imaging
NAS	nonsense associated splicing
NGRL	National Genetics Reference Laboratory
NHEJ	non-homologous end-joining
NMD	nonsense mediated decay
dNTPs	dinucleotide triphosphates
PARP	poly (ADP-ribose) polymerase
PCR	polymerase chain reaction
PPT	polypyrimidine tract
PTC	premature termination codon
PR	progesterone receptor
RNA	ribonucleic acid
RRM	RNA recognition motif
RT-PCR	reverse transcription polymerase chain reaction
SNP	single nucleotide polymorphism
snRNP	small nuclear ribonuclear protein
SR	serine rich
SSB	single strand break
TTP	thymidine triphosphate

UV	unknown variant
WCGS	Wessex Clinical Genetics Service
WRGL	Wessex Regional Genetics Laboratory

List of Contents

List of Tables

List of Figures

1.0 Note regarding nomenclature of BRCA1 mutations

The reference sequences used for the purposes of this research were GenBank U14680 for the cDNA transcript and GenBank L78833 for the genomic DNA sequence. Following the recommendations of the HUGO Working Group [1] in this thesis all BRCA1 nucleotide positions are referred to according to a numbering system in which the "A" of the first translated ATG represents the number 1 position. This numbering system (the HUGO system) therefore does not include the first 119 bases of untranslated 5' BRCA1 sequence and effectively commences numbering at the 20th position of exon 2. The HUGO system has the advantage of accommodating multiple different versions of the 5' untranslated region. It should be noted however that BRCA1 mutations documented in the Breast Information Core Database and the majority of published manuscripts to date have been numbered according to a system in which the first nucleotide of the GenBank entry U14680 is designated as number 1 [2]. For ease of reference, here each previously reported BRCA1 mutation is referred to firstly by the HUGO system (designated with a "c" preceding the nucleotide number) with the original mutation designation also included in brackets.

2.0 Introduction

2.1 Hereditary breast cancer

Breast cancer remains one of the leading health problems of the Western World, currently being the most common malignant diagnosis in the UK. Over 44,000 new cases of breast carcinoma were diagnosed in the UK in 2003 and in the same year it accounted for 12417 deaths [3]. Whilst metastatic breast cancer remains incurable, reducing the number of deaths caused by breast cancer is dependent on either preventing the development of the disease or detecting cases at an early stage so that radical treatment can be performed before micro-metastases have spread from the primary site. Trials of putative preventative agents in the general population are on going, whilst screening by mammography is currently offered only to women aged between 50 to 65 years. An important aspect of breast cancer risk management is the identification of individuals with a genetic predisposition to a high lifetime risk.

Population studies indicate that 20-30% of patients with a diagnosis of breast cancer report a positive family history of this condition [4]. There is growing evidence that the majority of these cases are likely to be associated with exposure to hormonal and environmental risk factors and stochastic genetic events and genetic variants in low-penetrance cancer susceptibility genes [5]. Over the last few years a number of large case-control association studies have identified variants in the DNA repair genes CHEK2, ATM, BRIP1 and PALB2 which appear to double the risk of breast cancer [6-9]. The population prevalence of these variants however seems to be only in the order of 1-5%. A recently published genome–wide association study has now identified a further five novel breast cancer susceptibility loci, four of which lie within plausible causative genes: FGFR2, TNRC9, MAP3K1 and LSP1 [10]. Easton *et al.* estimate that these new loci explain approximately 3.6% of the excess familial risk of breast cancer.

Clinical Syndrome	Gene	Location	Mutation/ Status	Population frequency	Breast Cancer Risk	Ref.
Hereditary breast cancer	BRCA1	17q21	Multiple variants Heterozygous	1: 1000	46-85% lifetime risk	[11]
Hereditary breast cancer	BRCA2	13q12	Multiple variants Heterozygous	1: 750	43-84% lifetime risk	[12]
Li-Fraumeni	TP53	17p13.1	Multiple variants Heterozygous	<400 families worldwide	28-56% by age 45	[13]
Cowden	PTEN	10q23.3	Multiple variants Heterozygous	1: 300,000	25-50% lifetime risk	[14]
Peutz-Jeghers	LKB1/ STK11	19p13.3	Multiple variants Heterozygous	1:8900-1:280000	29-54% lifetime risk	[15]
HDGC	CDH1	16q22.1	Multiple variants Heterozygous	Unknown	20-40% lifetime risk	[16]
Ataxia Telangiectasia (homozygotes)	ATM	11q22-23	Multiple variants Heterozygous	1:250	RR 2.3	[7]
	CHEK2	22q12.1	1100delC Heterozygous	1:100	RR 2.0	[6]
	TGFβ1	19q13.1	C-509T T/T T-29C C/C	1:10-1:20 1: 6-1:16	OR 1.25 OR 1.21	[17]
	BRIP1	17q	Multiple variants	1:1000	RR 2.0	[8]
	PALB 2	16p	Multiple variants	<1:1000	RR 2.3	[9]
	FGFR2	10q	rs2981582	MAF 0.38	Homo. OR 1.63	[10]
	TNRC9	16q 16q	rs12443621 rs8051542	MAF 0.46 MAF 0.44	Homo. OR 1.23 OR1.19	
	MAP3K1	5q	rs889312	MAF 0.28	Homo. OR 1.27	
	LSP1	11p	rs3817198	MAF 0.30	Homo. OR 1.17	

Table 2.1: *List of known breast cancer susceptibility genes. (MAF=minor allele frequency)*

Approximately 5-10% of all breast cancer cases are due to inheritance of a germ-line mutation in a cancer pre-disposition gene that is transmitted as a highly penetrant autosomal dominant trait [18]. Clinically, such cases are typically characterised by young age at presentation, an increased incidence of bilateral breast carcinomas and a family history including non-breast malignancies [19]. The Li Fraumeni, Cowden, Peutz-Jeghers and Hereditary Diffuse Gastric Cancer (HDGC) syndromes are characterised by multiple primary tumours due to inherited mutations of the tumour suppressor genes TP53, PTEN, LKB1/STK11 and CDH1 respectively [13-16]. Lifetime breast cancer risk for carriers of these mutations is in the order of 20-56%, but families with these syndromes are very rare. The two most highly penetrant breast cancer susceptibility genes so far identified are BRCA1 and BRCA2 [12, 20].

2.2 BRCA1 and BRCA2

Initially tracked to chromosome 17q21 in 1990 as a result of gene linkage studies, [20], BRCA1 was subsequently isolated by positional cloning methods, and has been found to contain 24 exons that encode a protein of 1,863 amino acids [21]. A second breast cancer susceptibility gene, BRCA2, was localised to the long arm of chromosome 13 through linkage studies of 15 families with multiple cases of breast cancer that were not linked to BRCA1 [22]. BRCA2 is also a large gene with 27 exons encoding a protein of 3,418 amino acids [12]. Although not homologous genes, both BRCA1 and BRCA2 have an unusually large exon 11 encoding some 60% of the resultant protein and translational start sites in exon 2.

BRCA1 and BRCA2 both appear to act as classical tumour suppressor genes; inheritance of mutations occurs in an autosomal dominant fashion and loss of the unmutated allele is found in tumour specimens. The lifetime risk of breast cancer in carriers of BRCA1 or BRCA2 mutations varies between 37-82% whilst that of ovarian cancer is approximately 54% for BRCA1 mutation carriers and 20% for BRCA2 mutation carriers [19, 23, 24]. It is estimated that some 45% of families affected by multiple episodes of breast cancer have an underlying BRCA1

mutation, whilst up to 90% of families with multiple occurrences of breast and ovarian cancer are a result of this particular genetic anomaly [19]. BRCA2 is thought to account for approximately 35% of families with multiple female breast cancers but is also associated with male breast cancer, ovarian cancer, prostate cancer, and pancreatic cancer [25]. Studies of tumor tissue from sporadic breast cancers have detected no somatic BRCA1 mutations and a very low frequency of BRCA2 mutations [26, 27].

2.3 Functions of BRCA1 and BRCA2

The precise functions of BRCA1 and BRCA2 remain unclear but evidence to date indicates that these proteins are involved primarily in DNA repair and transcription regulation, with additional roles in cell cycle checkpoint control and cytokinesis. Early studies indicated that the mammalian tissue distribution of these proteins is similar to that of the DNA repair protein RAD-51 and both genes encode nuclear proteins that exhibit cell cycle regulated expression and co-localise in multiple tissues during proliferation and differentiation [28]. Cells lacking BRCA1 or BRCA2 accumulate chromosomal abnormalities including chromosomal breaks, severe aneuploidy and centrosome amplification. Analysis of BRCA1 associated breast tumours by comparative genomic hybridisation (CGH) reveals increased loss and gain of a variety of chromosomal regions, particularly 4p, 5q and 12q deletions and gains of 3q, 8q and 10p [29, 30].

Further support for the proposed role of BRCA1 and 2 in DNA damage repair comes from preclinical studies which have shown that reduction of BRCA2 expression in mouse embryonic fibroblasts results in hypersensitivity to ionising radiation [31] whilst increased levels of BRCA1 have been found in cisplatin resistant breast and ovarian cancer cell lines [32]. BRCA2 has also now been identified as one of the nine genes involved in the pathogenesis of Fanconi's Anaemia (FANCD1) [33], with biallelic BRCA2 mutations found in patients with the D1 version of this condition. This rare recessive disease is characterised by abnormal development, bone marrow failure and high risk of malignancy.

Homozygous mutations of BRCA1, as demonstrated in knockout mice, invariably result in embryonic lethality, due to increased cell death and growth retardation [34].

It is now established that cells deficient in BRCA1 or 2 have a defect in the repair of DNA double strand breaks (DSB) by the method of Homologous Recombination (HR), a conservative and potentially error-free mechanism. This results in BRCA1/2 depleted cells relying on DSB repair by the error-prone and potentially mutagenic method of Non-Homologous End Joining (NHEJ) [35-37]. It seems likely that the genomic instability caused by this activity underlies the predisposition towards cancer seen in BRCA1/2 mutation carriers [38].

Recent work suggests that the BRCA1 protein plays an important role in the early response to DNA double strand breaks with phosphorylation by protein kinases such as ATM and CHEK2 inducing its migration to the site of DNA damage [39]. Here it co-localises with other components of the multi-protein BRCA1 associated genome surveillance complex (BASC) [40] and initiates repair by modifying local chromatin structure so that DNA repair proteins can gain access to the damaged material. In particular, BRCA1 seems to regulate the initial processing of DSBs by the MRE11/RAD50/NBS1 (MRN) complex [41]. BRCA1 also complexes with BARD1 (BRCA associated ring domain) to produce E3 ubiquitin ligase activity which may influence the stability and activity of the BRCA1 multiprotein complexes [42]. The BRC repeats of the BRCA2 protein bind RAD51 directly and, following DNA damage and the formation of DSBs, transport it to the site of DNA repair [43]. Here, BRCA2 binds to the MRN processed single stranded 3' overhang, displacing replication protein A (RPA) and facilitating the loading of RAD51. RAD51 subsequently forms a nucleoprotein filament which seeks the homologous sequence of the sister chromatid to use as a template for repair in the HR process known as Gene Conversion [44].

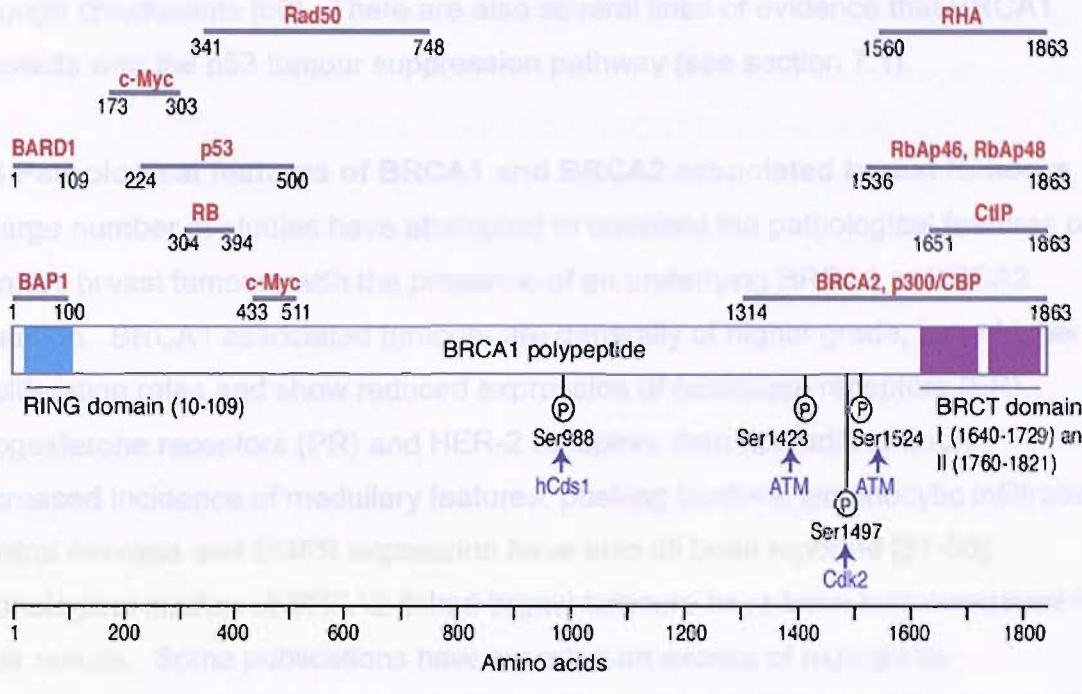


Figure 2.1: *A schematic representation of the BRCA1 polypeptide, and sites of its interaction with other proteins, from Lee and Chung, [45].*

BRCA1 and BRCA2 act as transcriptional co-regulators through direct interaction with sequence specific transcription factors, with evidence suggesting that the BRCA1 gene product is a component of the RNA polymerase II holoenzyme [46]. The amino-terminal of BRCA1 (amino acids 20-68) contains a ring finger domain containing the cys3-his-cys4 motif characteristic of the ring finger family of transcriptional regulatory proteins [11]. This interacts with BARD1 which itself inhibits the activity of a RNA polyadenylation factor (Cst-50), suggesting that BARD1 may be implicated in regulation of RNA processing [47]. Fusion of the C-terminus (amino acids 1528-1863) of BRCA1 to the GAL4 DNA-binding domain activates transcription of the GAL4 dependent promoters in transfected cells [48]. The BRCA1 ring domain also binds to a number of cell cycle proteins, including cyclins and cyclin-dependent kinases [49] and there is increasing evidence that BRCA1 is involved in all phases of the cell cycle, regulating the progression of cells

through checkpoints [50]. There are also several lines of evidence that BRCA1 interacts with the p53 tumour suppression pathway (see section 7.1).

2.4 Pathological features of BRCA1 and BRCA2 associated breast tumours

A large number of studies have attempted to correlate the pathological features of primary breast tumours with the presence of an underlying BRCA1 or BRCA2 mutation. BRCA1 associated tumours are generally of higher grade, have higher proliferation rates and show reduced expression of oestrogen receptors (ER), progesterone receptors (PR) and HER-2 receptors than sporadic tumours. Increased incidence of medullary features, pushing borders, lymphocytic infiltrates, central necrosis and EGFR expression have also all been reported [51-55]. Pathological studies of BRCA2 linked breast tumours have been less consistent in their results. Some publications have reported an excess of high grade carcinomas and there may be an increased number of lobular/ tubular tumours [51, 56, 57]. Whilst several reports suggest that a significantly increased frequency of p53 mutations is seen in breast carcinomas of BRCA1 mutation carriers, no such association has been found for BRCA2 mutation carriers [58-60].

The search for a distinct histological pattern specific to hereditary breast cancer moved forward in 2003. A number of microarray studies of sporadic breast cancers identified a basal epithelial phenotype that was associated with ER/ HER-2 negative tumours and represented no more than 15% of invasive breast cancers [61-63]. This subgroup was characterised by the expression of high levels of the stratified epithelial cytokeratins (e.g. Cytokeratins 5,6,14,15 and 17), which are commonly found in the basal epithelial cells of the normal mammary gland. Foulkes *et al.* and Sorlie *et al.* subsequently demonstrated that an underlying BRCA1 mutation was significantly associated with this basal epithelial phenotype as identified by either positive immunohistochemical evidence of cytokeratin 5/6 [64] or gene expression profile analysis [65]. Analysis by Lakhani *et al.* confirmed that 69.6% of BRCA1 associated tumours were ER negative/ cytokeratin 5/6 +/- 14 positive compared with 8.8% of matched controls [66]. Palacios *et al.* have now

described the use of a microarray system incorporating 37 immunohistochemical markers to examine tissue cores from 20 BRCA1, 14 BRCA2 and 59 age-matched sporadic breast carcinomas [67]. Most of the BRCA1 tissue samples grouped in a collection of ER negative, HER2 negative tumours expressing basal cell markers (e.g. P-cadherin and basal keratins) and/or p53. P-cadherin positivity is a poor prognostic factor in small node negative breast cancers [68]. Higher expression of activated caspase 3 and the cell cycle proteins E2F6, cyclins A, B1 and E, SKP2 and Topo II alpha was also associated with an underlying BRCA1 mutation. In contrast, most BRCA2 tumours were ER/PR and BCL2 positive with higher expression of cell cycle proteins cyclin D1, D3, p27, p16, p21, CDK4, CDK2 and CDK1.

There is now increasing evidence that the basal-like subgroup of sporadic breast carcinomas exhibit so-called "BRCAness" by sharing many of the phenotypical traits described above [69]. This has led to speculation that an acquired deficiency of the BRCA pathway might underlie the sporadic basal-like phenotype and Turner *et al.* have recently reported significantly lower expression of BRCA1 mRNA in 37 basal-like sporadic cancers than age and grade matched controls [70]. This concept may have important implications for the clinical management of such sporadic tumours, which are now often referred to as "triple negative" breast cancers. It is recognised that breast tumours with features of the basal cell-subtype but positive ER and HER-2 status do occur. Thus, although there is a significant overlap between BRCA1 associated tumours, basal-like breast tumours and triple negative breast tumours, these terms cannot be used entirely interchangeably.

Use of the "triple-negative" breast tumour phenotype as an economical and rapid means of identifying patients with underlying BRCA1 mutations is now under investigation and early reports suggest that screening patients with tumours of this phenotype will detect a significantly higher frequency of BRCA1 mutations than predicted by the patient's family history [71]. Although basal-like sporadic breast

cancers are generally accepted to follow an aggressive clinical course, with studies recording inferior survival compared with ER positive controls, [65, 72], the outcome of BRCA1 associated breast cancer is more variable. A number of studies have reported poorer outcomes than sporadic breast cancer cases but several series describe a similar prognosis [54, 55, 73-77]. This could be due to over-representation of medullary tumours in these series, ascertainment bias, or greater efficacy of adjuvant chemotherapy. As all series to date have reported retrospective data, survival bias is a significant problem in interpreting results from these studies. Several series have also relied upon short follow-up periods to draw their conclusions. One recently published population based survival study of Israeli breast cancer patients found no significant difference in ten-year survival between 76 BRCA1 mutation carriers, 52 BRCA2 mutation carriers and 1189 non-carriers [77].

2.5 Management of known BRCA1 and BRCA2 mutation carriers

Regular breast screening by mammography, together with clinical breast examinations, has been recommended for all patients with known BRCA1 or BRCA2 mutations for some years [5, 78, 79]. A number of studies have now demonstrated that yearly magnetic resonance imaging (MRI) offers improved sensitivity for the detection of malignancy in high risk women and detects tumours at an earlier stage than mammography [80-82]. The 2006 National Institute for Health and Clinical Excellence (NICE) guidelines for the management of familial breast cancer therefore include a recommendation for annual MRI surveillance of all BRCA1/2 mutation carriers aged 30-49, (www.NICE.org/CG041). Screening for ovarian cancer remains under investigation.

In terms of primary prevention, most benefit is currently achieved from prophylactic oophorectomy, which reduces breast cancer rates in BRCA1/2 mutation carriers by 60% and ovarian cancer rates by 95% [83], and prophylactic double mastectomy, which can reduce the incidence of breast cancer by up to 90% [84], as shown in figure 2.2. The use of chemoprevention remains controversial. Tamoxifen has

been reported to reduce the risk of contralateral breast cancer by up to 50% in BRC1/2 mutation carriers [85, 86], but data from the NSABP1 primary prevention trial showed no protective effect from this drug, [87]. The IBIS-II International Breast Cancer Intervention Study is currently examining the effect of the aromatase inhibitor anastrazole as a prophylactic agent in patients with a family history of breast cancer in a placebo-controlled randomised trial (www-ibis-trials.org).

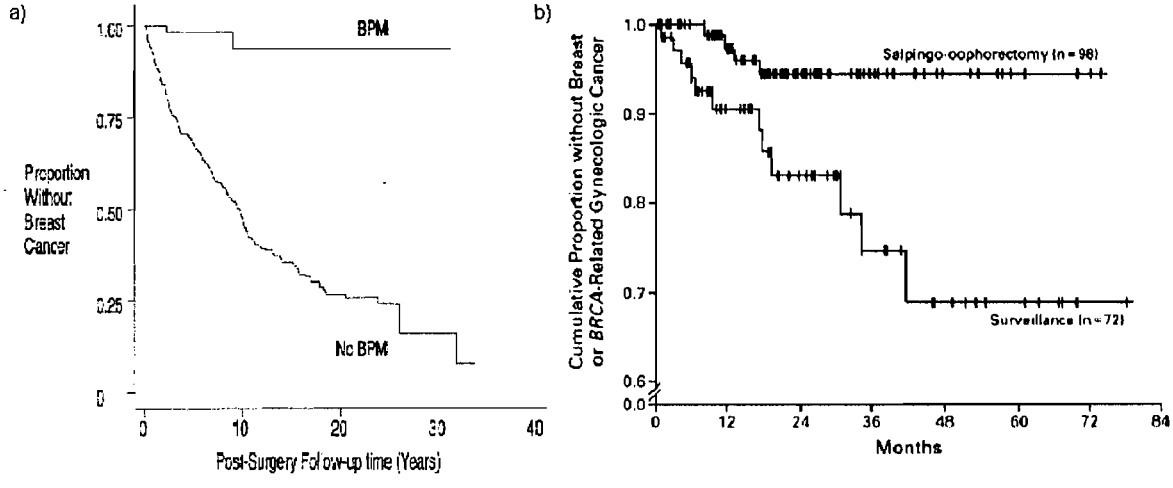


Figure 2.2: *a) Time to breast cancer diagnosis in female BRCA1 mutation carriers with and without bilateral prophylactic mastectomy (BPM), from Rebbeck et al. 2004 [84], b) Kaplan-Meier estimates of the time to breast cancer or BRCA-related gynecologic cancer among women electing risk-reducing salpingo-oophorectomy and women electing surveillance for ovarian cancer, from Kauf et al. 2002 [88].*

There has been considerable discussion in the literature as to whether BRCA1/2 carriers who do develop invasive breast cancer should be offered breast conserving surgery (BCS) because of the potentially higher risk of local disease

recurrence. A recent publication reporting the use of BCS in 160 BRCA mutation carriers describes a 10 year local recurrence rate of 12% [86], slightly higher than that seen in five trials of sporadic cancer cases (8.7%) [89]. Comparisons with age-matched controls however suggest that local recurrence rates of BRCA related breast tumours are similar to those of sporadic cancer, providing the BRCA mutation carriers have undergone prophylactic oophorectomy [86]. Risk of contralateral breast cancer remains significantly higher in BRCA mutation carriers than controls [86]. This increased risk does not however seem to be due to excess radiation toxicity in BRCA mutation carriers, despite the theoretical heightened sensitivity of BRCA1 cells to irradiation, [86, 90].

At present, systemic treatment for breast cancer in BRCA1/2 mutation carriers is identical to that of sporadic breast cancer. The increased proportion of ER negative tumours does however mean that more BRCA mutation carriers receive adjuvant chemotherapy. *In vitro* studies suggest that BRCA associated breast tumours are particularly vulnerable to cytotoxic agents that promote cross-linking of DNA, such as platinums and mitocycin C but show resistance to the mitotic spindle poisons docetaxel and vinorelbine [91-93]. A national study of docetaxel versus carboplatin as first-line therapy for metastatic BRCA related breast cancer is currently under way (www.geneticbreastcancertrial).

The observation that BRCA1/2 mutant cells have an impaired ability to repair double stranded DNA breaks (DSB) has recently led to the recognition that poly (ADP-ribose) polymerase-1 (PARP) inhibitors may offer a selective method of killing tumour cells with underlying BRCA deficiencies. PARP-1 is a central component of the single-strand break (SSB) repair pathway. Disruption of the pathway results in accumulation of SSBs which degenerate to DSBs. In the absence of a functioning BRCA1 mediated repair pathway these result in the collapse of replication forks, growth arrest and apoptosis (reviewed by [94]). Data suggest that mouse embryonic stem cells carrying BRCA1 or BRCA2 mutations show increased sensitivity to PARP-1 inhibitors whilst *in vivo* work indicates that

PARP-1 inhibitors block the growth of BRCA1/2 deficient xenografted tumours, but not wild-type tumours [95,96]. These reports promote the hope that increased understanding of the biology underlying BRCA1/2 mutation associated tumours will permit the development of drugs which specifically attack BRCA deficient tumour cells. Phase I trials of PARP inhibitors are now in progress.

2.6 Diagnosis of BRCA1 and BRCA2 mutations

A number of models exist to calculate the likely risk of an underlying BRCA1/2 mutation given a particular family history but in order to take advantage of the progress being made in the management of families with defective BRCA1 or BRCA2 genes, BRCA mutations have to be correctly diagnosed. Almost 2,000 distinct mutations and sequence variations in BRCA1 and BRCA2 have now been described [97]. Although some mutations are found repeatedly, many are limited to just a few families. DNA analysis is a time consuming process and the large size of both BRCA genes presents a technological and financial challenge to diagnostic molecular genetic laboratories. The gold standard for diagnosis of a BRCA1/2 mutation in a previously untested family is direct sequencing of the entire genomic DNA sequence, but this service is currently offered only by a small number of molecular genetics laboratories. Direct sequencing does not detect large genomic deletions and the efficacy of this method is also limited by the accuracy of the software used to read the sequence traces.

The majority of pathogenic mutations in genes causing hereditary cancer result in the production of an abnormal truncated protein, and are relatively easily recognised. Most European laboratories have therefore traditionally provided a limited screening service focussing on the regions of the gene most commonly found to include pathogenic truncating mutations [98]. Until September 2005 the Wessex Regional Genetics Laboratory (WRGL), for example, performed heteroduplex analysis of exons 2, 20 and the ends of exon 11 of BRCA1 and of exons 10 and 11 of BRCA2 only [99]. Additional information regarding truncating mutations in exon 11 of both genes was provided by the Protein Truncation test

[100]. Multiplex ligation probe amplification (MLPA) was added to these routine tests in 2003 to examine for exonic deletions, which account for up to 5% of pathogenic BRCA1/2 gene abnormalities.

During the course of this project, the WRGL developed the use of conformational sensitive capillary electrophoresis (CSCE) as a rapid method of examining the entire exonic sequences of BRCA1 and BRCA2 for mutations. The SCOBEC (Southampton, Cambridge, Oxford, Bristol, Exeter and Cardiff) high throughput genetic screening facility is now using this method to provide routine screening of the BRCA 1 and 2 genes for patients from a wide geographical area, and systems are in place to ensure that all high risk patients with negative routine BRCA1/2 screen results are re-tested using this method. Inevitably it will take time for this backlog of samples to be processed and it therefore will remain important to take into account the limitations of traditional BRCA1/2 mutation screening for some time to come. It also should be noted that even this method of testing does not identify all intronic sequence variations.

Testing for BRCA1/2 mutations can therefore result in one of three possible outcomes. Firstly, a mutation with definite pathogenic potential may be identified. Such mutations generally consist of deletions, frameshift, or nonsense alterations that are predicted to result in truncation of the encoded protein. Secondly, a so-called negative result may be documented, meaning that no genetic alteration was detected within the areas of BRCA1 and BRCA2 examined. Thirdly, a subtle DNA variant may be detected which would be predicted to either result in only a minor change to the protein (missense mutations), or to not change the protein at all (silent mutations). These are often classified as "DNA sequence variants of unknown significance" and such findings are generally not reported out of the laboratory as the clinical implications of this type of genetic change within BRCA1 or 2 are currently unknown.

2.7 The trouble with unknown variants

Approximately one-third of all BRCA1 mutations and one-half of all BRCA2 mutations notified to the Breast Cancer Information Core (BIC) database fall into the category of unknown variant (UV). This has profound implications for the genetic counselling of affected families as the evidence grows of the benefits that confirmed BRCA1 and BRCA2 mutation carriers can receive from a coherent management strategy. The beneficial effects of prophylactic bilateral mastectomy and/or oophorectomy have been well proven but these two interventions are major surgical procedures with potentially significant psychological sequelae and thus the need for an accurate estimate of lifetime risk of breast and ovarian cancers is highly important in order for the patient to make an informed choice about their management. The incidence of UVs varies significantly between different ethnic groups, occurring ten times more frequently in African-Americans than white American women according to one report [101].

2.8 Methods of analysing unknown variants

Several different approaches have been adopted in attempts to assess the pathogenicity of unknown variants but so far conclusions have only been drawn on a handful. A relatively simple method involves the analysis of patterns of co-segregation of a given mutation with disease in an affected family, with the aim of evaluating the evidence for causality. Peterson *et al.* described a limited Baysian method which did not include data from unaffected individuals or allow for variation in risk as a function of age and sex [102]. A more complex method based on full pedigree likelihood has subsequently been published with evidence that this provides a more accurate assessment of causality than the Peterson model [103].

Comparative analysis of homologous genetic sequences from multiple organisms permits identification of highly conserved sequences that are likely to be of high functional significance. The hypothesis that missense mutations at highly conserved residues are more likely to be deleterious than those at more variable sites was validated by Miller and Kumar in their analysis of 7 genes including the

cystic fibrosis transmembrane conductance regulator (CFTR) [104]. In the largest evolutionary studies of BRCA1 to date, Fleming *et al.* initially examined exon 11 of BRCA1 in 57 divergent eutherian mammalian species with sequences available from GenBank [105]. Seven "highly conserved regions", defined as having greater than 80% homology, were found between amino acids 282-1103. All conserved regions were also present in birds and amphibians. Analysis of the 130 missense changes between a-a 282 and 1103 reported to the BIC database identified thirty-eight missense changes "likely to disrupt gene function; these were so-called because they affected fixed sites, were nonconservative substitutions at conservative sites or affected rapidly evolving or recently evolved residues in humans. A subsequent analysis including 37 of the above 57 ancestral sequences plus an additional 8 marsupial sequences (in order to identify conserved regions unique to mammals) identified 13 highly conserved exon 11 sequences [106]. Four of the eight regions conserved in mammals, birds and amphibians occur in the 5' portion of exon 11 (a-a 282-553), which includes domains involved in transcription. A further 2 of the highly conserved mammalian sequences involve the RAD51 binding region, whilst 7 potentially affect the putative STAT1 interaction domain. This larger analysis reduced the number of missense changes with likely pathogenic potential to 21.

Other studies have also identified the 5' end of exon 11 and the RAD-51 interacting domain as possible sites of positive selection. Hurst and Pal detected a high ratio of non-synonymous to synonymous substitutions at codons 200-300 on comparing BRCA1 sequence variations between human and dog with those between rat and mouse, along with a further hotspot at codons 944-949 [107]. A similar finding of a peak non-synonymous substitution at codon 900 was reported by Huttley *et al.* [108].

More recently Abkevitch *et al.* combined analysis of interspecific sequence variation (performing multiple sequence alignment of 7 orthologous sequences) with measurement of the chemical difference between the amino acids present at

individual residues (using the Grantham chemical difference matrix). This resulted in classification of 50 of the 314 BRCA1 missense changes detected at that time by Myriad Genetic Laboratories as potentially deleterious, whilst a further 92 were deemed to be neutral [109].

The determination of the x-ray crystal structure of the human BRCA1 BRCT domain [110] has permitted assessments of missense mutations based on their effect on protein structure. Williams *et al.* used a protease-based assay to examine the sensitivity of folding of the BRCT domain to a series of truncating mutations and single amino-acid acid substitutions derived from breast cancer screening databases [111]. The BRCA1 BRCT tolerated truncations of up to 8 amino acids but those any greater in length resulted in dramatic folding abnormalities and 14 of the missense mutations tested proved to be incompatible with normal protein folding. Mirkovic *et al.* used 37 BRCA1 mutations with a known impact on transcriptional activation and the three-dimensional structure of the BRCT domain to develop a set of rules for predicting the functional consequences of a given missense mutation [112]. This model was subsequently used to analyse the functional effect of 57 missense mutations; 29 were predicted to reduce transcription activation.

Functional examinations of BRCA1 have also largely been restricted to assessments of mutations within the BRCT domain, the region of the gene that apparently displays intrinsic activity as a transcription activator (see above). Introduction of known disease-causing mutations within the BRCT domain, but not benign polymorphisms, has been shown to abolish this activity [113-115]. Yeast and/ or mammalian assays of transcription activation have now been used to assess approximately 30 different missense mutations [115-117]. Such studies have confirmed that missense mutations with a deleterious effect on transcription activation are scattered throughout the BRCT coding region rather than being clustered in a particular "hot-spot" [115]. Importantly, one report noted discrepancies between results from yeast and mammalian based assays,

suggesting that the simpler yeast–based assay may lack the complex transcription regulatory controls necessary to fully examine the effect of a human missense mutation on transcription [113].

More recently Phelan *et al.* demonstrated that regions adjacent to the BRCT domains contribute to full transcriptional activity, developing a transcription assay capable of analysing variants in exons 13-24 [117]. On testing 7 UVs within the BRCT domain and 6 UVs lying outside they reported significant loss of transcription activation for 7 UVs (similar to that observed with high risk mutation controls) whilst 4 UVs showed transcriptional activity similar to wild type BRCA1 and 2 variants gave intermediate results.

A rather different approach has been reported by Quaresima *et al.* who have specifically examined the effect of BRCA1 missense mutations on protein –protein interactions [118-119]. They initially demonstrated that three missense mutations (Met 1775Arg, Trp1837Arg and Ser1841Asn) within a MLH1 binding domain of BRCA1 in the BRCT domain produced a microsatellite instability phenotype. This was achieved by transfecting HeLa cells with either wildtype BRCA1 cDNA or with BRCA1 cDNA containing one of the above mutations and detecting the presence of genomic instability using eight standard markers. Subsequently, the same group has shown that Trp1837Arg and Ser1841Asn also affect binding activity to p53 [119].

Most of the most recent publications on the analysis of missense mutations have described a combination of the above approaches to characterise BRCA1 missense mutations. In addition to functional assays, Phelan *et al.* used segregation analysis and assessment of interspecific variation to characterise their unknown variants [117]. Analysis of interspecific variation provided conclusive results for 7 UVs; these data matched the results of the functional analysis in 6 cases but contradicted in 1 case. Pedigree analysis (using the Thompson model [103]) matched functional assay results for 3 of the 4 variants analysed. The

results of this transcription activation assay are in accordance with protein characteristic data published by Williams *et al.* and Mirkovic *et al.* for 5 out of 5 and 6 out of 7 variants respectively [111, 112]. Goldgar *et al.* have developed a multifactorial likelihood model using data on co-occurrence of the unclassified variant with pathogenic mutations in the same gene, cosegregation of the UV with affected status, the Grantham Matrix score and evolutionary conservation of the affected amino-acid [120] This model estimates the "odds of causality" (ratio of likelihood of observed data under hypothesis of causality to that under hypothesis of neutrality) and considers UVs with an odds of causality of >1000:1 to be pathogenic whilst those with a score of <1:100 to be neutral. These thresholds are arbitrary. This model has subsequently been both extended to include additional histopathological and immunohistochemical features and loss of heterozygosity of associated breast tumours [121], and has been limited to rely on only on immunohistochemical and pathological data, LOH and familial phenotype [122]. Each version of this model has so far only been used to classify a small number of UVs, but Osorio *et al.* claim that their simple approach is sufficient to classify variants as pathogenic or neutral providing that tumours from three carriers of each variant can be assessed [122].

2.9 RNA; An underused tool in genetic analysis?

Detection of mutations associated with inherited susceptibility to cancer is currently undertaken using techniques that aim to relate alterations in DNA sequence to their effect at the protein level. Production of a protein from DNA cannot however occur without the transcription and translation of an intermediary - messenger RNA (mRNA). Traditionally, mRNA/ cDNA has not been used as a template for mutation screening because of its relative instability and the technical difficulties associated with its' analysis such as the presence of spurious bands. Over the last few years there has been increasing evidence in the literature to indicate that "benign" alterations in the DNA sequence, which would not be predicted to affect protein structure, can interfere with correct gene expression via their effect on the

sequence or stability of the mRNA transcript [123]. This phenomenon can be explained by the complex series of genetic events involved in RNA processing.

2.10 RNA processing

Human RNA, as with that of all higher eukaryotes, requires processing from its initial precursor (pre-) mRNA form into a mature mRNA transcript before it can successfully engage in protein translation. RNA splicing removes non-coding introns and joins adjacent coding regions (exons) together. This complex procedure involves multiple interactions between a variety of regulatory elements. The stepwise splicing reaction occurs in the spliceosome which consists of the pre-mRNA sequence, 5 small nuclear ribonucleoprotein particles (snRNPs U1, U2, U4, 5 and 6) and numerous non-snRNP polypeptides [124]. Accurate identification of exon-intron boundaries by the spliceosome requires recognition of a number of *cis*-consensus sequences in the pre-mRNA: 5' and 3'splice sites, defined by a GU and AG dinucleotide respectively in 95-99% of cases, plus a branchpoint sequence (BPS, usually containing an adenosine residue) and a polypyrimidine tract (PPT) upstream of the 3' splice site. Current models suggest that splicing is initiated by the binding of U1 to the 5' splice site whilst the heterodimeric U2 snRNP (U2AF35) splicing factor binds to the AG sequence at the 3' splice site to generate the early (E) complex. U2AF65 snRNP binds initially to the PPT whilst splicing factor 1 binds to the BPS. Association of U2AF65 and SF1 with U2 result in the formation of a pre-spliceosomal A complex as U2 snRNP is guided to the branch point adenosine. U2AF65, U2AF35 and SF1 subsequently disassociate from the PPT, 3' splice site and BPS respectively and the U4-U6-U5 tri-snRNP join the spliceosome to form the B complex. A series of RNA-RNA and RNA-protein rearrangements result in the formation of the catalytic C complex with the release of U1 and U4 [125-129]. The limited evidence available today suggests that cutting of the RNA molecule is accomplished by protein-assisted RNA catalysis [130].

Although these participants are all necessary, they are not sufficient for accurate splicing to occur, being poorly conserved sequences that are readily recognised

and matched by non-splice site intronic sequences which define pseudo-exons. Additional regulatory cist- elements are required to ensure exact recognition of genuine splice sites and prevent inclusion of pseudo-exons; these are present in the form of exonic and intronic splicing enhancers and silencers (ESEs, ISEs and ESSs, ISSs) which activate and repress splicing respectively. Exonic enhancers consist of binding sites for specific serine/arginine rich (SR) proteins, members of a structurally related and highly conserved family of splicing factors [131]. They are characterised by a modular structure containing 1 or 2 copies of an RNA recognition motif (RRM) that mediates sequence specific binding to RNA, and a carboxyl terminal domain rich in alternating serine and arginine dipeptides (the RS domain) [132]. This region is involved in protein-protein interactions. SR proteins bound to ESEs promote exon definition either by direct recruitment of the U2AF to the PPT (via the RS domain) [133,134], or by antagonising snRNPS recognising local ESIs [135]. Their precise prevalence is unknown but they appear to be present in most, if not all exons, including constitutive ones [136,137].

There is less published information on exonic (or intronic) splicing silencers. They appear to work by interacting with negative regulators, mainly proteins of the heterogeneous nuclear ribonucleoprotein (hnRNP) family such as hnRNP I (better known as poly-pyrimidine-tract-binding protein, PTT) [138-140]. Like SR proteins, hnRNP proteins have a modular structure incorporating one or more RNA binding region and another domain responsible for protein-protein interactions [141].

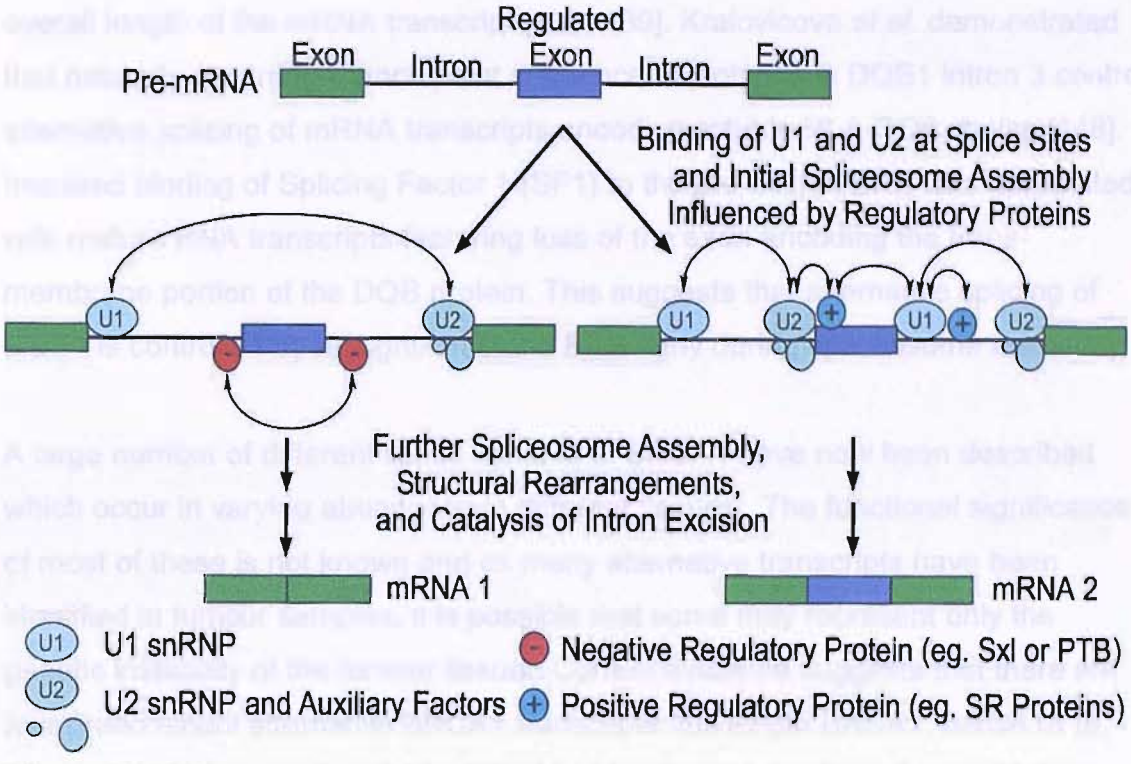


Figure 2.3: *Regulation of pre-mRNA processing (from:* [142]*)*

2.11 Alternative Splicing

This highly complex but flexible control mechanism permits the phenomenon of "alternative splicing", the production of multiple isoforms of mature RNA from each pre-mRNA sequence. The extreme example is that of Drosophilia which is capable of producing some 38016 different mature RNA transcripts from its axon guidance receptor gene [143]. Studies suggest that 35-74% of human genes [144-146] produce more than one transcript. This fact goes some way to explaining the surprising finding of the Human Genome Project that the human genome consists only of 20000-25000 genes [147].

A number of different variations in mRNA structure may occur. Exons can be skipped or introns may be retained and exon length may be altered by adjusting the site of either the 3' or 5' splice site. Furthermore, the position of the transcriptional start site or the polyadenylation signal may be shifted to change the

overall length of the mRNA transcript [138, 139]. Kralovicova *et al.* demonstrated that naturally occurring branch point sequence haplotypes in DQB1 intron 3 control alternative splicing of mRNA transcripts encoding soluble HLA-DQβ chains [148]. Impaired binding of Splicing Factor 1 (SF1) to the pre-mRNA BPS was associated with mature RNA transcripts featuring loss of the exon encoding the trans-membrane portion of the DQB protein. This suggests that alternative splicing of DQB1 is controlled by recognition of the BPS early during spliceosome assembly.

A large number of different splice variants of BRCA1 have now been described which occur in varying abundance in different tissues. The functional significance of most of these is not known and as many alternative transcripts have been identified in tumour samples, it is possible that some may represent only the genetic instability of the tumour tissue. Current evidence suggests that there are four predominant alternative BRCA1 transcripts: full length BRCA1, BRCA1Δ (9, 10) (deletion of exons 9 and 10), BRCA1Δ11q, (partial deletion of exon 11 from c.788-4096) and BRCA1 (Δ9, 10, 11q) (deletion of exons 9, 10 and 11q,) (reviewed by Orban and Olah [149]). All of these transcripts maintain the wildtype open reading frame. Functional studies of the proteins encoded by the Δ11q and Δ9, 10, 11q variants suggest that, despite lacking the exon 11 nuclear localisation sequences, these phosphoproteins can still enter the nucleus, possibly by using p53, BAP1 or full-length BRCA1 as a "shuttle" protein bound to their BRCT domain [150]. These two variants maintain their ability to transactivate the p21 promoter and can still interact with E2F proteins, various cyclins and cyclin dependent kinases and the EKL-1 transcription factor [150, 151].

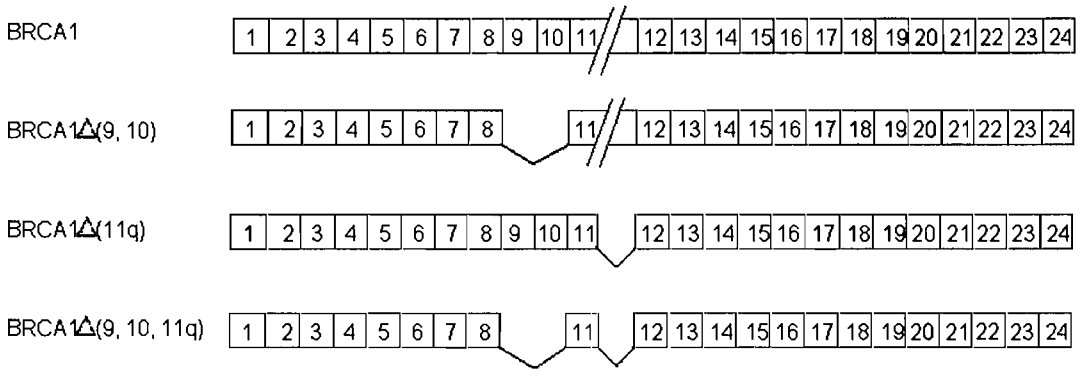


Figure 2.4: *Schematic representations of the four predominant BRCA1 splice variants: full-length BRCA1, Δ (9, 10) with deletion of exons 9 and 10, Δ (11q) with partial deletion of exon 11, and Δ (9, 10, 11q).*

Recently, ElShamy *et al.* have described "BRCA1-IRIS" (in-frame reading of BRCA1 intron-11 splice variant), a splice variant encoding a 1399 amino-acid protein which, unlike the wildtype BRCA1 protein (BRCA1-p220), appears to have a role in DNA replication [152]. The BRCA1-IRIS transcript consists of exons 2-11 with the addition of the first 34 triplets of intron 11. An alternative exon 1 is utilised. The use of a BRCA1-IRIS specific antibody indicated that BRCA1-IRIS was more highly expressed than BRCA1-p220 in spleen and white blood cells and cell-cycle analysis confirmed that, whilst BRCA1-p220 was restricted to cycling cells, BRCA1-IRIS was expressed in quiescent and cycling cells. Cells depleted of BRCA1-IRIS showed reduced DNA synthesis rates. The exact role of BRCA1-IRIS remains under discussion, but the discovery that expression of BRCA1-p220 and BRCA1-IRIS can be uncoupled has led to the suggestion that selective silencing of BRCA1-p220 with continued expression of BRCA1-IRIS may result in enhanced DNA replication rates, and potentially predispose towards tumourigenesis [153].

2.12 The pathogenic potential of alternative splicing

Thus, whilst alternative splicing provides functional diversity, it is also a potential source of disease as aberrant transcripts could result in unstable RNA sequences or deleterious proteins. For example, functional studies of the Δ(11) BRCA1

transcript, which completely lacks exon 11 have shown that this protein cannot interact with Rad 51 and that knockout mice expressing only the Δ(11) variant cannot suppress tumourigenesis [154, 155] .

Krawczak *et al.* determined that approximately 15% of mutations that cause genetic disease affect pre-mRNA splicing [156]. Single base-pair substitutions are by far the commonest cause of human hereditary disease and the majority (84% in the 2003 update) are currently classified as missense or nonsense mutations in the Human Gene Mutation Database (HGMD) [157]. Only 10% of the mutations contained in this data set are categorised as splicing mutations. To date, the effect of DNA point mutations on mRNA expression has only been analysed in a handful of cases, but studies on the Ataxia-Telangiectasia and NeuroFibromatosis genes suggest that approximately 40-50% of disease associated mutations in these two genes affect splicing [158].

The discovery of exonic splicing enhancers and silencers, and particularly their intronic counterparts, has led to the realisation that mutations outside the immediate splicing donor/ acceptor zones may also affect RNA splicing. Thus so-called silent DNA mutations that would not be predicted to have an effect on protein structure may exert an effect on RNA stability or sequence through alterations in splicing regulators. The potential for silent mutations to effect transcription regulation has now gained strong support from a study by Pagini *et al.* They systematically introduced 19 synonymous mutations into exon 12 of the CFTR and found that 25% of these reduced the normal transcript to less than 25% of the normal level, with a potentially pathogenic effect [159].

A survey of 11014 mRNAs from the RefSeq database and some 3.47 million expressed sequence tags resulted in the identification of 26,258 alternative splice variants, 3.2% of which were significantly associated with cancer [160]. Lopez-Bigas *et al.* have developed a mathematical model to examine the importance of splicing mutations in disease [161]. They estimated that the proportion of mutations

affecting splicing over all disease causing mutations is 0.62 ± 0.07%. They also calculated that the average number of introns in disease related genes is 12.4 whilst the average number of introns in the remaining human genes is 9.2 ($p<2.2 \times 10^{-16}$). It is interesting to note that BRCA1 and BRCA2, with 22 and 23 exons respectively are both considerably larger than the average human gene and are therefore potentially at a higher risk of splicing errors.

There is also increasing evidence that coding single nucleotide polymorphisms (cSNP) previously classified as "silent" can directly affect splicing regulation by disrupting or creating exonic splicing enhancers and/ or silencers. Following a recent survey of 106 human genes that estimated the frequency of cSNPs to be 1 in 346bp [162], it has been calculated that 50% of exons may include at least 1 cSNP [129]. It is feasible that the influence of cSNPs on splicing could contribute to the phenotypic variation within a population and to the variable penetrance of mutation. cSNPs in cancer predisposition genes could be responsible for a low penetrance increased cancer risk. Again, large genes such as BRCA1 and BRCA2 are theoretically more at risk of SNP mediated effects on splicing.

Several bioinformatic systems have been designed to predict the effect of sequence changes on splicing, particularly with respect to mutations in ESEs and ESSs. These include ESEfinder and Rescue-ESE [163-164]. Pettigrew *et al.* used ESE-finder to demonstrate that 61% (14/ 23) of putative BRCA1 ESEs conserved or shared between human and mouse would be affected by BIC classified UVs [165]. However, these approaches are theoretical in nature and are not sufficiently reliable for clinical diagnostic purposes. The definitive analysis of splicing abnormalities requires direct analysis of mRNA.

Several groups have used standard RT-PCR approaches to identify aberrant transcripts arising from alterations to nucleotides within 3 bases of the exon/ intron boundary which would be predicted to disrupt normal splicing by directly changing highly conserved splice donor or acceptor sites [166]. Brose *et al.* reported that

IVS 17+1G>T (ATGn 5074 +1) results in inclusion of part of intron 17 after exon 17 and 252del5insT (ATGn133) produces skipping of exon 5 via disruption of the exon 3 3'splice site [167]. Similarly, aberrant splicing associated with the sequence variants, IVS22-2 A>G, IVS 4-1G>T, IVS9-2A>C, IVS20+1G>A, Ala1693del (ATGn5077del3), IVS10-2A>C has been described by Campos *et al.*, Tesoriero *et al.* and Keaton *et al.* [168-170]. Sequencing of the aberrant transcripts for each of these six variants confirmed skipping of the exon adjacent to the altered splice site. In all cases except IVS4-1G>T and Ala1693del the resultant mRNAs were predicted to carry a frameshift resulting in a truncated protein.

Very few BRCA1 missense mutations located outside the splice site consensus sequences have been analysed in terms of their effect on the cDNA transcript. Sharp *et al.* reported an abnormal transcript associated with the c736 T>G missense mutation, Yang *et al.* found a partial deletion of exon 5 associated with the missense mutation 309 T>G (c190) [171,172]. More data on the influence of missense mutations on splicing regulation comes from investigation into other genes. The disruption of an ESE in exon 3 of MLHI by a missense mutation found in a Quebec family with HNPCC was reported by McVety *et al.* whilst a systematic analysis of mRNA from subjects carrying 60 distinct missense mutations of MLH1 revealed that 5 of the 15 exonic mutations produced aberrant transcripts; 4 of these mutations were however situated within 3 nucleotides of the exon boundary [173, 174]. Pagani *et al.* described two missense mutations of the CFTR gene which induce skipping of exon 12 in a significant proportion of the resultant transcripts [159]. Interestingly, these mutations were associated in several cases with non-classical phenotypes of cystic fibrosis.

Although the web-based bioinformatic systems such as ESEfinder and rescue-ESE have been designed to predict the effect of DNA sequence variations on potential ESEs and ESSs [163,164], there is increasing evidence to suggest that there are significant discrepancies between predicted ESEs and actual splicing regulating regions [173-175]. Work by Auclair to analyse the correlation between aberrant

splicing and prediction of ESE by ESE finder produced a sensitivity of 80% and a specificity of 42% [174]; thus this type of approach is theoretical in nature and currently not sufficiently reliable for clinical diagnostic purposes.

2.13 BRCA mutations and splicing

There are now a small amount of published data reporting the effect of BRCA 1 and BRCA2 mutations on splicing. Liu *et al.* were the first group to describe the disruption of an ESE in BRCA2, by the nonsense mutation GLu1694Ter in exon 18. This caused inappropriate skipping of the constitutive exon 18 [123]. Similarly the BRCA2 variant 8204G>A has been found to cause skipping of exon 17 [176]. The largest study to date of missense mutations describes the RNA analysis of mRNA for 4 BRCA1 UVs and 2 BRCA2 UVs; three of the studied variants (IVS22-2 A>G, IVS 18+ 5G>A and Ala1693del (c5077del3) were shown to affect splicing [168]. A further 7 BRCA1 missense mutations resulting in aberrant splicing regulation have since been identified: IVS 17+1G>T (c5074 +1), 252del5insT (c133), IVS 4-1G>T, IVS9-2A>C, IVS20+1G>A, 309 T>G (c190) and IVS10-2A>C [166-170, 172]. Sequencing of the aberrant transcripts for each of these six variants confirmed skipping of the exon adjacent to the altered splice site. In all cases except IVS4-1G>T and Ala1693del the resultant mRNAs were predicted to carry a frameshift resulting in a truncated protein.

Recent work within the Wessex Regional Genetics Laboratory used the method of allele-specific RT-PCR to study the quantity and sequence integrity of the RNA product of hereditary cancer genes in 12 patients who were previously known to have a DNA sequence variant of unknown significance in a cancer predisposition gene (BRCA1 in three cases) [171]. In 4 of the 12 cases the missense alterations were associated with disruption of the RNA sequence due to splicing errors. Additionally, one patient was found to have monoallelic expression of BRCA1, as determined by polymorphism analysis. Monoallelic expression of a cancer predisposition gene is clear evidence of a pathogenetic abnormality, such as a

mutation in a promoter/ enhancer element or as a consequence of nonsense-mediated decay.

2.14 Nonsense Mediated Decay (NMD)

This is the mechanism by which RNA transcripts containing premature stop codons (PTCs) as a result of truncating DNA mutations are specifically degraded, a process designed to remove aberrant mRNA sequences. Current evidence suggests that NMD is triggered by the persistence of the exon-junction complex (EJC), a multiprotein complex including the nuclear proteins UPF2 and UPF3, deposited 20 nucleotides upstream from each exon-exon boundary by the spliceosome. During translation of "healthy" RNA, the EJCs are displaced by elongating ribosomes and the mRNA is protected from NMD. If the translating ribosomes encounter a stop codon upstream of an EJC, UPF1 is recruited by translation release factors and interacts with the UPF2 and UPF3 proteins. This initiates a cascade of cellular events resulting in destruction of the aberrant RNA transcript. A further phenomenon by which point mutations introducing a PTC have resulted in exon skipping has recently been described: nonsense –associated altered splicing (NAS) [177-178]. The exact mechanism by which this occurs has yet to be fully elucidated.

2.15 NMD as a diagnostic marker

A study of the relative amount of transcripts encoded by BRCA1 alleles harbouring 30 different truncating mutations revealed that NMD was triggered by 80% of PTC positive alleles and resulted in a 1.5-5 fold reduction in RNA abundance [179]. Tournier *et al.* described the use of RT-PCR, single nucleotide primer extension and denaturing high performance liquid chromatography to measure allele-specific MLH1 expression in mRNA of 40 heterozygous controls and 12 patients with known mutations of MLH1 that could result in NMD [180]. Whilst their control group showed a narrow range of biallelic expression of MLH1 (44.7-55.3%), in 5 out of 8 patients with known truncating mutations significantly reduced levels of one allele were detected. Such techniques may therefore provide a sensitive tool capable of

identifying genetic defects that escape detection in genomic DNA screening but result in a quantitative change at the mRNA level.

2.16 Disease modifying genes

In addition to the problem of diagnosing pathogenic mutations of BRCA1 and BRCA2, there is significant interest in the identification of alternative genetic profiles that may influence the course of disease in a BRCA mutation carrier. Inheritance of a BRCA1 truncating mutation carries a lifetime risk of breast cancer of between 50 and 80% with age at onset of first malignancy varying from 2nd to the 8th decade [19, 23]. No differences in penetrance have yet been documented for different BRCA1 truncating mutations [181]. Much of the published variation in lifetime cancer risk in BRCA1 mutation carriers may be accounted for by differences in the study populations and in the methods used to calculate risk. However inter-individual variation in the risk of developing cancer may also be affected by co-inheritance of other disease modifying genes. Plausible candidates for genetic modifiers are genes that interact with the BRCA1 cellular pathway. Identification of disease modifiers would facilitate clinical risk management of patients with diagnosed BRCA1 mutations.

It has been noted that BRCA1 breast tumours have a higher frequency of p53 mutations than sporadic breast tumours and several lines of evidence now point to a role for the tumour suppressor p53 in the development of BRCA1 cancers [56, 67, 182-184]. Mice carrying homozygous mutations of BRCA1 fail to develop beyond an early embryonic stage but have their survival prolonged by the co-inheritance of p53 mutations [185, 186]. tBRCA1-/- mice carry a targeted null mutation of BRCA1 restricted to the T-cell lineage resulting in depletion of thymocytes and peripheral T-cells, and activation of p53 [187]. Thymocyte development in these mice is restored in the absence of p53. Such observations have led to the hypothesis that truncating mutations in BRCA1 lead to accumulation of DNA damage and activation of cell cycle checkpoints including activation of p53, with the subsequent suggestion that disruption of the p53

pathway may be required before a BRCA1 mutation results in tumour development [18].

The co-inheritance of alterations in the p53 tumour suppressor pathway would therefore be expected to modify the clinical effects of a BRCA1 mutation. Checkpoint kinase 2 (Chk2) is a mediator of cell cycle arrest and apoptosis induced by DNA damage. Pre-clinical studies by McPherson *et al.* have provided evidence of the role Chk2 plays in the accumulation and activation of p53 in BRCA1 deficient mice [188]. A deficiency of Chk2 partially rescued the defective development and growth of BRCA1 deficient T-cells but enhanced tumour development.

MDM2 encodes a protein which directly binds to and inhibits p53 [189] but also has p53 independent negative effects on DNA double strand break repair [190, 191]. MDM2 gene amplification is found in a number of tumour types, including sporadic breast cancers, [192] and overexpression of MDM2 is associated with accelerated tumour development and failure to respond to treatment [193]. Bond *et al,* have described a T to G exchange at nucleotide 309 within the promoter region of MDM2, which increases the affinity of the transcriptional activator Sp1 and results in higher levels of MDM2 protein with subsequent attenuation of the p53 tumour suppression pathway [194]. Their study of 88 individuals with inherited p53 mutations demonstrated that carriers of this SNP in either its' homozygous or heterozygous form developed cancer on average 7 years before their wildtype counterparts. Additionally, individuals homozygous for SNP309 but with no known hereditary cancer predisposition showed a significantly earlier age of onset of sporadic soft tissue sarcoma than those with the T/T genotype. Given the evidence for interaction between the p53 stress response and the BRCA1 cellular pathway, MDM2 could be a plausible disease modifying gene in BRCA1 mutation carriers.

2. 17 Aims of this research project

This research project aimed to extend the above pilot project, using the technique of RT-PCR to investigate the effect on RNA of missense DNA alterations of BRCA 1 and BRCA2 in up to 50 patients with previously reported DNA variations of unknown significance. As pathological studies are unable to identify an underlying BRCA mutation with complete certainty and routine screening methods available at the time did not include the whole of the BRCA genes, it was intended that the RNA of patients with known truncating mutations would be analysed by rapid throughput Pyrosequencer™ technology (see section 4.1) to look for alterations in allele frequency at single nucleotide polymorphism (SNP) sites that are consistent with nonsense-mediated mRNA decay. Such changes in mRNA could potentially act as easily detectable indicators of an underlying DNA truncating mutation. RNA from patients with positive family histories but previous negative (limited) BRCA1/2 screens were also analysed to seek previously undetected genetic abnormalities.

It was anticipated that this study would detect new pathogenic mutations in cancer predisposition genes that were previously classified as benign sequence variants by current DNA screening methods, potentially improving the detection of families with a high lifetime risk of cancer. In addition, whatever the cause of the abnormal RNA transcript, any variation in a gene that allows one allele to be distinguished from another is a potentially useful flag for predicting the presence of a protein truncating mutation. If the use of Pyrosequencing™ technology proved to be a robust method of analysing cDNA then it could provide a rapid throughput method of genetic analysis that could potentially lead to the development of novel diagnostic tests which would be applicable to many genetic disorders.

In addition, this project aimed to identify and describe splicing abnormalities in BRCA1 using a series of multiplex RT-PCRs to cover the entire coding regions of the gene. The cDNA products were analysed by gel electrophoresis and any PCR products of unpredicted molecular weight were identified as alternative transcripts. Again, subjects included patients with known truncating mutations, patients with

previously identified missense mutations and those with family histories strongly suggestive of a BRCA1/2 mutation but negative limited BRCA1/2 screens. As well as identifying the potentially pathogenic missense mutations, this analysis could also provide information about splicing regulation sites in BRCA1 and BRCA2. The introduction by the Wessex Regional Genetics Laboratory of CSCE as a high throughput method of detecting BRCA1/2 exonic mutations subsequently permitted the identification of further subjects with missense mutations suitable for further investigation of the impact of missense mutations on splicing regulation.

As the project progressed it became apparent that a lack of availability of RNA from patients with known BRCA1 missense mutations would significantly limit this investigation into the effects of these sequence variants on splicing of BRCA1. The initial plan to analyse RNA to find mutations which would be missed on routine limited screening of genomic DNA also effectively became overtaken by the advent of high throughput methods capable of screening all exons of BRCA1 DNA within a few days, such as CSCE. These methods are likely to identify more missense mutations and it therefore became appropriate to adopt an alternative technique to investigate the effect of reported BRCA1 missense mutations on splicing. The latter part of this project therefore focused on the development of a BRCA1 minigene system in which previously reported BRCA1 missense mutations could be artificially recreated. Transfection of these mutated minigenes into cultured cells subsequently permitted examination of the effects of these sequence changes on splicing regulation via analysis of the various transcripts produced.

In addition I used the Pyrosequencing™ technique to investigate the effect of the MDM2 SNP309 on clinical outcome in a cohort of patients with germline mutations of BRCA1 with established pathogenic potential.

3.0 Materials and Methods

3.1 General materials

3.1.1 Chemical and laboratory reagents

Acrylamide (40% stock)	Fisher Scientific, Manchester, UK.
Agar	Fisher Scientific, Manchester, UK.
Agarose – multipurpose molecular grade	Bioline, London, UK.
Ammonium persulfate	Sigma-Aldrich Ltd. Poole, UK.
Ampicillin	Sigma-Aldrich Ltd. Poole, UK.
Bovine serum albumin (BSA)	Promega, Madison, USA.
Chloroform (99%)	Sigma-Aldrich Ltd. Poole, UK.
Dinucleotriphosphates (dNTPs), (100mM)	Invitrogen Ltd, Paisley, UK.
Dithiothreitol (DTT), (0.1M)	Invitrogen Ltd, Paisley, UK.
Dulbecco's modified eagle's medium	Sigma-Aldrich Ltd. Poole, UK.
EDTA	Fisher Scientific, Manchester, UK.
Ethidium bromide solution (10mg/ml)	Promega, Madison, USA.
Ethanol (99+%)	BDH Laboratory Supplies, Poole, UK.
Foetal bovine serum	Sigma-Aldrich Ltd. Poole, UK.
Formamide	Fisher Scientific, Manchester, UK.
IPTG, dioxane-free	Promega, Madison, USA.
Isopropranol (99%)	Sigma-Aldrich Ltd. Poole, UK.
Histopaque	Sigma-Aldrich Ltd. Poole, UK.
LB Broth, Lennox	Fisher Scientific, Manchester, UK.
Magnesium chloride (25mM)	Promega, Madison, USA.
Mineral oil	Sigma-Aldrich Ltd. Poole, UK.
Penicillin/ Streptomycin	PAA Labs., Pasching, Austria.
Phenol-chloroform (50%:48%)	Sigma-Aldrich Ltd. Poole, UK.
Phosphate buffered saline tablets	Sigma-Aldrich Ltd. Poole, UK.
2-Propanol (99%)	Sigma-Aldrich Ltd. Poole, UK.
Proteinase K	Sigma-Aldrich Ltd. Poole, UK.
Sodium chloride (5M)	Sigma-Aldrich Ltd. Poole, UK.
Sodium dodecyl sulphate	Sigma-Aldrich Ltd. Poole, UK.

Sodium lauryl sulphate	Sigma-Aldrich Ltd. Poole, UK.
Temed (N,N,N',N',-Tetramethylethylenediamine)	Sigma-Aldrich Ltd. Poole, UK.
Triton-X100	Fisher Scientific, Manchester, UK.
Tris base	Fisher Scientific, Manchester, UK.
Tris-acetate-EDTA buffer (TAE) (50x)	Promega, Madison, USA.
Tris-borate-EDTA buffer (TBE) (10x)	Promega, Madison, USA.
Water - ultra PURE distilled	Invitrogen Ltd. Paisley, UK.
Water – nuclease free	Promega, Madison, USA.
Tris-acetate-EDTA buffer (10x)	Sigma-Aldrich Ltd. Poole, UK.
TRI®zol	Invitrogen Ltd. Paisley, UK.
TRI® Reagent	Ambion, Austin, USA.
X-Gal (5bromo-4chloro-3indolyl-β-D-galactopyranoside)	Promega, Madison, USA.

3.1.2 Enzymes

All restriction enzymes were obtained from Promega (Madison, USA). RQ RNase-free DNase was obtained from Promega and TURBO-DNase was obtained from Ambion, (Austin, USA). SuperScript II and SuperScript III First Strand Synthesising systems for RT-PCR were obtained from Invitrogen (Paisley, UK). M-MLV Reverse Transcriptase was obtained from Promega (Madison, USA).

3.1.3 DNA molecular markers

The following molecular markers were used:

Benchtop 1kb ladder Promega (Madison, USA)
(250bp, 500bp, 750bp, 1000bp, 1500bp and 2000-10,000bp in 1000bp increments)

100bp ladder (1.0µg/ml) Promega (Madison, USA)
(100bp – 1.5kb in 100bp increments, 2027bp)

50 bp ladder (1.0µg/ml) Pharmacia (Pfizer, New York, USA)
(50bp-1000bp in 50bp increments)

3.2 Clinical samples

All patient blood samples were obtained from patients who had previously attended appointments with the Wessex Clinical Genetics Service (WCGS) for genetic screening for BRCA1/2 mutations. Ethical approval was obtained from Southampton Local Research Ethics Committee prior to the commencement of this research and informed consent was obtained from all subjects. Following review of the results of routine BRCA1/2 mutation screening (performed by the Wessex Regional Genetics Laboratory), patients were classified as having a known truncating mutation of BRCA1 or BRCA2, a DNA variation of unknown significance of BRCA1 or BRCA2, or as having no known genetic alteration in BRCA1/2. The blood samples were categorised appropriately and then encoded. Anonymous genomic DNA samples and linked RNA samples stored within WRGL and referred for screening of non-neoplastic conditions were obtained to provide a control group.

3.3 *General methods*

3.3.1 Preparation of nucleic acids

3.3.1.1 DNA extraction from whole blood

The 5ml anticoagulated blood sarnple was first chilled on ice. The blood sample was then poured into a sterile 50 ml tube and 20 ml sterile distilled water was added to it. The sample was centrifuged at 2000g for 10 minutes at 4°C in a Sorvall RT 6000B centrifuge. The supernatant was discarded and the cell pellet was resuspended in 10 ml cold sucrose lysis buffer (0.32M sucrose, 10mM Tris base, 5mM magnesium chloride, 1% Triton; pH adjusted to 7.5 with concentrated hydrochloric acid). The sample was centrifuged at 2000g for 10 minutes at 4°C and the supernatant was discarded. One point five ml Resuspension Buffer (0.075 M sodium chloride and 0.024 M EDTA), 75 µl 10% sodium lauryl sulphate and 8 µl proteinase K (50 mg/ml) were added to the pellet. The tube was then agitated vigorously to resuspend the pellet and then incubated for 16 hours at 37°C. Seven hundred and fifty µl 6M sodium chloride solution was added and the sample was shaken for 20 seconds. The sample was then centrifuged at 2500g for 30 minutes at room temperature. The supernatant was collected in a 15 ml tube and

7.5 ml 100% ethanol was added. The sample was inverted until a mass of DNA was seen to form. The DNA sample was then washed in 1 ml of 70% ethanol. The DNA was stored in 1.5 ml of 1 x Tris EDTA at -20°C.

3.3.1.2 Isolation of peripheral blood mononuclear cells for RNA extraction

Five ml of whole anticoagulated blood was layered over 10 ml Histopaque in a sterile 50 ml tube. The sample was centrifuged at 2000g for 20 minutes at room temperature in a Sorvall RT 6000B centrifuge. The layer containing mononuclear cells (MNCs) was collected and transferred to a sterile 15 ml centrifuge tube. The sample was centrifuged at 500g for 5 minutes at room temperature and the supernatant was discarded. Two ml of PBS was added and the sample was centrifuged at 500g again for 5 minutes. The supernatant was carefully removed from the cell pellet and 750 µl of TRI®zol was added to it. The TRI®zol was gently mixed with the pellet and then transferred to a sterile DNA tube containing a further 750 µl of TRI®zol. The sample was stored at -80°C.

3.3.1.3 RNA extraction from peripheral blood mononuclear cells (MNCs): manual chloroform/ phenol precipitation technique

The frozen lysate of MNCs stored in TRI®zol was thawed at room temperature for one hour. Three hundred and fifty µl of the TRI®zol lysate was transferred to a clean 1.5 ml microcentrifuge tube and a further 650 µl TRI®zol was added. The sample was mixed by inversion and then centrifuged at 15000g in a Micromax RF (IEC) centrifuge for 10 minutes at room temperature. Nine hundred µl of the supernatant was transferred to a clean microcentrifuge tube and 300 µl chloroform was added. The sample was vortexed for 15 seconds and then left for 2 minutes at room temperature. The sample was centrifuged at 15000g for 15 minutes at 4°C and the upper aqueous phase was then transferred to a further clean microcentrifuge tube. Seven hundred and fifty µl of isopropanol was added to this and the sample was mixed by inversion. The tube was incubated at -20°C for 20 minutes and then centrifuged at 15000g for 10 minutes at 4°C. The supernatant was discarded and the remaining RNA pellet was washed with 1 ml of 75%

ethanol. The sample was vortexed for 3 seconds and then centrifuged at 15000g for 5 minutes at 4°C. Nine hundred µl of the supernatant was discarded and the sample was centrifuged at 15000g, 4°C for 10 seconds. The remainder of the supernatant was removed and the RNA pellet was air dryed for 10 minutes at room temperature. The pellet was then resuspended in 30 µl of sterile water and stored at -80°C.

3.3.1.4 RNA extraction: RNeasy® Column Method (Qiagen)

RNeasy® columns (Qiagen Ltd. Crawley, UK) were used according to the manufacturer's instructions. The frozen lysate of MNCs stored in TRI®zol was thawed at room temperature for one hour. Three hundred and fifty µl of the TRI®zol lysate was transferred to a clean 1.5 ml microcentrifuge tube containing 350 µl chloroform/ phenol. The sample was mixed by inversion and then centrifuged at 13000g for 5 minutes at room temperature. The supernatant was transferred to a further microcentrifuge tube containing 350 µl chloroform/ phenol and the sample was mixed by inversion. The sample was centrifuged for 10 minutes at 13000g and the upper aqueous layer was transferred to a microcentrifuge tube containing 350 µl of 70% ethanol. The sample was mixed and then applied to a RNeasy® column placed within a collecting tube. The RNeasy® column was left to stand for 30 minutes. The column and collecting tube was centrifuged for 15 seconds at 7000g and the flow through was discarded. Six hundred and fifty µl of RW1 solution was applied to the column. The column and collecting tube was centrifuged for 15 seconds at 7000g and the flow through was discarded. 500 µl RPE solution was applied to the column, the column and collecting tube was centrifuged at 7000g for 15 seconds and the flow through was discarded. A further 500 µl RPE solution was applied to the column and the column and collecting tube was centrifuged at 13000g for 2 minutes. The flow through was discarded and the spin column was removed from the collecting tube and left to dry for 2 minutes. The spin column was then placed in a fresh 1.5 ml microcentrifuge tube, the column cap was opened and the column was left for 10 minutes. Twenty-five µl sterile water was applied to the centre of the columnn and

the column was centrifuged at 13000g for 1 minute to elute the RNA. The RNA was subsequently stored at -80°C.

3.3.1.5 Extraction of RNA from whole blood collected into PAXgene® Blood RNA Tubes (Qiagen)

Isolation of RNA was performed using the reagents contained within the PAXgene® Blood RNA Kit (Qiagen Ltd. Crawley, UK) according to the manufacturer's instructions. To achieve complete cell lysis, the PAXgene® Blood RNA tube was incubated for a minimum of 2 hours at room temperature prior to further processing. The PAXgene® Tube was centrifuged for 10 minutes at 3000g in a Centra CL centrifuge (Thermo IEC). The supernatant was discarded and 5 ml RNase free water was added to the residual pellet. The tube was then closed using a fresh Hemogard closure. The pellet was resuspended by vortexing and centrifuged for 10 minutes at 3000g. The supernatant was again discarded and 360 µl Buffer BR1 was added. The tube was vortexed until the pellet had visibly dissolved and the sample was pipetted into a 1.5 ml microcentrifuge tube. Three hundred µl Buffer BR2 and 40 µl Proteinase K were added and mixed with the sample by vortexing. The sample was then incubated for 10 minutes at 55°c in a water bath. The sample was vortexed for 30 seconds and centrifuged for 10 minutes at 10000g. The supernatant was transferred to a clean 1.5 ml microcentrifuge tube and 350 µl 100% ethanol was added. Seven hundred µl of the sample was applied to a PAXgene® spin column placed within a collecting tube. The column was centrifuged for 1 minute at 10000g and the flowthrough was discarded. The remaining sample was then applied to the spin column and the column was centrifuged for 1 minute at 10000g. The spin column was transferred to a new collecting tube and the flowthrough was discarded. Seven hundred µl of BR3 buffer was added to the spin column and it was centrifuged for 1 minute at 10000g. The column was transferred to a new collecting tube and the flowthrough was discarded. Five hundred µl of BR4 buffer was applied to the spin column and it was centrifuged for 1 minute at 10000g. The column was transferred to a new collecting tube and the flowthrough was discarded. A further 500 µl of BR4 buffer

was applied to the spin column and it was centrifuged for 3 minutes at maximum speed to dry the column member. The column was transferred to a 1.5 ml elution tube and 40 µl of BR5 buffer was added directly onto the spin column membrane. The column was centrifuged for 1 minute at 10000g to elute the RNA. A further 40 µl of buffer BR5 was added onto the spin column membrane and the column was again centrifuged for 1 minute at 10000g to elute further RNA. The elute was incubated for 5 minutes at 65°C in a water bath and then chilled on ice.

3.3.1.6 Extraction of RNA from cell culture preparations; chloroform/ propanol precipitation technique

One hundred µl chloroform was added to each 0.5 ml volume of TRI® Reagent (Ambion, Austin, USA)/ cell lysate in a microcentrifuge tube and the tubes were vigorously shaken for 15 seconds. They were then centrifuged at 10500g for 15 minutes at room temperature. The supernatants were carefully transferred to clean microcentrifuge tubes and 250 µl propanol was added to each. The tubes were vortexed for 10 seconds and then incubated at room temperature for 10 minutes. After centrifugation at 10500g for 8 minutes at 4°C the propanol was removed from the RNA pellets by gentle aspiration. Each RNA pellet was washed with 500 µl 75% ethanol before being centrifuged again at 10500g for 5 minutes at 4°C. The ethanol was carefully removed and the RNA pellets were allowed to air dry at room temperature for 5 minutes. Each RNA pellet was dissolved in 30 µl nuclease free water (Promega, Madison, USA). The RNA samples were subsequently kept on ice for immediate use or stored at -20°C.

3.3.1.7 Measurement of nucleic acid concentrations

The concentration of all DNA and RNA samples was measured by analysing a 1.5 µl sample on a NanoDrop® ND-1000 Spectrophotomotomotor, (NanoDrop Technologies, Inc. Wilmington, USA), according to the manufacturer's instructions.

3.3.2 cDNA Synthesis

3.3.2.1 cDNA synthesis: allele specific RT-PCR

Sequential primers were designed to amplify the cancer predisposition genes APC, BRCA1, BRCA2, MLH1, MSH2 (see Appendix C). A RT primer mix containing the reverse-strand primer of each of the above genes at a concentration of 33 ng/µl was prepared by adding 5 µl of each reverse-strand primer stock solution (660ng/µl) to 35 µl of sterile water. Two µl RQ1 RNase-Free DNase, 4 µl 5x RT buffer, 0.2 µl RNase out and 1.8 µl sterile water were added to 10 µl RNA. The sample was incubated for 30 minutes at 37°C. Two µl DNase stop buffer was then added and the sample was heated at 65°C for 10 minutes to inactivate the DNase. Four µl of the sample was removed at this point to act as a negative control. To the remaining sample was added 2 µl of the RT primer mix. The sample was centrifuged at 10000g for 5 seconds and 1 drop of mineral oil was overlaid onto the sample. The sample was incubated at 65°C for 5 minutes and then chilled on ice. Eleven point five µl of a reaction mix containing 3 µl 100mM DTT, 2 µl 10mM dNTPs, 0.5 µl RNase OUT, 3 µl 5x RT buffer and 3 µl sterile water was added to each sample and the reactants were mixed thoroughly by pipetting. The sample was incubated at 42°C for 2 minutes and 1.5 µl Superscript II reverse transcriptase was then added. The sample was incubated at 42°C for 60 minutes and then heated to 70°C for 15 minutes before being cooled to 4°C. The cDNA sample was subsequently stored at -20°C.

3.3.2.2 cDNA synthesis: RT-PCR with random hexamers using Superscript First Strand Synthesis System (Invitrogen)

The Invitrogen SuperScript First Strand Synthesis System for RT-PCR Invitrogen Ltd. Paisley, UK), was used in accordance with the manufacturer's recommendations. Eight µl of RNA was added to 1 µl random hexamers (50ng/µl) and 1 µl 10 mM dNTPs in a 0.5ml tube. This was then incubated at 65°C in a hot block for 5 minutes before being chilled on ice for 5 minutes. Ten µl of a cDNA mix containing 2 µl 10x RT buffer, 4 µl 25 mM magnesium chloride, 2 µl 0.1 M DTT, 1 µl RNaseOUT (40 U/µl) and 1 µl Superscript III (200U/µl), was added to the RNA

mix. The reactants were mixed thoroughly by pipetting and transferred to a sterile 0.2 ml tube. The sample was then sequentially incubated at 25°C for 10 minutes, 50°C for 50 minutes, 85°C for 5 minutes and then cooled to 4°C for 20 minutes in a Peltier 225 thermal cycler (MJ research, Bio-Rad Laboratories, Massachusetts, USA). One µl RNaseH was added to the cDNA and it was incubated at 37°C for 20 minutes in a hot block. The cDNA sample was subsequently stored at -20°C.

3.3.2.3 cDNA synthesis: RT-PCR with oligonucleotides using M-MLV Reverse Transcriptase (Promega, Madison, USA)

The concentration of the RNA sample was adjusted to approximately 180ng/ µl. 16.5 µl RNA was added to 1 µl TurboDNAse (Ambion, Austin, USA), 2 µl 10x buffer and 0.5 µl nuclease-free water (Promega, Madison, USA). After a 25 minute incubation at 37°C, 2 µl DNase stop was added. The sample was centrifuged at 12,000 rpm for 2 minutes and 8 µl RNA was then added to 1 µl oligo-$d(T)_{15}$ (50 pmol/ µl, -BIOTECH AG, Abersberg, Germany) and 1 µl RNase free water in a 96 well PCR plate on ice. The PCR plate was incubated at 70°C for 5 minutes and then chilled on ice for 5 minutes. Fifteen µl of a reaction mix consisting of 5 µl 10x buffer, 5 µl 10 mM dNTPs, 0.5 µl M-MLV Reverse Transcriptase, 0.2 µl RNaseIn (Promega) and 4.3 µl nuclease-free water was added to each sample and mixed thoroughly by gentle pipetting. The samples were incubated at 43°C for 2 hours, 95°C for 5 minutes and then held at 15°C. The cDNA samples were subsequently stored at -20°C.

3.3.3 Analysis of nucleic acids by gel electrophoresis

3.3.3.1 Analysis of nucleic acids by agarose gel electrophoresis

DNA was analysed on agarose gels of a range of concentrations dependent upon the size of the DNA fragment to be visualised. PCR products greater than 300 base pairs were separated on 1-2% agarose gels and those smaller than this were separated on 2.5% gels. Agarose was weighed into a conical flask and 100 ml 1x tris-acetate-EDTA (TAE) was added. The flask was then heated in a microwave oven (Hinari) on high heat until the agarose had fully dissolved. The contents of

the flask were cooled under running cold water and ethidium bromide (2 µl/100ml) was added. The gel was poured into a gel tray and allowed to set for 30 minutes. The tray was then placed in the gel tank containing 1x tris acetate-EDTA (TAE). The gel comb was removed and the appropriate volume of sample was loaded into the wells. PCR products were loaded in orange G loading dye (50% glycerol in water, 0.25% orange G). DNA molecular markers were run alongside the samples to permit determination of DNA fragment length. The gel was run at 100 volts until the loading dye had migrated a sufficient distance. The gel was subsequently viewed with a GeneGenius Bioimager (SynGene, Cambridge, UK) and photographed using a Polaroid direct screen instant camera system.

3.3.3.2 Analysis of nucleic acids by polyacrylamide gel electrophoresis

A 6.4% polyacrylamide gel was prepared as follows: 4.8 mls of 40% acrylamide, 3 mls 10 % tris-borate-EDTA (TBE) and 22 mls sterile water were mixed with a magnetic stirrer for 5 minutes at room temperature. One hundred and fifty µl of 10% ammonium persulfate solution (APS) and 22 µl TEMED (N,N,N',N',-Tetramethyl-ethylenediamine) were added to the solution and it was mixed for a further minute. The solution was poured into a vertically positioned gel form, a gel comb was pushed into the top edge of the gel and the gel was allowed to set for 60 minutes. The gel comb was then removed and the gel was positioned in a gel tank which was subsequently filled with 1% TBE. Care was taken to ensure that no air bubbles had lodged within the wells. PCR products were loaded in blue/ orange loading dye and DNA molecular markers were run alongside the samples to permit determination of DNA fragment length. The gel was run at 100V until the loading dye had migrated a sufficient distance. The gel was then removed from the tank and soaked for 15 minutes in a solution of 200 ml 1% TBE containing 15 µl ethidium bromide before being visualised by a UV light.

3.3.4 Recovery of DNA from agarose and polyacrylamide gels

The PCR product or restriction enzyme digest product to be purified was loaded onto an agarose or polyacrylamide gel of the appropriate concentration as

described above. The gel was then run at 100 volts until the band of interest had migrated the required distance. The gel was viewed by UV light and the relevant band was cut out of the gel using a sterile razor blade and placed in a pre-weighed 1.5ml microcentrifuge tube.

3.3.4.1 Extraction of DNA from agarose gel using GENECLEAN® Kit

(MP-Biomedicals, Solon, USA)

The volume of each gel fragment was estimated from its weight and three volumes of sodium iodide were added to the gel fragment. The samples were heated at 50°C for 10 minutes, mixing by gentle inversion every 2 minutes to ensure complete dissolution of the gel fragment. GLASSMILK® was resuspended by vortexing for 1 minute. Four µl of GLASSMILK® was added to each sample and the sample was incubated for 10 minutes at room temperature with gentle inversion of each tube every 2 minutes. The samples were then centrifuged at 12000g for 60 seconds and the supernatant was removed with a pipette and discarded. Four hundred µl of New Wash was added to each pellet and the pellet was resuspended by pipetting. The samples were centrifuged at 12000g for 60 seconds and the supernatant was discarded. The pellet was then washed again with 400 µl of New Wash, centrifuging the tubes at 12000g for 60 seconds before discarding the supernatant. The samples were centrifuged again for 10 seconds at 12000g and any remaining supernatant was removed using a small bore pipette tip. The pellet was then dried in air for 10 minutes before being resuspended in 20 µl of sterile water by gentle pipetting. To elute the DNA, the samples were centrifuged at 12000g for 60 seconds. The supernatant was carefully transferred to a fresh mini-Eppendorf tube and centrifuged for 5 seconds at 12000g to pellet any residual GLASSMILK® before storing the DNA at -20°C.

3.3.4.2 Extraction of DNA using QIAquick® Gel Extraction Kit (Qiagen Ltd. Crawley, UK)

DNA was purified from agarose gel using the QIAquick® Gel Extraction Kit (Qiagen Ltd. Crawley, UK) in accordance with the manufacturer's instructions. Total

recovery of DNA was estimated by running 1 µl of the extracted DNA on an agarose gel and quantified by comparison with DNA molecular weight markers of known concentration.

3.3.4.3 Recovery of DNA from polyacrylamide gels

The PCR product of interest was excised from the acrylic gel using a sterile scalpel blade and was transferred to a clean Eppendorf tube. One hundred µl of a diffusion buffer consisting of 50 µl 20% sodium dodecyl sulphate (SDS), 20 µl 0.5 M EDTA, 1660 µl 3 M sodium acetate and 8270 µl sterile water was added to the gel fragment. The gel fragment was mixed with the diffusion buffer with a pipette tip and the Eppendorf tube was then incubated at 50°C for 40 minutes. The supernatant was transferred to a fresh Eppendorf tube and 300 µl sodium iodide was added to each sample. Four µl GLASSMILK® was then added to each sample and DNA was extracted using the GENECLEAN® Kit as described in section 3.3.4.1.

3.3.5 Polymerase Chain Reaction (PCR) [195]

3.3.5.1 PCR reagents

PCR Primers: Oligonucleotide primers (in the range of 20-25 bases) were obtained from biomers.net (Ulm, Germany) or MWG-BIOTECH AG (Abersberg, Germany). The primers were designed by eye using the BRCA1 gene sequence (GenBank accession number U14680). The primers were supplied as a concentrated stock solution with an OD260 reading (units of nucleic acid to concentration). The stocks were diluted to a concentration of 100 pmol/µl in sterile water.

dNTPs: 100 mM solutions of dATP, dCTP, dGTP and dTTP were purchased from Invitrogen Ltd. (Paisley, UK). A stock 10 mM solution of dNTPs were made by adding 50 µl of each 100 mM NTP to 800 µl sterile water and a 2 mM solution was made by adding 2 µl of each 100 mM NTP to 192 µl sterile water.

AmpliTaq Gold and Reaction Buffer: AmpliTaq Gold DNA Polymerase (5U/ µl, Applied Biosystems, California, USA) was stored at -20°C and used in accordance with the manufacturer's instructions. The 10x PCR Buffer II used with this consisted of 500 mM KCl, 100 mM Tris-HCl, pH 8.3.

AmpliTaq Gold 2x PCR Master Mix (250 U/ 5ml) was obtained from Applied Biosystems (California, USA). This included AmpliTaq Gold DNA Polymerase (0.05 U/µl), Gold Buffer (30 mM Tris-HCl, pH 8.0, 100 mM KCl), 400 µM each dNTP and 5 mM $MgCl_2$.

Platinum Taq DNA Polymerase (5 U/µl) was obtained from Invitrogen Ltd. (Paisley, UK). This was used with a 10x PCR Buffer (200 mM Tris-HCl, pH 8.4 and 500 mM KCl).

Pfu DNA Polymerase (a high fidelity polymerase) was obtained from Promega (Madison, USA). This was used in combination with AmpliTaq Gold in a 1:1 ratio with a 10x PCR buffer containing 200 mM Tris-HCl, pH 8.8, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1.0% Triton® X-100, and 1 mg/ml nuclease-free BSA.

3.3.5.2 Prevention of contamination

To reduce the risk of possible DNA contamination, either from genomic DNA samples or human DNA, all PCR reaction mixes were prepared in a designated "pre-PCR" room into which no genomic DNA samples or post-PCR products were permitted to be taken. Specific "pre-PCR" laboratory coats and gloves were worn whilst working in this environment. All pipette tips and other consumables used in the preparation of the PCR reaction mix were first exposed to UV radiation using a Stratalink 2400 cross linker (Stratagene, California, UK). This cross-links any contaminating DNA and prevents amplification of contaminating molecules. Target DNA/ cDNA was subsequently added to the PCR reaction mix in a second designated "pre-PCR" laboratory using filter tips and a set of Gilson pipettes that were not used to handle any products of PCR reactions.

3.3.5.3 PCR reactions

PCR using AmpliTaq Gold Polymerase

One µl of the DNA (30 ng/µl) or 2 µl of the cDNA (10 ng/µl) sample to be amplified was added to 5 µl of 10x PCR Buffer II, 4 µl 25 mM $MgCl_2$, 1 µl 10 mM dNTPs, 1 µl 3' primer (10 pmol/µl), 1 µl 5' primer (10 pmol/µl) and 0.2 µl AmpliTaq Gold Polymerase. The volume of the reaction was made up to 50 µl with sterile water. The plate was then sealed with an adhesive PCR film (ABgene, Epsom, UK) before being placed in PTC-0225 DNA Engine Tetrad thermal cycler (MJ research, Bio-Rad Laboratories, Massachusetts, USA).

PCR using Platinum Taq Polymerase

Two to four µl of the cDNA sample to be amplified was added to 5 µl of 10x PCR buffer, 1.5 µl 50 mM $MgCl_2$, 1 µl10 mM dNTP s, 1 µl 3' primer (10 pmol/µl), 1 µl 5' primer (10 pmol/µl) and 0.4 µl Platinum Taq Polymerase. The volume of the reaction was made up to 50 µl with sterile water. The plate was then sealed with an adhesive PCR film before being placed in a thermal cycler.

PCR using *Pfu* DNA Polymerase

Five µl of the DNA (25 ng/µl) sample to be amplified was added to 20 µl of a master mix containing 59.5 µl sterile water, 10 µl of 10x PCR Buffer II, 8 µl 10 mM dNTPs, 1 µl 3' primer (10 pmol/µl), 1 µl 5' primer (10 pmol/µl), 0.25 µl of AmpliTaq Gold Polymerase and 0.25 µl *Pfu* DNA Polymerase prepared on ice. The plate was then sealed with an adhesive PCR film (ABgene) before being placed in PTC-0225 DNA Engine Tetrad thermal cycler (MJ research) using a "hot start" technique.

Cycle parameters for different sets of PCR primers are included in the relevant results chapters with details of adjustments to the generic PCR methods. Sequences for all of the primers used are provided in the relevant sections.

3.3.5.4 Digestion of PCR products or plasmids with restriction enzymes

Ten µl of the PCR product or plasmid was added to 16.2 µl sterile water, 0.3 µl of the restriction enzyme, 0.5 µl BSA and 3 µl of the enzyme specific 10x buffer. The reaction was incubated at the temperature specified by the enzyme manufacturer for 60 minutes.

3.3.6 Automated sequencing

Automated sequencing was performed using either an ABI Prism 3100 Genetic Analyser (Applied Biosystems, California, USA) or a Beckman CEQ 8800 Capillary Sequencer (Beckman Coulter, Inc. California, USA).

3.3.6.1 Exosap clean up

To remove non-incorporated nucleotides from the PCR product prior to sequencing all PCR products were treated with an "Exosap" clean up. Five µl of the PCR product was added to 2 µl of EXOSAP (prepared by mixing 50 µl Exonuclease I with 40 µl Antarctic Phosphatase, 16 µl Antarctic Phosphatase Buffer [all from New England Biolabs, Hitchin, UK] and 144 µl sterile water) in a 96 well PCR plate. The plate was centrifuged at 3000g for 5 seconds to mix the EXOSAP with the PCR product. The plate was then placed in a thermal cycler and incubated at 37°C for 30 minutes and then 80°C for 15 minutes. This was followed by a hold at 15°C.

3.3.6.2 Sequencing PCR

One µl of the purified PCR product was added to 0.5 µl BigDye® Terminator Version 3.1, (Applied Biosystems, California, USA), 1.5 µl BigDye® Terminator buffer (5x) and 1 µl forward or reverse primer (10 pm/µl). The reaction volume was made up to a total of 10 µl with sterile water. The reaction mix was transferred to a 96 well PCR plate and transferred to a thermal cycler. PCR conditions were 24 cycles with denaturation at 94°C for 30 seconds, annealing at 50°C for 15 seconds and elongation at 60°C for 120 seconds, followed by a hold at 15°C. PCR products were purified using Montage SEQ cleanup plates (Millipore Ltd. Watford, UK).

3.3.6.3 Purification of Sequencing PCR Products: Montage SEQ_{96} Sequencing Reaction Cleanup Kit (Millipore)

The Montage SEQ_{96} Sequencing Reaction Cleanup Kit (Millipore Ltd. Watford, UK) was used in accordance with the manufacturer's instructions. Thirty µl of Injection Solution was added to each PCR product and mixed gently before transferring the sample to a well of the SEQ_{96} plate. The SEQ_{96} plate was placed on the vacuum manifold and a 25-30mmHg vacuum was applied until 15 seconds after all the wells had emptied. The SEQ_{96} plate was removed from the vacuum and excess liquid was blotted from the bottom of the SEQ_{96} plate. A further 25 µl of Injection Solution was added to each well and the SEQ_{96} plate was replaced on the vacuum manifold. A 25-30mm Hg vacuum was reapplied for 15 seconds after all the wells had emptied. The SEQ_{96} plate was removed from the vacuum and again excess fluid was blotted from the bottom of the plate. Twenty µl of Injection Solution was added to each sample and the purified sequencing products were resuspended by repeated pipetting. Fifteen µl of the sequencing products were then transferred to a sequencing plate and 15 µl formamide (99%, 1.13 g/ml, Sigma-Aldrich, Poole, UK) was added to each sample. The sequencing plate was then loaded into the Applied Biosystems 3100 DNA Analyser (Applied Biosystems, California, USA).

3.3.6.4 Automated Sequencing- Beckman

Sequencing was performed using an automated Beckman CEQ 8800 Capillary Sequencer. For plasmids, 5 µl of template DNA (200 ng/ µl) and 5 µl of primer (3.2 pmol/ µl) were used and for PCR products, 5 µl of DNA (10 ng/ µl) and 1 µl of primer (3.2 pmol/ µl) were used. Preparation of the samples and loading of the sequencer was performed by the Southampton University Human Genetics Department Sequencing Service.

3.3.6.5 Analysis of sequencing data

Analysis of sequencing data was performed using Sequence Navigator software (Applied Biosystems, California, USA) or Finch TV software (www.geospiza.com/finchTV/).

3.3.7 Pyrosequencing™

3.3.7.1 Preparation of PCR products for Pyrosequencing™

Single-stranded biotinylated PCR products were prepared for Pyrosequencing™ using a Vacuum Prep Tool (Biotage AB, Uppsala, Sweden). Three µl Streptavidin Sepharose™ HP (Amersham, New Jersey, USA) was added to 37 µl Binding buffer (10 mM Tris-HCl pH 7.6, 2 M NaCl, 1 mM EDTA, 0.1% Tween 20) and mixed with 20 µl PCR product and 20 µl high purity water for 10 minutes at room temperature using a Variomag Monoshaker (Camlab, Cambridge, UK). The beads containing the immobilised templates were captured onto the filter probes after applying the vacuum and then washed with 70% ethanol for 5 seconds, denaturation solution (0.2M NaOH) for 5 seconds and washing buffer (10 mM Tris-Acetate pH 7.6) for 5 seconds. The vacuum was switched off and the beads released into a PSQ 96 Plate Low containing 45 µl annealing buffer (20 mM Tris-Acetate, 2 mM $MgAc_2$ pH 7.6) and 5 µl of 5mM sequencing primer. The samples were heated to 80°C for 2 minutes and then allowed to cool to room temperature.

3.3.7.2 Pyrosequencing™ reactions

Pyrosequencing™ reactions were performed according to the manufacturer's instructions using the PSQ 96 SNP Reagent Kit (Biotage AB, Uppsala, Sweden) which contained the enzyme and substrate mixture and nucleotides. Assays were performed using the nucleotide dispensation orders detailed in the relevant section. The sample genotype and allelic ratio was determined using the Allele Frequency Quantification function in the SNP Software (Biotage AB).

3.3.8 Cloning procedures

3.3.8.1 Preparation of LB broth

Eight grammes LB Broth was weighed into a sterile culture jar and 400 ml sterile water was added. The jar was shaken until the powders had completely dissolved and then sterilized in a Rodwell Ensign autoclave at 121°C for 15 minutes (Rodwell Scientific Instruments, Basildon, England). The solution was permitted to cool until it reached approximately 37°C. Two hundred µl of ampicillin (50mg /ml) was

added to each 400 ml batch of LB Broth required for the growth of selected bacterial colonies. All batches of LB broth were stored at 4°C.

3.3.8.2 Preparation of agar plates

Six grammes Agar and 8.0g LB Broth were weighed into a sterile culture jar and 400mls of sterile water was added. The jar was shaken until the powders had completely dissolved and then sterilized in a Rodwell Ensign autoclave, (Rodwell Scientific Instruments, Basildon, UK), at 121°C for 15 minutes. The solution was permitted to cool until it reached approximately 37°C. Two hundred µl ampicillin (50 mg /ml) was added to the solution and it was shaken to mix thoroughly. Fifty ml of solution was poured into each of 8 sterile Petri dishes and the lids were immediately placed on them. The Petri dishes were left at room temperature until the agar had completely set and were then stored at 4°C until required. Ten µl X-gal (50 mg/ml, Promega, Madison, USA) and 100 µl 0.1 M IPTG (Promega, Madison, USA) were spread on plates required for the growth of pGEM®-T Easy Vector (Promega) colonies 30 minutes before use.

3.3.8.3 Ligation reactions

TOPO TA Cloning Kit (Invitrogen Ltd, Paisley, UK).

On ice, 1-3µl PCR product (depending on the strength of the product observed), 1 µl salt solution (1.2 M NaCl, 0.06 M $MgCl_2$), 1 µl pCR®4 vector (10 ng/µl) (Invitrogen Ltd. Paisley, UK), and sterile water to a final volume of 5 µl were combined. The reaction was then mixed gently, and incubated at room temperature for 5 minutes.

pGEM®-T Easy vector ligations (Promega, Madison, USA)

Three point five µl PCR product was mixed with 0.5 µl pGEM®-T Easy vector, 5 µl 2x ligation buffer and 1 µl T4 ligase (Promega). The reaction was mixed gently, transferred to a PCR plate and incubated at 15 °C overnight.

pCR® 3.1 vector ligations (Promega. Madison, USA)

Eighteen µl of the purified PCR product was digested with 1 µl of each appropriate restriction enzyme in a 30 µl reaction containing 3 µl 10x buffer, 0.5 µl BSA and 9.5 µl water. Four µl of pCR® 3.1 vector was also digested by the same enzyme pair in a reaction mix consisting of 2.5 µl 10x buffer, 1 µl of each restriction enzyme, 0.5 µl BSA and 16 µl sterile water. The digested products were analysed on an agarose gel as above and extracted using GENECLEAN® into 20 µl (vector) or 12 µl (minigene inserts and released vector fragment) volumes of sterile water. One point five µl vector was mixed with 3-4.5 µl PCR product (depending on observed strength of product), 1 µl 10x buffer, 1 µl T4 DNA ligase and sterile water to a total volume of 10 µl. The reaction was mixed gently, transferred to a PCR plate and incubated at 15°C overnight. Positive and negative ligation controls were set up for each ligation reaction using 3-4.5 µl vector fragment or water respectively in the place of the PCR product.

Ligation of pGEM®-T Easy inserts with pCR® 3.1 vector

Ten µl of the pGEM®-T Easy plasmid was digested for 3.5 hours at 37°C with 1 µl of each appropriate restriction endonucleases in a 30µl reaction mix containing 3 µl 10x buffer, 0.5 µl BSA and 9.5 µl water. For each ligation reaction, the pCR® 3.1 vector was also digested by the same pair of restriction enzymes in a reaction containing 4 µl of vector, 2.5 µl 10x buffer, 1 µl of each restriction enzyme, 0.5 µl BSA and 16 µl sterile water. The digest products were run on a 2% agarose gel in order to confirm that digestion was complete and that the products were of the anticipated size. The pCR® 3.1 vector and pGEM®-T Easy plasmid inserts were excised and extracted using GENECLEAN® into 20 µl (vector) or 12 µl (minigene inserts) volumes of sterile water. One µl of each sample was run on an agarose gel to ensure that DNA concentrations were similar. One point five µl vector was mixed with 3-4.5 µl PCR product (depending on observed strength of product), 1 µl 10x buffer, 1 µl T4 DNA ligase) and sterile water to a total volume of 10 µl. The reaction was mixed gently, transferred to a PCR plate and incubated at 15°C overnight.

3.3.8.4 Transformations

TopoTA Cloning (Invitrogen Ltd, Paisley, UK)

For TopoTA cloning, transformations were performed using the chemically competent TOP10 cells provided with the Invitrogen kit. The chemically competent cells were slowly thawed on ice. Two µl of the ligation reaction was added and stirred gently with a pipette tip to mix. The cells were then incubated in ice for 30 minutes. The cells were then heat-shocked at 42°C for exactly 30 seconds and immediately placed on ice for 2 minutes. Two hundred and fifty µl of SOC medium (2% Tryptone, 0.5% Yeast extract, 10 mM NaCl, 2.5 mM $MgSO_4$, 20 mM glucose) was added to each vial, and the cells were then shaken horizontally for 1 hour. LB, agar plates with ampicillin added were spread with 25 µl of 50mg/ml X-Gal (Promega, Madison, USA) and allowed to dry for 30 minutes before being spread with 100 µl of transformed cells. Once the liquid had been absorbed, the plates were inverted and placed in a 37°C incubator overnight.

pGEM®-T Easy and pCR® 3.1 cloning

For these vectors, transformation was performed using chemically competent high efficiency (> x 10^8 cfu/ µg DNA) JM109 cells (Promega, Madison, USA). For each transformation, 50 µl of JM109 cells were thawed on ice. Ten µl of the ligation reaction was added and stirred gently with a pipette tip to mix. The cells were incubated on ice for 20 minutes and then heat shocked at 42°C for exactly 45 seconds before being placed on ice for a further 2 minutes. Two hundred µl of SOC medium was added to each sample of cells and the cells were shaken horizontally for 90 minutes at 37°C. The entire volume of transformed cells was spread on LB/agar plates with ampicillin. For pGEM®-T Easy cloning the plates were spread with 10 µl Xgal (50mg/ml) and 100 µl 0.1M IPTG 30 minutes prior to plating out of the cells to permit blue /white colour screening of colonies. The plates were incubated at 37°C for 18 hours.

3.3.8.5 Selection and culture of colonies

The agar plates were subsequently stored at 4°C to permit colour development. Selected colonies from each plate were transferred to separate 10ml culture tubes

containing 5 ml L-Broth with ampicillin (0.025 mg/ml) and incubated at 37°C overnight in a shaker. Five hundred µl culture medium was mixed with 150 µl 30% glycerol and stored at -80°C. Plasmid DNA was extracted from the remaining culture medium as described below.

3.3.8.6 Extraction of plasmid DNA

Qiagen™ Miniprep Kit (Qiagen Ltd. Crawley, UK)

One point five ml of culture medium was spun down in an Eppendorf at 7000g for 10 minutes in a Thermo IEC microcentrifuge. The supernatant was then carefully removed to leave only the DNA pellet. The pellet was completely resuspended in buffer P1 before adding the P2 lysis buffer. The tubes were then inverted gently 4-6 times to ensure complete lysis before addition of buffer P3. Following addition of buffer P3 each tube was immediately inverted 4-6 times until the solution became cloudy and white. The tubes were then spun at 13000g for 10 minutes in a Thermo IEC microcentrifuge and the supernatant was carefully transferred to an appropriately labeled QI Prep spin column. The columns were spun at 13000g for 1 minute (to allow the DNA to bind to the columns) and then washed twice for 1 minute at 13000g using firstly 500 µl of buffer PB and then 750 µl of buffer PE. The column was then spun for an additional minute at 13000g to remove all residual PE buffer. The column was placed in a new sterile labeled microcentrifuge tube. To collect the DNA, 25 µl of distilled water was added to the centre of the column and it was spun at 14000 rpm for a further 1 minute.

Wizard® *Plus* SV Minipreps DNA Purification System (Promega, Madison, USA)

Plasmid DNA was extracted using the Wizard® Plus SV Minipreps DNA Purification System, according to the manufacturer's instructions. The culture tubes were centrifuged at 10,000g for 5 minutes in a tabletop centrifuge and the supernatant was then carefully removed to leave only the pellet of DNA. The pellet was resuspended in 250 µl Cell Resuspension Solution by pipetting and the samples were transferred to sterile Eppendorf tubes. Two hundred and fifty µl of

Cell Lysis Solution was added and the tubes were then inverted gently 4-6 times and allowed to incubate at room temperature for 5 minutes to ensure complete lysis. Ten µl of Alkaline Protease Solution was then added to each sample and each tube was immediately inverted 4-6 times and incubated at room temperature for 5 minutes. Three hundred and fifty µl of Neutralisation Solution was added and each tube was immediately inverted 4 times. The tubes were then spun at 12000g for 10 minutes in a Thermo IEC microcentrifuge and the supernatant was decanted into a spin column placed within a collection tube. The columns were spun at 12000g for 1 minute (to allow the DNA to bind to the columns) and the supernatant was discarded. Seven hundred and fifty µl wash solution was added to each column. The columns were spun at 12000g for 1 minute before discarding the supernatant. A second wash was then performed using 250 µl of Wash Solution and a 2 minute spin at 12000g. The column was then spun for an additional minute at 12000g to remove all residual wash solution. The column was placed in a new sterile labelled Eppendorf tube. To elute the DNA, 80 µl of distilled water was added to the centre of the column and it was spun at 12000g for a further 1 minute.

3.3.8.7 Analysis of plasmid DNA

As EcoRI sites flank the 3' and 5' cloning sites of the pCR®4.0 and pGEM®-T Easy vectors, products of cloning into these vectors were analysed by a single digest with EcoRI in a reaction mix containing buffer H. DNA sequences cloned into the pCR® 3.1 vector were analysed by double digests using the appropriate restriction enzymes and buffers. For each reaction, 8 µl of plasmid DNA was digested with 0.5 µl each enzyme (or 1 µl EcoRI) in a 20 µl reaction containing 2 µl 10x Buffer, 0.5 µl BSA and 9.5 µl sterile water for 1 hour at 37°C. The digested clones were run on a 1% agarose gel to determine the insert size and cloning efficiency. The plasmid DNA was then sequenced according to the protocol in section 3.3.6.

3.3.9 Transfection Procedures

3.3.9.1 Preparation of cells

All procedures involving cell culture were performed in a filtered hood under strict sterile conditions. HEK-293 adherent cells were grown up and maintained in varying volumes of Dulbecco's modified eagle's medium with 10% foetal bovine serum (FBS) and penicillin 100 u/ml, streptozocin 0.1 mg/ml (Dul/FBS/P/S) in $75cm^3$ cell culture flasks at 37°C. Twenty-four hours prior to transfection procedures, a flask of cells was selected and the cells were checked for viability using a light microscope. Cells present in five separate field views were counted, an average cell concentration was calculated and the total number of cells in the flask was then estimated. The medium was aspirated and cells were washed with 10 ml of PBS. One point five ml of Trypsin -EDTA (Sigma-Aldrich Ltd. Poole, UK.) was added to the culture flask and the cells were agitated in this for 5 minutes to break adherent protein strands. Sufficient fresh Dulbecco's modified eagle's medium with FBS, penicillin and streptozocin to produce a cell concentration of $0.25\text{-}1x10^6$/ ml was added to the cells using a sterile pipette and the cells were gently resuspended by repeat pipetting. One hundred µl of cells was transferred to each well of a sterile 12 well cell culture plate (well size = $3.8cm^3$). A further 900µl of Dul/FBS/P/S was added to each well resulting in a final cell concentration of $0.25\text{-}1x10^5$/ ml. The cell culture plates were then incubated at 37°C for 24 hours.

3.3.9.2 Transfection with plasmid DNA using siPORT™XP-1 Transfection Agent (Ambion, Austin, USA)

Prior to transfection, the cell culture plates were examined by light microscopy to ensure that the adherent cells were 30-60% confluent. Medium was removed from each well of the prepared cell culture plate by gentle aspiration and 1ml fresh Dul/FBS/P/S warmed to 37°C was added to each well. The cell culture plates were then again incubated at 37°C whilst the transfection reaction mix was prepared. A master mix of 50 µl Dulbecco's medium per reaction plus 1.3 µl SI Port (Ambion) per reaction was prepared and vortexed briefly. Aliquots of 51.3 µl were transferred to sterile Eppendorf tubes. Five µl of plasmids diluted to a

concentration of 100 ng/ml were added to separate medium/ SIport aliquots and mixed by gentle pipetting. The transfection reaction mixes were incubated at room temperature for 25 minutes and then each was added to one well of the cell culture plate, dotting the reaction mix onto the surface of the medium. The cells were subsequently incubated at 37°C for 24 hours.

3.3.9.3 Extraction of RNA from cell cultures

Prior to RNA extraction, the transfected cells were examined by light microscopy to check that they were still viable. The medium was removed from each cell culture well by gentle aspiration and 0.5 ml TRI® Reagent was added to each well. The cells were resuspended in the TRI® Reagent by pipetting and the resulting lysate was transferred to a fresh 1.5 ml microcentrifuge tube. RNA extraction was performed as described in section 3.3.1.6. If RNA extraction could not be performed immediately, the cell lysate/ TRI® Reagent samples were stored at -20°C and then allowed to thaw completely immediately prior to the RNA extraction procedure.

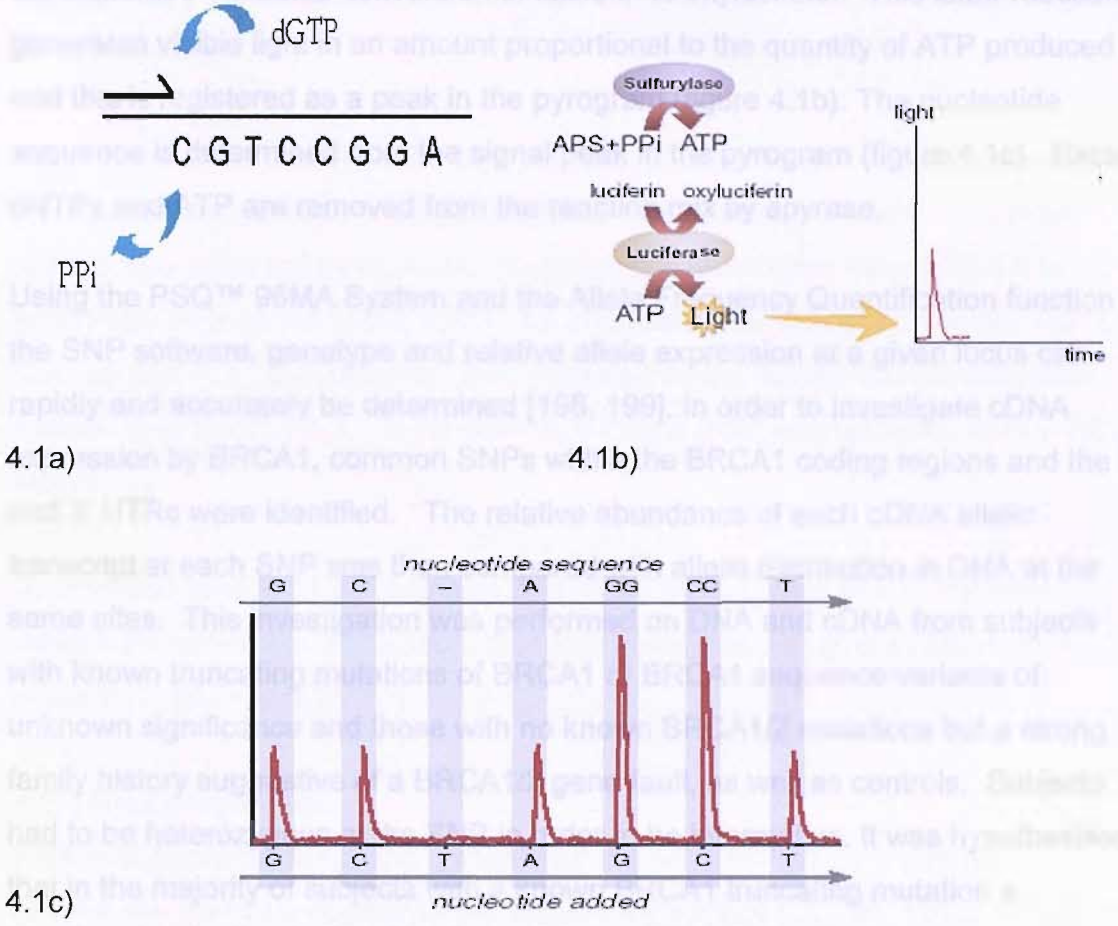
4.0 Use of Pyrosequencing™ technology to quantitate allele expression in patients with BRCA1 mutations

4.1 Introduction

The principle that the NMD pathway will degrade most mRNA transcripts acquiring a PTC from a frameshift or nonsense mutation can potentially be exploited by analytical methods capable of detecting a reduction in the quantity of mRNA expressed by the mutant allele. Perrin-Vidoz *et al.* demonstrated that a significant proportion (80%) of BRCA1 alleles carrying a PTC exhibit a 1.5-5 fold reduction in mRNA abundance [179]. Montagna *et al.* have also used the effects of NMD to develop a method of allele-specific gene expression analysis of BRCA1 and BRCA2 [196]. These studies suggest that such techniques can be used reliably to detect underlying truncating BRCA1/2 mutations.

Recently, Sharp *et al.* reported putative monoallelic expression of BRCA1 cDNA in 2 patients with BRCA1 DNA sequence variants of unknown clinical significance [171]. As discussed in the introduction, there is increasing evidence that missense mutations can have previously unsuspected effects on mRNA expression via disruption of splicing regulation. Quantitative analysis of mRNA allelic expression could therefore provide a means of screening for BRCA1/2 missense mutations with pathogenic potential in patients who have previously had a negative routine BRCA1/2 screen.

Pyrosequencing™ is a real time sequencing method used for the analysis of short to medium length DNA sequences which is based on the detection of pyrophosphate (PPi) release during DNA synthesis [197].



4.1a) 4.1b)

4.1c)

***Figure 4.1**: The principles of Pyrosequencing™ (see text for description)*

A single stranded, PCR amplified DNA template is hybridised to a sequencing primer and incubated with the enzymes, DNA polymerase, ATP sulfurylase, luciferase and apyrase, and the substrates, adenosine 5´ phosphosulfate (APS) and luciferin. Deoxynucleotide phosphates (ATP, CTP, GTP and TTP) are repeatedly added to the reaction mix, one at a time, in an order determined by the DNA sequence. Incorporation of a dNTP complementary to the template is catalysed by DNA polymerase and accompanied by the release of PPi equimolar in quantity to the amount of nucleotide incorporated (figure 4.1a). In a reaction

catalysed by sulfurylase, APS reacts with PPi to produce ATP, which then drives the luciferase-mediated conversion of luciferin to oxyluciferin. This latter reaction generates visible light in an amount proportional to the quantity of ATP produced and this is registered as a peak in the pyrogram (figure 4.1b). The nucleotide sequence is determined from the signal peak in the pyrogram (figure 4.1c). Excess dNTPs and ATP are removed from the reaction mix by apyrase.

Using the PSQ™ 96MA System and the Allele Frequency Quantification function of the SNP software, genotype and relative allele expression at a given locus can rapidly and accurately be determined [198, 199]. In order to investigate cDNA expression by BRCA1, common SNPs within the BRCA1 coding regions and the 5' and 3' UTRs were identified. The relative abundance of each cDNA allelic transcript at each SNP was then compared with allelic distribution in DNA at the same sites. This investigation was performed on DNA and cDNA from subjects with known truncating mutations of BRCA1 or BRCA1 sequence variants of unknown significance and those with no known BRCA1/2 mutations but a strong family history suggestive of a BRCA1/2 gene fault, as well as controls. Subjects had to be heterozygous at the SNP in order to be informative. It was hypothesised that in the majority of subjects with a known BRCA1 truncating mutation a significant reduction in the abundance of one cDNA allele (if not complete allele dropout) would be seen as a consequence of NMD of the mutant allele. It was also anticipated that cDNA allelic imbalances would be seen in subjects with missense mutations of BRCA1 that resulted in splicing disruption.

4.2 Methods

4.2.1 Subjects

Potential study participants were identified from the WCGS database:
a) Patients with a previously identified truncating mutation of a BRCA1 or BRCA2,
b) Patients with a previously identified DNA sequence variation of unknown significance of BRCA1 or BRCA2 (including two patients whose DNA and cDNA had previously been assessed by Sharp *et al.* [171].

c) Patients with a personal history of breast or ovarian cancer and a very strong family history of cancer who had undergone genetic screening and had a negative BRCA1/2 screen.

Patients were recruited into this study by the WGCS. A fresh 5ml sample of blood was collected from each patient who consented to enter the study. The blood sample was collected by standard venepuncture techniques into an EDTA bottle which was transported to the WRGL for immediate isolation of MNCs into TRI®zol (see section 3.3.1.2). Genomic DNA, prepared by the standard DNA extraction method (see section 3.3.1.1) for each study participant was retrieved from WRGL stores.

Anonymous genomic DNA samples and linked MNC pellets in TRI®zol from individuals referred to WRGS for screening of non-neoplastic conditions were obtained to provide a control group.

4.2.2 RNA preparation

RNA was extracted from MNC pellets stored under TRI®zol by a manual chloroform/ phenol precipitation technique (see section 3.3.1.3). BRCA1 cDNA for each subject and control was subsequently prepared by allele-specific RT-PCR (see section 3.3.2.1). The quality of each cDNA sample was verified by PCR amplification with previously tested BRCA1 cDNA primers (see appendix D).

4.2.3 Identification of BRCA1 SNPs and design of primers

Seven SNPs within BRCA1 were identified as occurring in the general population with a frequency of approximately 30% (WRGL BRCA1 database; www.ncbi.nlm.nih.gov/SNP/index.html; www.genome.utah.ed/genesnps), see table 4.1.

Exon	Nucleotide Number	Δ Nucleotide	q Frequency	NCBI SNP reference number
11	c2082	C > T	0.343	rs1799949
11	c2311	T > C	0.332	rs16940
11	c2612	C > T	0.381	rs799917
11	c3113	A > G	0.275	rs16941
11	c3548	A > G	0.297	rs16942
13	c4308	T >C	0.253	rs1060915
16	c4837	A > G	0.304	rs1799966

Table 4.1: *The seven selected SNPs within the BRCA1 gene*

Primers of 15-25bp in length were designed to amplify DNA/cDNA in amplicons of 100-250bp in length across the site of the SNP (table 4.2).

SNP	Forward and Reverse Primers	Sequence 5'to 3'	Start position (HUGO)	End position (HUGO)	Amplicon size
C2082T	C2082T FB C2082T R	biotin-ctg caa ctg gag cca aga ag- -tgg aag gct agg att gac aaa-	2022 2161	2041 2181	161bp
T2311C	T2311C FB T2311C R	biotin-atg ctg aag aac cca aag at- -ttt gcc ttc cct aga gtg ct-	2228 2364	2247 2384	157bp
C2612T	C2612T F C2612T RB	-tga act tga tgc tca gta ttt gc- biotin-gtt gca cat tcc tct tct gc-	2550 2633	2572 2652	102bp
A3113G	A3113G F A3113G RB	-cag tga gca caa tta gcc gta- biotin-act gga gcc cac ttc att agt-	3065 3151	3085 3171	107bp
A3548G	A3548G FB A3548G R	biotin-tga cct gtt aga tga tgg tga a- -tat ggg tga aag ggc tag ga-	3453 3566	3474 3586	134bp
T4308C	T4308C Fnew T4308C Rnew	biotin-cca tgc aac ata acc tga taa a- -gtg ctt tgt tct gga ttt cg-	4196 4327	4217 4346	151bp
A4837G	A4837G FB A4837G R	biotin-aag aca gag ccc cag agt ca- -ccc tgc tca cac ttt ctt cc-	4732 4885	4751 4903	162bp

Table 4.2: *Primers for amplification of DNA/cDNA at BRCA1 SNP sites*

4.2.4 PCR reaction

One µl target DNA (30ng/µl) or 2µl cDNA (10-15ng/µl) was added to 5µl 1x Buffer II, 4µl 25mM $MgCl_2$, 1µl 10mM dNTP s, 1µl 3' primer (10pmol/µl), 1µl 5' primer (10pmol/µl) and 0.2µl AmpliTaq Gold Polymerase. The volume of the reaction was made up to 50µl with sterile water. PCR conditions were 94°C for 7 minutes; 50 cycles with denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds

and elongation at 72°C for 30 seconds; 1 cycle at 72°C for 7 minutes; and a final hold at 15°C.

4.2.5 Preparation of PCR products for Pyrosequencing™

Single-stranded biotinylated PCR products were prepared for Pyrosequencing™ as described in section 3.3.7.1.

4.2.6 The Pyrosequencing™ reaction

Pyrosequencing reactions were performed as described in section 3.3.7.2. Assays were performed using the nucleotide dispensation orders shown in table 4.3. The sample genotype and allelic ratio was determined using the Allele Frequency Quantification function in the SNP Software (Biotage AB).

SNP	Sequencing primer	Sequence of primer 5' to 3'	Sequence to be analysed	Dispensation order
C2082T	C2082Tseq	-ttc agc tct ggg aaa g-	tatcg/actgtcatgtc	gtatcgagc
T2311C	T 2430Cseq	-aat cag tac gcca-	g/atgaaatactgctact	cgactgata
C2612T	C2612Tseq	-cgc cag tca tttgct-	c/tgttttcaaatccaggaaa	actcgttcat
A3113G	A3113Gseq	-aga gaa aat gtt tttaaa g-	g/aagccagctcaagc	cgatgagc
A3548G	A3548Gseq	-tcc tgc taa gct ctc ct-	c/ttctggacgcttttgc	gctgctgac
T4308C	T4308Cseq	-gtc ctc aag ggc aga-	g/agagtcacttatatgatgga	cagcagtca
A4837G	A4837GSeq	-gag cag cag ctg gac-	c/tctgggcagattctgcaac	gtcatgcag

Table 4.3: *Sequencing primers, sequences analysed and dispensation orders for Pyrosequencing™ reactions*

4.3 Results

4.3.1 Optimisation of technique: DNA analysis

Pyrosequencing at each SNP was initially tested on pooled genomic DNA from anonymous controls. Pilot studies of Pyrosequencing using the above primer pairs had a good pass rate, but results for heterozygotes failed to achieve any exact 50%: 50% allelic ratios. The mean ratios for pilot studies using primers for A4596G resulted in a mean ratio of 53.7%: 46.3% (range 63.6%-36.4). Attempts were made to improve the allelic ratio towards 1:1 by altering the concentration of the PCR product and of the forward and reverse primers – these measures had no significant effect on the allelic ratio.

To determine the informativeness of each SNP in the general population (i.e. the percentage of individuals heterozygous at each SNP), 80 anonymous pooled control genomic DNA samples (concentration 100ng/µl) held by NGRL (Wessex) were analysed at each of the seven BRCA1 SNP sites together. A single 96 well PCR plate was used for each SNP, with 10 of the controls being run in duplicate and 6 negative controls. This confirmed that all SNPs are informative in approximately 40% of the general population. Significant linkage disequilibrium does exist between these SNPs with 20% of the samples proving to be heterozygous at all seven SNPs and 20% of the population proving non-informative at any of the seven SNPs. The failure rate for this procedure varied between 11.1-23.3% (table 4.4).

SNP	Total No. Of Valid Results	Failure rate	No. Homozygotes	No. Heterozygotes	% Informative
C2082T	77	14.4%	39 G/G 6 A/A	32 G/A	41.6%
T2311C	74	17.7%	7 G/G 33 A/A	34 G/A	45.9%
C2612T	77	14.4%	9 T/T 33 C/C	35 T/C	45.5%
A3113G	69	23.3%	33 A/A 6 G/G	30 A/G	43.5%
A3548G	78	13.3%	38 T/T 8 C/C	32 C/T	41.0%
T4308C	78	13.3%	38 A/A 8 G/G	32 G/A	41.0%
A4837G	80	11.1%	40 T/T 8 C/C	32 C/T	40.0%

Table 4.4: *Results of analysis of genomic DNA at each SNP site in group of 90 anonymous individuals representative of general population*

4.3.2 Optimisation of technique; analysis of cDNA

Analysis of cDNA from a group of control subjects at all SNPs was initially unsuccessful owing to problems with DNA contamination of cDNA, despite the incorporation of a DNase step into the RT-PCR protocol. It was established that any further attempts to purify the cDNA would be most likely to simply result in loss of RNA rather than contaminating DNA. Instead, further primers were designed to exploit the difference in capacity of DNA and cDNA to amplify across introns for SNPs in exons 13 and 16 (i.e. trans-exonic primers were adopted for PCR of the cDNA) and plans were made to develop a "nested PCR" technique to amplify cDNA relevant to exon 11 SNPs. Trials of the new T4308C and A4837G primers (table 4.5) were successful: these primers failed to amplify DNA but produced a strong PCR product when using cDNA as their template.

SNP (ATGn)	Forward and Reverse Primers	Sequence 5'to 3'	Amplicon size	Start HUGO	End HUGO
T4308C	T4308C FB T4308C R	biotin-gca tct ggg tgt gag agt ga- -tct gat gtg ctt tgt tct gga-	248bp	4105 4332	4124 4352
A4837G	A4837G cDNA FB A4837G cDNA R	biotin-ctg aag aca gag ccc cag ag- -tcc gtt cac aca caa act cag-	360bp	4739 5078	5757 5098

Table 4.5: *Revised primers for trans-exonic amplification of SNPs in exon 13 and 16*

4.3.3 Comparison of allelic ratios at exon 13 and exon 16 SNPs in subjects and controls (cDNA synthesised by allele-specific RT-PCR)

cDNA samples were available for: 13 patients with known truncating mutations of BRCA1, 5 patients with BRCA1 DNA variants of unknown clinical significance and 5 patients with strong family histories of breast/ovarian cancer but confirmed negative routine screens for mutations of BRCA1/2 (see table 4.6), plus 80 anonymous controls.

Analysis of genomic DNA of all subjects at each SNP revealed that 7 of the 13 patients with known truncating mutations, 4 of the 5 patients with DNA variations of unknown significance and 2 of the 5 patients with negative routine BRCA1/2 screening results were informative at each of these seven SNPs. Analysis of the ratio of allelic frequency at the exon 13 SNP (T4308C) and the exon 16 SNP (A4837G) in DNA and cDNA samples of these informative patients was repeated on three occasions. No complete allelic dropout was identified but recurrent alterations in allelic frequency within cDNA compared with genomic DNA were found at both the exon 13 and 16 SNPs in two patients with known truncating mutations of exon 11 (patients T02 and T12) and at the exon 16 SNP only in one further patient (T06), who is also known to have a truncating mutation within exon 11.

Code	Result of BRCA1 Screening	Comment
T01	AA deletion at ATGn3706, exon 11	Non-informative at T4308C & A4837G
T02	T>A at ATGn1326, exon 11	Informative at T4308C & A4837G
T03	GTCT deletion at ATGn1143, exon 11	Non-informative at T4308C & A4837G
T04	G>A at ATGn4185, exon 12	Non-informative at T4308C & A4837G
T05	G>A at ATGn4185 12	Non-informative at T4308C & A4837G
T06	4bp deletion at ATGn3726, exon 11	Informative at T4308C & A4837G
T07	A>T at ATGn1141, exon 11	Informative at T4308C & A4837G
T08	C>T at ATGn 3817, exon 11	Informative at T4308C & A4837G
T09	5bp deletion at ATGn3700, exon 11	Non-informative at T4308C & A4837G
T10	5bp deletion at ATGn3700, exon 11	Informative at T4308C & A4837G
T11	Deletion of exons 9-12	Non-informative at T4308C & A4837G
T12	4bp deletion at ATGn4065, exon 11	Informative at T4308C & A4837G
T13	2bp deletion at ATGn 66, exon 2	Informative at T4308C & A4837G
T14	1bp deletion at ATGn3066, exon 11	Non-informative at T4308C & A4837G
U01	A >G at ATGn 1067, exon 11	Informative at T4308C & A4837G
U02	G >A at ATGn 2077 exon 11 – missense mutation Monoallelic BRCA1 expression reported by Sharp *et al.* [161]	Informative at T4308C & A4837G
U03	A >G at ATGn 5198, exon 20- missense mutation	Informative at T4308C & A4837G
U04	T > G at ATGn 53, exon 2 – missense mutation	Informative at T4308C & A4837G
U05	T > G at ATGn 736, exon 11 - missense mutation Monoallelic BRCA1 expression or constitutive skipping of exon 11 reported by Sharp *et al.* [161]	Non-informative at T4308C & A4837G
N01	Negative BRCA1/2 screen	Non-informative at T4308C & A4837G
N02	Negative BRCA1/2 screen	Non-informative at T4308C & A4837G
N03	Negative BRCA1/2 screen	Non-informative at T4308C & A4837G
N04	Negative BRCA1/2 screen	Informative at T4308C & A4837G
N05	Negative BRCA1/2 screen	Informative at T4308C & A4837G
N06	Negative BRCA1/2 screen	Informative at T4308C & A4837G

Table 4.6: *Results of routine BRCA1/2 DNA screening and DNA analysis at BRCA1 SNPs T4308C and A4837G for 24 subjects recruited to study*

SNP	Exon 16 A4837G		Exon 13 T4308C	
Pt.	DNA A:G	cDNA A:G	DNA T:C	cDNA T:C
T02	53.8%: 46.2%	66.9%: 33.1%	48.8%: 51.2%	66.7%: 33.3%
T06	53.5%: 46.5%	60.7%: 39.3%	47.8%: 52.2%	55.9%: 44.1%
T12	52.3%: 47.7%	61.5%: 38.5%	49.4%: 50.6%	65.1%: 34.9%

***Table 4.7**: Mean results from repeat analysis of allelic frequency at SNPs within exon 13 and exon 16 of 3 patients exhibiting alterations in allelic frequency in cDNA compared with DNA*

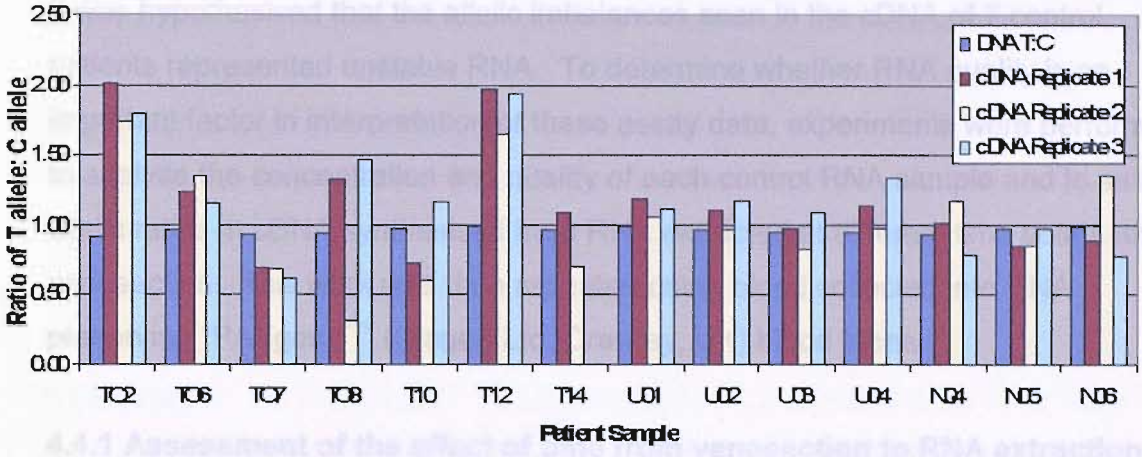
Of the 80 control DNAs tested, 38 were informative (heterozygous) at the A4837G and T4308C SNPs. Significant problems were encountered when attempting to analyse the control cDNA, due to extremely poor PCR yields. Attempts were made to optimise the PCR by adjusting magnesium concentration, elongation time and number of cycles, but PCR products were still inadequate for successful Pyrosequencing™. The A4837G and T4308C cDNA primers were subsequently tested on reference cDNA samples of known concentration produced by the random hexamer method of cDNA synthesis. These cDNA samples were amplified readily using the above PCR conditions. It was therefore concluded that poor quality cDNA was the cause of the Pyrosequencing™ failures. Fresh RNA samples were subsequently prepared for each control and subject using the RNeasy® method (see section 3.3.1.4) and cDNA was synthesised by the random hexamer method (3.3.2.2).

4.3.4 Comparison of allelic ratios at exon 13 and exon 16 SNPs in subjects and controls at T4308C (cDNA synthesised by random hexamer method)

Fourteen of the subjects were informative at the exon 13 and exon 16 SNPs, including 7 patients with known truncating mutations, 4 with point mutations of unknown significance and 3 with negative BRCA1/2 screens but strong family histories (see table 4.6).

DNA and cDNA from the informative subjects was analysed at the T4308C SNP in triplicate. DNA samples had a mean T: C ratio of 0.96 (+/- 0.02). Six out of 7 cDNA samples from known truncating mutations had a T: C allele ratios that varied significantly from a ratio of 1:1 (defined as 0.85-1.15). CDNA samples from patients with point mutations of unknown significance or negative BRCA1/2 screens but strong family history (n=7) showed no variation from an equal allelic ratio. The mean T: C allele ratio for the control DNA samples (n=38) was 0.94 (+/- 0.02). Unexpectedly, seven control cDNA samples (indicated by * on figure 4.1) also had T: C allele ratios that also varied significantly from an equal ratio.

a)



b)

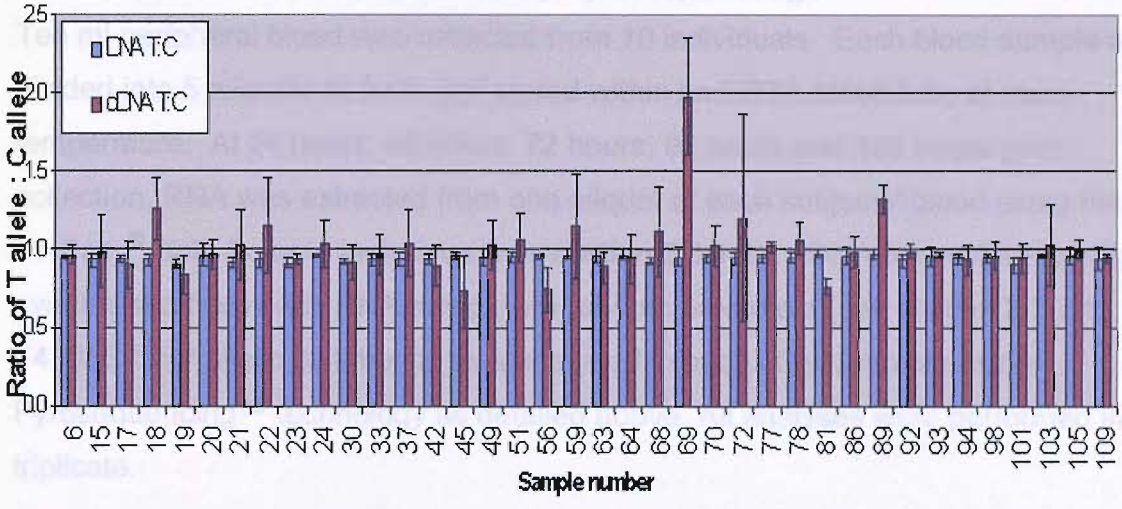


Figure 4.2: *Bar charts showing ratio of T allele: C allele for BRCA1 SNP T4308C for DNA and cDNA samples from:*

a) 14 putative familial breast cancer patients with no known truncating mutations (T), point mutations of unknown significance (U) and patients with strong family history of familial breast cancer but no detectable BRCA1/2 mutation

b) 38 healthy controls (error bars indicate standard deviation for triplicate samples.

Of note, one of the subjects (U02) described by Sharp *et al.* [171] as showing monoallelic expression of BRCA1 was in fact found to be heterozygous at T4308C and A4837G on Pyrosequencing™ analysis of cDNA.

4.4 Analysis of effect of RNA quality on Pyrosequencing™ results

It was hypothesised that the allelic imbalances seen in the cDNA of 7 control patients represented unstable RNA. To determine whether RNA quality is an important factor in interpretation of these assay data, experiments were performed to analyse the concentration and quality of each control RNA sample and to assess allelic ratios in cDNA synthesised from RNA extracted at different time points after venesection. This work was then repeated using blood collected into RNA preserving "PAXgene™" (Qiagen Ltd. Crawley, UK) blood tubes.

4.4.1 Assessment of the effect of time from venesection to RNA extraction on cDNA allelic ratios as determined by Pyrosequencing™

Ten ml peripheral blood was collected from 10 individuals. Each blood sample was divided into 5 aliquots of 2mls and stored within an EDTA blood tube at room temperature. At 24 hours, 48 hours, 72 hours, 96 hours and 168 hours post collection, RNA was extracted from one aliquot of each subjects' blood using the RNEasy® Spin Column technique (see section 3.3.1.4). cDNA was subsequently synthesised from each RNA sample via random hexamers (see section 3.3.2.2). T4308C SNP allelic distribution was analysed in each cDNA sample using Pyrosequencing™ technology as detailed above. All analyses were performed in triplicate.

4.4.2 Assessment of PAXgene™ RNA preserving system

The PAXgene™ Blood RNA Tube (Qiagen) is a plastic evacuated tube for the collection of whole blood that contains 6.9ml of an additive ("RNA additive") that stabilises cellular RNA for up to 5 days at room temperature. 10ml peripheral blood was collected from 2 individuals known to be informative at the T4308C SNP. Each blood sample was divided into 5 aliquots of 2mls and stored within a PAXgene™ Blood RNA Tube blood tube at room temperature. At 24 hours, 48 hours, 72 hours, 96 hours and 168 hours post collection, RNA was extracted from one aliquot of each subject's blood using the RNeasy® Spin Column technique (see section 3.3.1.4) and cDNA was subsequently synthesised using random

hexamers (see section 3.3.2.2). T4308C SNP allelic distribution was analysed in each cDNA sample using Pyrosequencing™ technology as detailed above.

4.4.3 Analysis of RNA quality and concentration

To assess concentration and quality, the 18s ribosomal RNA to 28s ribosomal RNA ratio of all control RNAs was analysed using an Agilent 2100 Bioanlyser. This work was performed by Dr. Feng Lin, University of Southampton.

4.5 Results

4.5.1 Assessment of effect of time from venesection to RNA extraction on cDNA allelic ratios as determined by Pyrosequencing™

Five of the samples were informative at T4308C. At 24 hours, the mean T: C allele ratio was 0.99 (+/- 0.1) and all 5 samples had reproducible ratios. At 48 hours or greater T: C ratios were not reproducible and the variability increased with time. 20/20 samples had T: C ratios that varied significantly from an equal ratio.

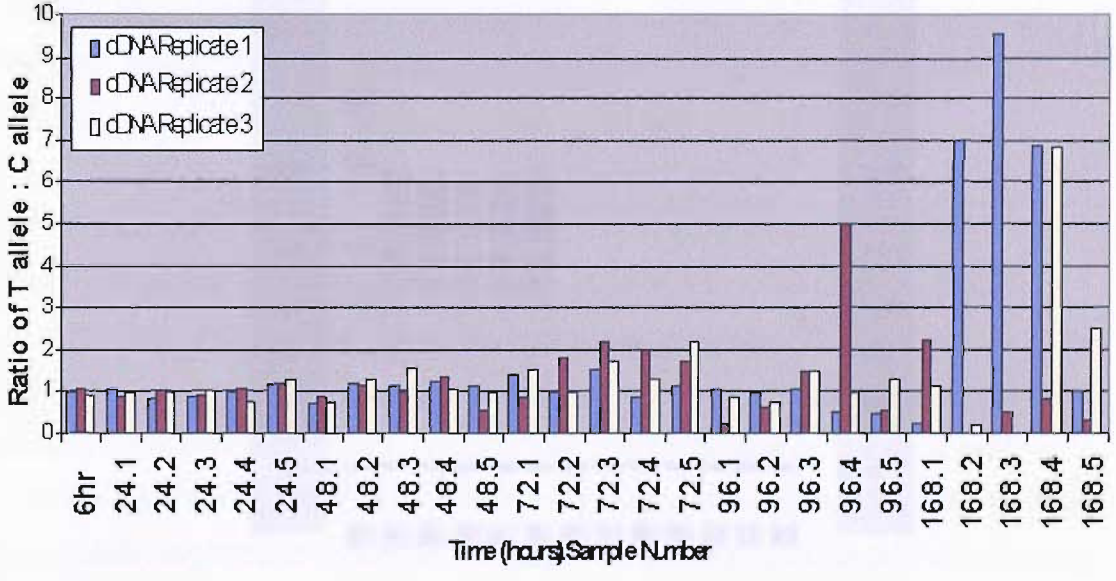


***Figure 4.3**: Allelic ratios of 5 normal subjects where RNA has been extracted 6, 24, 48, 72, 96 and 168 hours after blood collection.*

4.5.2 Assessment of PAXgene™ RNA preserving system

The mean T: C allele ratio was 1.01 (+/- 0.06) at 24 hours, 1.04 (+/-0.04) at 48 hours, 1.04 (+/- 0.05) at 96 hours and 1.11 (+/-0.05) at 168 hours. Only one 96-hour sample varied significantly from an equal ratio.

4.5.3 Analysis of RNA quality and concentration

Thirty control RNA samples were analysed. See figure 4.3 for sample trace. Mean concentration was 28.19ng/µl (range 5-156ng) with a mean 28s/18s ratio of 0.973 (range 0-2.6). There was no correlation between reduced RNA quality and presence of allelic imbalance. Analysis of RNA quality for samples extracted at different time points after venesection showed a mean 28s/18s ratio of 0.82 at 48 hours and 0.26 at 96 hours.

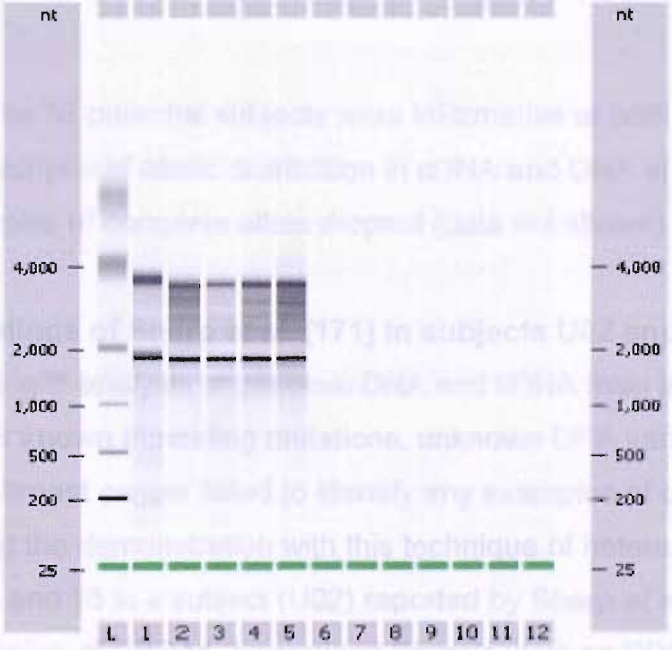


Figure 4.4: *Sample traces from RNA quality assessment study; 28s/ 18s ratios analysed by electrophoresis*

4.6 Search for complete allele dropout in patients with a negative routine BRCA1/2 screen but strong family history suggestive of BRCA1/2 mutation

To determine how much further to take this area of research I proceeded to look to see if I could find any examples of complete allele dropout at the A4837G and T4308C SNPs in subjects with a strong family history of breast/ovarian cancer but a negative routine BRCA1/2 screen.

4.6.1 Method

Genomic DNA from 55 subjects recruited into the above study with strong family histories of breast and/or ovarian cancer consistent with an underlying BRCA1/2 abnormality was analysed at the BRCA1 A4837G and T4308C SNPs using Pyrosequencing™ technology. CDNA was subsequently synthesized (random hexamer method) for all subjects who proved to be informative (heterozygous) at either SNP. A4837G and T4308C SNP allelic distribution in DNA and cDNA samples using the methods described above.

4.6.2 Result

Twenty-seven of the 55 potential subjects were informative at both A4837G and T4308C SNPs. Analysis of allelic distribution in cDNA and DNA at both SNPs revealed no examples of complete allele dropout (data not shown).

4.7 Review of findings of Sharp *et al.* [171] in subjects U02 and U05

The Pyrosequencing™ analysis of genomic DNA and cDNA from a total of 41 patients with either known truncating mutations, unknown DNA variants or strong family histories of breast cancer failed to identify any examples of complete allele dropout. Following the demonstration with this technique of heterozygosity at SNPs in exons 13 and 16 in a subject (U02) reported by Sharp *et al.* as having monoallelic expression of BRCA1, I examined Sharp's work on BRCA1 in more detail [171].

Sharp *et al.* described the expression of SNPs located at c736, c2077, c2082 and c4837 in the genomic DNA and cDNA of 2 patients with strong family histories of breast cancer. In a subject with a DNA T to G mutation at c736 (U05), allele-specific RT-PCR and sequencing revealed only the presence of the wild-type allele in cDNA. Similar analysis of two exon 11 SNPs (c2077 and c2082) and an exon 16 SNP (c4387) in a second subject (U02) also showed heterozygosity at each locus in the DNA but cDNA expression of the wildtype allele only. I repeated analysis of DNA and cDNA of U02 and U05 at these SNPs.

4.7.1 Methods

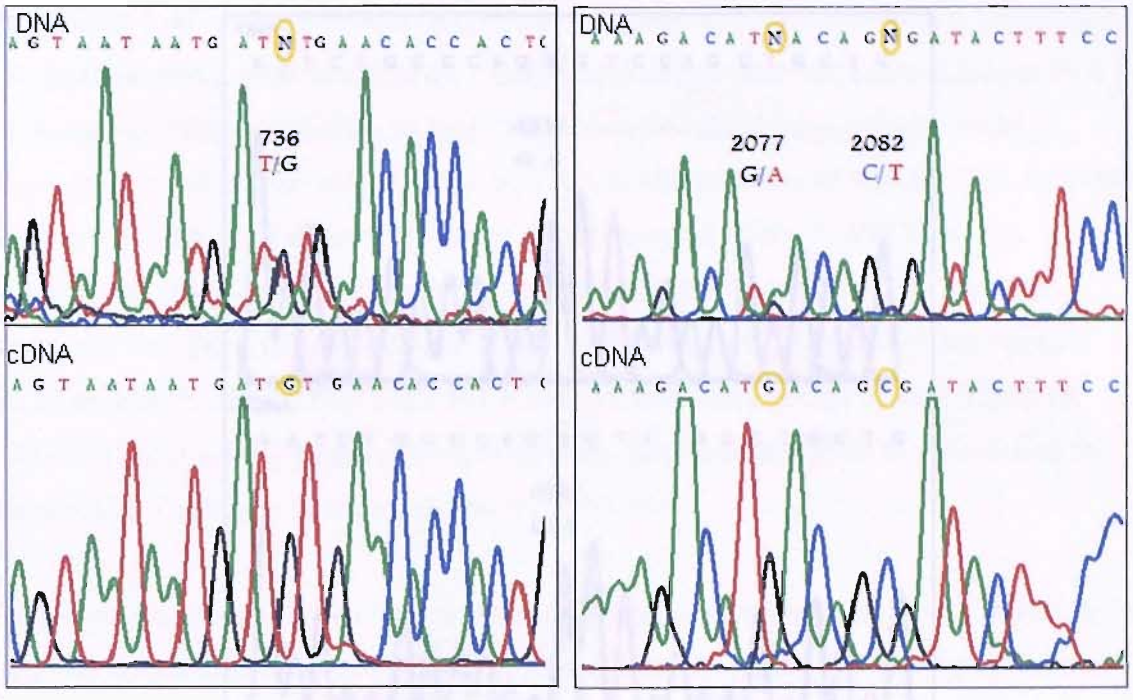
Primers of 15-25bp in length were designed to amplify DNA/cDNA in amplicons of 100-250bp in length across the site of each SNP (see table 4.8.1). To ensure that analysis of the cDNA sequence was not affected by contamination with DNA, primers for cDNA amplification were designed to be trans-exonic. PCRs to amplify genomic DNA from subjects U02 and U05 using the appropriate primer pairs were performed using AmpliTaq Gold Polymerase as described in section 3.3.5.3. PCRs to amplify cDNA from both subjects were performed using Platinum Taq Polymerase as described in the same section. PCR conditions were 94°C for 2 minutes; 35 cycles with denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds and elongation at 72°C for 30 seconds; 1 cycle at 72°C for 7 minutes; and a final hold at 15°C. PCR products were analysed on agarose gels to ensure that the amplicons sizes were correct. The c/DNA bands were excised, extracted with GENECLEAN® and then sequenced using the amplification primers as sequencing primers.

SNPs to be analysed	Forward and reverse primers for genomic DNA amplification	Forward and reverse primers for cDNA amplification
T736G	BRCA1.11SNPA F tgtaattttctgagacggatg BRCA1.11SNPA R gctgtaatgagctggcatga	BRCA1 RTx1-11F ctcgctgagacttcctggac BRCA1 RTx1-11R atggctccacatgcaagttt
G2077A C2082T	BRCA1.11SNPB F gcaactggagccaagaagag BRCA1.11SNPB R gatctttggggtcttcagca	BRCA1 RTx1-11F ctcgctgagacttcctggac BRCA1 RTx1-11R atggctccacatgcaagttt

Table 4.8: *Primer sequences used to amplify genomic DNA and cDNA sequences containing the SNPs T736G, G2077A, C2082T and A4837G*

4.7.2 Results

Sequencing of DNA and cDNA of U02 at the c736 SNP (exon 11) confirmed the presence of T/G heterozygous expression in DNA at this locus. CDNA was however homozygous for the non-wildtype allele (G). Analysis of DNA and cDNA of U05 at the c2077 and c2082 SNPs confirmed Sharp *et al.*'s report of heterozygosity in the DNA (G/A and C/T respectively) but not the cDNA (G and C). Sequencing of U05 DNA and cDNA at the exon 16 SNP (c4837) however showed heterozygous expression (G/A) in both DNA and cDNA.



a) **b)**

Figures 4.5 a and b: *Results of sequencing of DNA and cDNA from U02 and U05 at sites of SNPs in BRCA1 exon 11. U05 genomic DNA is heterozygous (G/T) at the c736 locus but U02 cDNA shows only the non-wildtype allele, G (figure a). U02 genomic DNA is heterozygous at c2077 (G/A) and c2082 (C/T); cDNA shows only the wildtype sequence with no evidence of the polymorphic variants (figure b).*

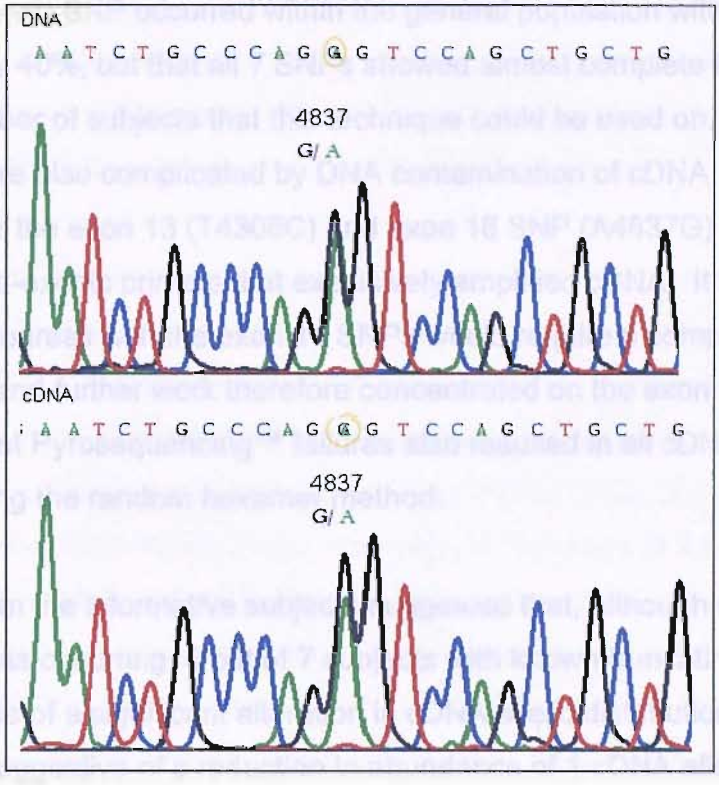


Figure 4.6: *Sequencing of U02 at the c4837 (exon 16) locus shows expression of both G and A alleles in DNA and cDNA*

4.8 Discussion

Pyrosequencing™ is a rapid and technically simple method of determining genotype and allelic distribution within short to medium sequences of DNA and cDNA. Studies have shown it to be a reliable and accurate technology [191]. It was anticipated that Pyrosequencing™ might prove to be a quick method of detecting NMD or impaired transcription of mutant allele cDNA transcripts, as evidenced by a significant reduction in the abundance of one allele or even complete loss of one allele.

The seven most common SNPs within BRCA1 were identified. In order to be informative, subjects had to be heterozygous at these SNPs. Investigations

confirmed that each SNP occurred within the general population with a frequency of approximately 40%, but that all 7 SNPs showed almost complete linkage, thus limiting the number of subjects that this technique could be used on. Initial experiments were also complicated by DNA contamination of cDNA. This problem was resolved for the exon 13 (T4308C) and exon 16 SNP (A4837G) by the adoption of trans-exonic primers that exclusively amplified cDNA. It became apparent that progress with the exon 11 SNPs would require a complex nested PCR approach and further work therefore concentrated on the exon 13 and 16 SNPs. Recurrent Pyrosequencing™ failures also resulted in all cDNAs being re-synthesised using the random hexamer method.

Initial results from the informative subjects suggested that, although no complete allele dropout was occurring, 6 out of 7 subjects with known truncating mutations showed evidence of a significant alteration in cDNA allelic distribution at the T4308C SNP, suggestive of a reduction in abundance of 1 cDNA allele. However, seven control cDNA samples showed a similar phenomenon. It was hypothesised that the alteration in allelic distribution could be the result of RNA degradation. Studies investigating the effect of a time delay in extracting RNA on allelic distribution at T4308C in normal subjects did show a clear association between increasing length of time from venesection to RNA extraction and alteration in allelic ratio. This effect was significantly less marked in cDNA synthesised from blood collected into PAXgene™ RNA preserving blood tubes. However, studies of RNA quality for the control samples showed that the abnormal control results could not easily be explained by poor quality RNA.

It is possible that the cDNA allelic imbalances seen in 6 of the 7 patients with known truncating mutations do represent an element of NMD of the mutant allele. Although the possibility of an unidentified and unsuspected BRCA1/2 mutation within the anonymous control population cannot be totally excluded, the frequency of BRCA1/2 mutations within the general population is only 5% and the incidence of 7 abnormal results out of 38 controls effectively rules this out as an explanation.

One alternative hypothesis is that the controls exhibiting unequal allelic expression carry rarer SNPs within a promoter region which are influencing the expression of 1 allele at the expense of the other. This explanation however could obviously equally apply to the subjects with changes in allelic frequency.

These results indicate that subtle changes in allelic frequency, as detected by Pyrosequencing™ cannot be used as a reliable indicator of underlying truncating or missense BRCA1 mutations. However, Sharp *et al.* [171] recently reported the putative occurrence of monoallelic cDNA expression in 2 patients with BRCA1 unknown variants. This technique could therefore still be of use if it were established that complete allele dropout is a regular sequelae of a BRCA1 mutation with pathogenic potential. To date I have analysed 7 patients with known truncating mutations, 4 with established unknown variants and 27 with family histories suggestive of an inherited BRCA1/2 gene abnormality but negative BRCA1/2 routine screens. No examples of complete allele dropout have been found. Interestingly, one of the subjects described by Sharp *et al.* [171] as showing monoallelic expression of BRCA1 was in fact found to be heterozygous at T4308C and A4837G on pyrosequencing™ analysis of cDNA. Subsequent analysis of this subject's DNA and cDNA using direct sequencing confirmed expression of both A and G alleles at this exon 16 SNP, although analysis of the two exon 11 SNPs examined by Sharp *et al.* in this subject verified their finding of loss of heterozygosity at these loci. The persistence of dual allelic expression at the exon 16 SNP indicates that subject U02 does not exhibit monoallelic expression of BRCA1. Re-analysis of the second subject described by Sharp as showing monoallelic BRCA1 expression (U05) confirmed the previous observation of loss of heterozygosity at c736, but my results showed persistence of the non-wildtype allele.

These results suggest that whilst loss of the whole or part of exon 11 of BRCA1 from one allele can occur, monoallelic expression of the entire germline gene is not yet a confirmed phenomenon. Review of the literature reveals no other confirmed

reports of inheritable complete monoallelic BRCA1 expression. Loss of heterozygosity at the BRCA1 locus in tumour cells has been found in breast and ovarian carcinoma specimens, but occurs in both sporadic cancers and those with proven underlying BRCA1 mutations [200, 201]. It is notable that each of the three exon 11 SNPs found here to demonstrate loss of an allele in cDNA lies within the c788-4096del (partial deletion of exon 11) or c548-4096del (deletion of exons 9, 10 and 11) regions that are absent in commonly occurring alternative transcripts of BRCA1 [149]. Similarly, the one other published report of heterozygous expression of a polymorphism in DNA but transcription of only one of the two alleles also refers to a SNP (T2312C) present in the region of exon 11 deleted in these alternative transcripts [202]. Therefore, it could be hypothesised that the failure to express both alleles of these exon 11 SNPs is due to selective loss of an allele within a common alternative transcript. This could be the result of these or linked SNPs resulting in up-regulation of cryptic splice sites by alterations of ESEs or ESIs.

To examine this hypothesis further and identify the length of the missing transcript, one could try to identify additional exonic SNPs 5' of c4215 and 3' of the exon 10/11 boundary which are present in the heterozygous form in U02 and U05 DNA. It would also be useful to analyse the DNA and cDNA of a number of controls at the exon 11 SNPs; similar examples of loss of the alternative alleles would give weight to the theory that these alleles are lost in the formation of the commonly occurring alternative transcripts described above. It is interesting that in the case of U02, the wildtype version of the analysed region of exon 11 was preserved, whilst in U05, the non-wildtype allele was expressed, suggesting that at certain sites the non wildtype allele may provide evolutionary advantages over the wildtype.

4.9 Conclusions

The results indicate that analysis of altered allelic expression using Pyrosequencing™ technology is not a reliable method of detecting BRCA1 mutations and is certainly not a candidate as a rapid screen for missense mutations not currently detected by routine screening. The development of the technique of Conformation Sensitive Capillary Electrophoresis is a far more promising strategy and the decision was therefore made to adopt this technique instead as a method of detecting previously undiagnosed BRCA1 missense mutations. Pyrosequencing™ is however a reliable, rapid and cost-effective method of establishing genotype at a known SNP locus.

My results also provide evidence against a previous report of complete monoallelic expression of BRCA1 and suggest that these observations may be explained by the well documented occurrence of alternative splicing of BRCA1 in the exon 11 region.

5.0 The Use of Multiplex RT-PCRs To Assess Splicing Abnormalities in BRCA1 and Identification of BRCA1 Missense Mutations using Conformation Sensitive Capillary Electrophoresis

5.1 Introduction

As described in the introduction, accurate assessment of the pathogenic potential of missense mutations via examination of their effect on splicing requires direct analysis of mRNA. One potentially simple method of identifying splicing abnormalities is to examine the RT-PCR products from a series of overlapping primers designed to cover the entire coding regions of BRCA1. The RT-PCR products can be rapidly analysed by gel electrophoresis and any alternative transcripts should be identifiable as PCR products of an unpredicted molecular weight. Abnormal PCR products can be sequenced to determine the nature of the rearrangement and to identify the likely causative mutation. Confirmation of a sequence change can then be obtained by sequencing the appropriate region in the matched DNA sample.

Such an approach has the advantage of being rapid and easily reproducible. It also has the potential to identify alternative transcripts resulting from intronic as well as exonic sequence variations. To date, this method has been used to analyse portions of the BRCA1 or BRCA2 mRNA transcript from a very small number of patients carrying BRCA1/2 missense mutations but, at the time of commencing this work there had been no published articles describing this process as a method of screening for aberrant splicing patterns caused by DNA sequence variants.

This piece of work was limited by lack of availability of RNA from patients with known missense mutations. As the project progressed, the decision was therefore made to use the conformational sensitive capillary electrophoresis system recently developed by the NGRL (Wessex) to identify additional patients with BRCA1 missense mutations. CDNA of these patients, where available, was then also analysed for aberrant splicing patterns using the multiplex RT-PCR method.

Conformation sensitive capillary electrophoresis (CSCE) is a relatively new method of indirect mutation detection that is technically straightforward, rapid and sensitive, and ideally suited to a high throughput system. The technique is based on the principle that the electrophoretic mobility of homoduplex DNA differs from that of heteroduplex DNA. A homozygous amplification product can only assume one single conformation post denaturation and re-annealing. However, when a heterozygous PCR product with a single base difference denatures and re-anneals there are four possible results: 2 heterozygous with a single base pair mismatch, 1 mutant homozygous and 1 homozygous wildtype. These four different duplex species are represented by 4 separate peaks on a CSCE analysis trace whilst the homozygous (wildtype) product produces a single peak. In practice a mutant product often produces an erratic waveform due to combination of peaks rather than 4 distinct peaks. Evidence suggests that this method is capable of detecting almost 100% of genomic DNA sequence variants [203].

The large size of the BRCA1 and BRCA2 genes and the failure of routine diagnostic techniques to examine the entire genetic sequence make them ideal candidates for this type of analytical approach. Additionally, CSCE performs optimally on genetic sequences with a GC percentage of approximately 25-60%, which completely encompasses the range exhibited by BRCA1/2 exons (27%-54%). The use of a CSCE based approach to screen BRCA1 and BRCA2 was first reported by Esteban-Cardenosa *et al.* [204]. To date they have screened 598 DNA samples from 431 unrelated families and have detected 114 different DNA variants, including 18 BRCA1 missense mutations, 5 BRCA1 splicing variants and 11 intronic BRCA1 variants [204]. Notably 26% of the detected BRCA1/2 sequence variants had not been detected on previous routine screening.

The National Genetics Reference Laboratory (NGRL), Wessex, has now developed a CSCE system to screen all 24 exons of BRCA1 using 33 separate fragments. A single PCR reaction is used to amplify each exon with the exception of exon 11, which is amplified as 12 overlapping fragments. All gene specific

primer pairs have been designed to carry one forward and one reverse universal primer (M13). The amplification reaction is performed as a single PCR using one pair of gene specific primers and a pair of universal primers, with the forward universal primer carrying a fluorescent tag (see figure 5.1). This permits further adaptation of this PCR for a wide range of downstream analysis techniques.

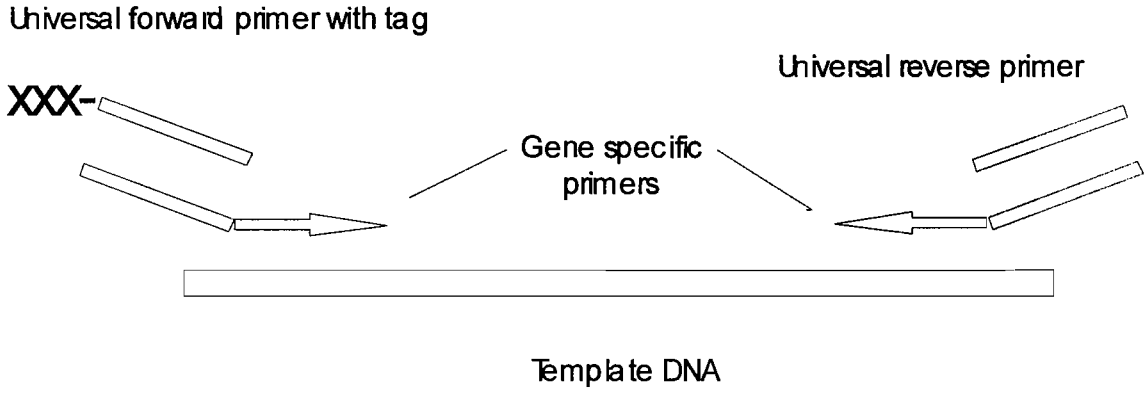


Figure 5.1: *The four primer system used to amplify BRCA1/2 gene fragments prior to analysis by CSCE, (with thanks to NGRL, Wessex)*

A pilot study of the NGRL CSCE system analysed 87 samples comprising 65 BRCA1 mutations within 6 different amplicons of exon 11 and 22 wild-type sequences. Of the samples yielding data deemed fit for analysis all mutations were detected with no false positives (see table 5.1.) This technique potentially provides a rapid and efficient means of searching the entire exonic BRCA1 sequence for missense mutations in patients with suspected BRCA1 genetic abnormalities who have previously had a negative routine BRCA1 limited screen.

Type of mutation	No. of samples	Fail		Call	
		No.	%	No.	%
Insertion	5	1	20	4	100
Deletion	32	3	9	29	100
Point	34	0	0	34	100
Normal	22	3	14	19	100
Total	87	7	8	80	100

Table 5.1: *Results of pilot study to assess ability of CSCE to identify 65 different mutations within 6 BRCA1 amplicons in 87 DNA samples (with thanks to NGRL, Wessex)*

In view of the evolutionary evidence that suggests that the 5' region of exon 11 is highly conserved, it is feasible that missense mutations with a pathogenic potential may be at a higher frequency in this region and I therefore initially searched this part of the gene. During the course of my project the WRGL diagnostic group used CSCE to analyse the entire coding sequence of the BRCA1 gene of 471 patients using the SCOBEC high throughput genetic screening facility. Inevitably this technique detects polymorphisms as well as missense (and nonsense/ frameshift) mutations.

5.2 Methods: CSCE identification of BRCA1 missense mutations

5.2.1 Subjects

Potential subjects were identified from a database of the Wessex Regional Genetics Service detailing patients with strong family histories of breast/ ovarian cancer in a pattern consistent with an underlying BRCA1 or BRCA2 pathogenic mutation who have previously undergone diagnostic screening for BRCA1/2 mutations with a "negative" result. All patients on this database were assessed to determine their relative risk of carrying a BRCA1 or BRCA2 mutation using the Manchester scoring system [205]. The 94 highest scoring patients for whom

genomic DNA was stored within the Wessex Regional Genetics Laboratory were selected for this study (see Appendix E).

5.2.2 Preparation of DNA samples

The concentration of the genomic DNA samples was assessed using a NanoDrop® ND-1000 spectrophotometer (NanoDrop technologies, Delaware, USA). A 1.6µl aliquot of each DNA sample was pipetted onto the lower measurement pedestal of the NanoDrop® Spectrophotometer and the DNA concentration was analysed using NanoDrop® software (version 3.1.0). Each DNA sample was diluted to a concentration of 10ng/µl using sterile water. The NanoDrop® Spectrophotometer was used to verify the final concentration of each sample. 4µl of each dilute DNA sample was pipetted into a well of a 96 well PCR plate and the DNA was lyophilised by heating the PCR plate at 40°C for 3 hours. The PCR plate was sealed with an adhesive cover and kept at room temperature until required.

5.2.3 Primers

Primers were designed by the National Genetics Reference Laboratory (Wessex) with the aim of amplifying exon 11 of BRCA 1 in 12 separate, overlapping segments:

Frag-ment	Forward and Reverse Primers	Sequences of gene specific primers	*Size of amplicon (bp)*	Start (HUGO)	End (HUGO)
11A	BC1_11AF_ext_M13F_02	tgagctacatcttcagtatacttggt	*439*	671-152	671-127
	BC1_11AR_ext_M13R_02	Gttatgttggctccttgctaa		937	957
11B	BC1_11BF_ext_M13F_01	cattacagcatgagaacagcagt	*407*	845	867
	BC1_11BR_ext_M13R_02	tttagaacgtccaatacatcagc		1228	1251
11C	BC1_11CF_ext_M13F_01	ccagaagtgatgaactgttaggt	*371*	1160	1182
	BC1_11CR_ext_M13R_01	tgatgtaggtctccttttacgc		1509	1530
11D	BC1_11DF_ext_M13R_02	aatacaagagcgtcccctcac	*435*	1479	1496
	BC1_11DR_ext_M13F_02	tcagtacaattaggtgggcttagatt		1888	1913
11E	BC1_11EF_ext_M13F_01	tatccacaattcaaaagcaccta	*268*	1796	1819
	BC1_11ER_ext_M13R_01	ygttaacttcagctctgggaaag		2040	2063
11F	BC1_11FF_ext_M13R_01	gcaactggagccaagaaga	*472*	1961	1983
	BC1_11FR_ext_M13F_01	tgcacactgactcacacatttat		2411	2432
11G	BC1_11GF_ext_M13R_01	tggtactgattatggcactca	*381*	2319	2339
	BC1_11GR_ext_M13F_01	gtgacttttggactttgtttctt		2677	2699
11H	BC1_11HF_ext_M13F_01	ggaaatgcagaagaggaatgtg	*345*	2626	2647
	BC1_11HR_ext_M13R_01	aacaratgacttgatgggaaaaa		2948	2970
11I	BC1_11IF_ext_M13R_01	ggcaacgaaactggactca	*394*	2878	2896
	BC1_11IR_ext_M13F_01	gttgcaaaacccctaatctaagc		3249	3271
11J	BC1_11JF_ext_M13R_02	aagcagaactaggtagaaacaga	*466*	3206	3228
	BC1_11JR_ext_M13F_02	ggaagctcttcatcctcac		3653	3671
11K	BC1_11KF_ext_M13F_01	araggagagcttagcaggagtc	*382*	3547	3568
	BC1_11KR_ext_M13R_01	tatttgcagtcaagtcttccaat		3906	3928
11L	BC1_11LF_ext_M13F_02	accgttgctaccgagtgtct	*307*	3818	3839
	BC1_11LR_ext_M13R_02	tgtaaaatgtgctccccaaaag		4096+10	4096+28

Table 5.2 *Primer sequences for amplification of BRCA1 exon 11 fragments for CSCE analysis*

5.2.4 PCR for CSCE

A PCR reaction mix was prepared by adding 5µl of AmpliTaq gold PCR master mix (2x) to 1.6µl of 50% glycerol, 1µl of a gene specific primer mix containing forward and reverse primers at a concentration of 1pmol/µl each, 0.5µl of a M13 primer mix

containing a 6-carboxy fluoroscein (FAM) labelled forward primer and an unlabelled reverse primer at a concentration of 10pmol/µl each and 1.9µl sterile water per DNA sample. 10µl of the PCR reaction mix was added to 40ng DNA lyophilised in the well of a PCR plate. The PCR plate was sealed with an adhesive cover and agitated gently. The plate was then placed in a thermal cycler and PCR conditions were 95°C for 10 min; 30 cycles with denaturation at 95°C for 5s, annealing at 60°C for 30s and elongation at 72°C for 60s, followed by 10 cycles with denaturation at 95°C for 5s, annealing at 55°C for 30s and elongation at 72°C for 60s followed by 1 cycle at 72°C for 5 minutes and a hold at 15°C. This was followed by heteroduplex formation commencing with heating to 96°C for 10 minutes followed by 1 cycle of heating at 96°C for 20s and a further 79 cycles in which the maximum temperature was dropped by 0.5°C with each cycle. This was completed with a hold at 15°C. 2µl of each PCR product was diluted with 18µl of sterile water and 2µl of this product was diluted with a further 18µl of water to produce a final concentration of one hundredth of the original PCR product. 20µl of the dilute PCR product was loaded onto the Applied Biosystems 3730 DNA Analyser.

5.2.5 Analysis of CSCE results

Analysis was performed using GeneMapper software (version 3.7, Applied Biosystems). Waveforms for each sample were examined by eye (see figure 6.2 for examples). The peak height (measure of intensity) of each CSCE result was adjusted as required to permit accurate examination of the profile. Any with a peak with a maximum intensity of less than 350 units was considered to have failed. All DNA samples resulting in a shift were subsequently sequenced using the forward primer for the appropriate fragment of exon 11.

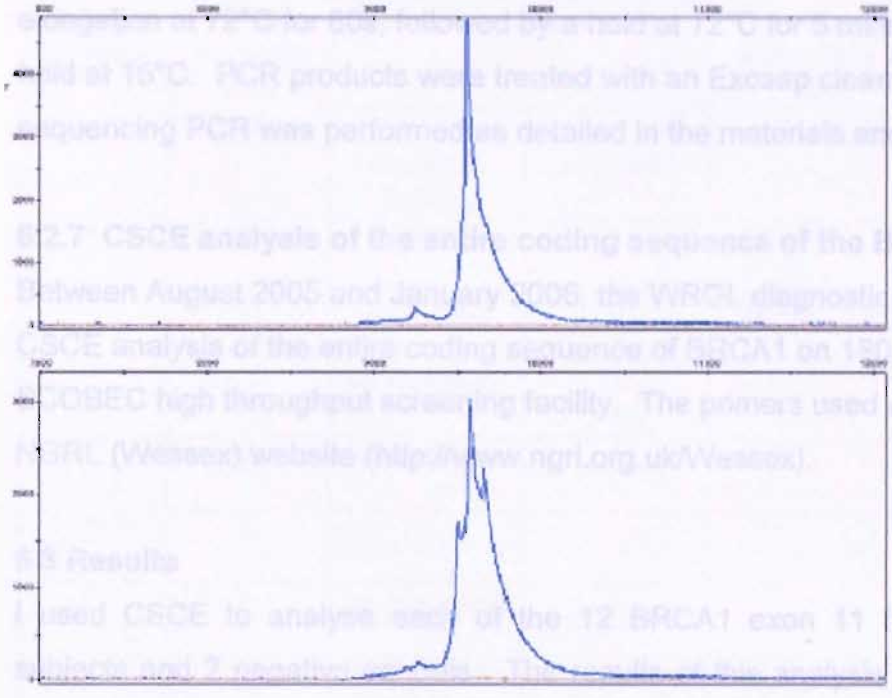


Figure 5.2: *Sample results from CSCE analysis of BRCA1 exon 11, fragment I*
Upper figure shows wildtype DNA with single CSCE peak
Lower figure shows heterozygous sequence with multiple peaks in CSCE waveform (subject N10).

All fragments producing shifts on CSCE analysis were sequenced using both forward and reverse gene specific primers. Any detected sequence variations were reviewed using the BIC database and any potential UVs were confirmed by repeat sequencing.

5.2.6 Gene specific PCR for sequencing

2µl of the target genomic DNA (10 ng/µl) was added to 5µl AmpliTaq gold PCR master mix, 1.6µl 50% glycerol and 0.2µl of a gene specific primer mix (forward and reverse primers at a concentration of 10pm/µl each). Sterile water was added to make the final volume up to 10µl. The reaction mix was transferred to a 96 well PCR plate which was placed in a thermal cycler. PCR conditions were 95°C for 10 min; 40 cycles with denaturation at 95°C for 5s, annealing at 60°C for 30s and

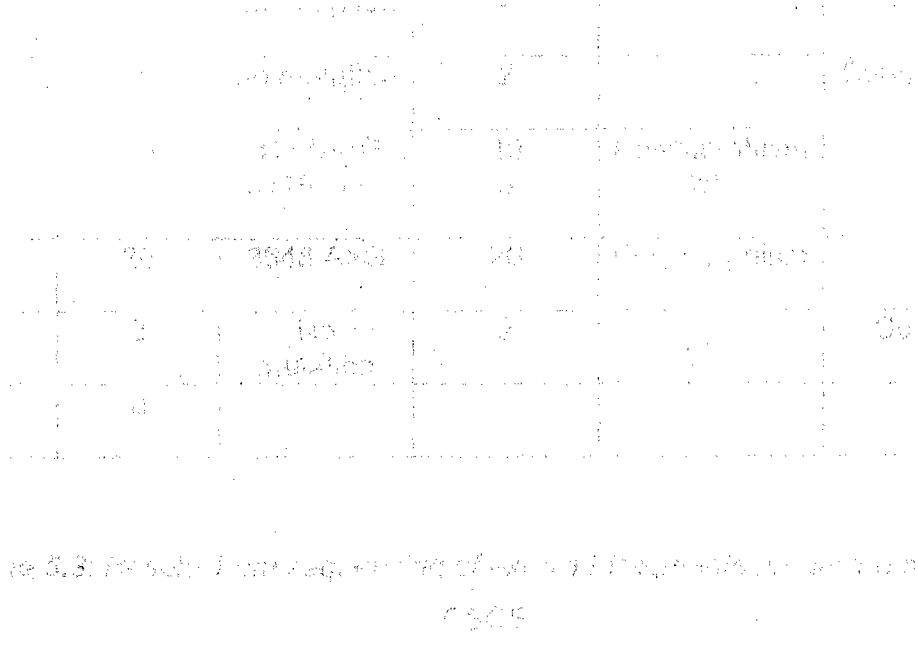
elongation at 72°C for 60s, followed by a hold at 72°C for 5 minutes and a final hold at 15°C. PCR products were treated with an Exosap clean up and a sequencing PCR was performed as detailed in the materials and methods section.

5.2.7 CSCE analysis of the entire coding sequence of the BRCA1 gene

Between August 2005 and January 2006, the WRGL diagnostic group performed CSCE analysis of the entire coding sequence of BRCA1 on 180 patients using the SCOBEC high throughput screening facility. The primers used are provided on the NGRL (Wessex) website (http://www.ngrl.org.uk/Wessex).

5.3 Results

I used CSCE to analyse each of the 12 BRCA1 exon 11 fragments in all 94 subjects and 2 negative controls. The results of this analysis are summarised in the table 5.3.



Location of shift	No. of subjects with shifts	Results of Sequencing		Classification according to BIC	Comment
		Sequence changes (ATGn)	No. of subjects		
A	1	736T>G	1	UV	
B	11	1067A>G no mutation	9 2	Polymorphism	 CSCE peak <400
C	2	no mutation	2		Subtle CSCE shift
D	0				
E	2	no mutation	2		CSCE peak <400
F	31	2077 G>A 2082 C>T 2083 G>T 2311 T>C no mutation	8 24 1 23 1	Polymorphism Polymorphism UV Polymorphism	 CSCE peak<400
G	41	2612 C>T no mutation	37 4	Polymorphism	 Low CSCE peaks
H	2	no mutation	2		Subtle shift only
I	43	3113A>G 3119 G>A	40 3	Polymorphism UV	
J	26	3548 A>G	26	Polymorphism	
K	2	No mutation	2		Subtle shift
L	0				

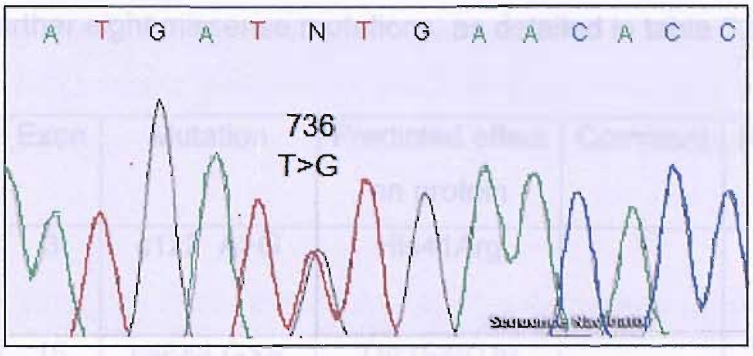
***Table 5.3**: Results from sequencing of exon 11 fragments producing shifts on CSCE*

A total of 161 exon 11 fragments were sequenced (see tables 5.3 and 5.4). The vast majority of these proved to be harbouring polymorphisms, but 3 distinct missense mutations were confirmed in five subjects: T>G at c736, G>T at c2083 and G>A at c3119.

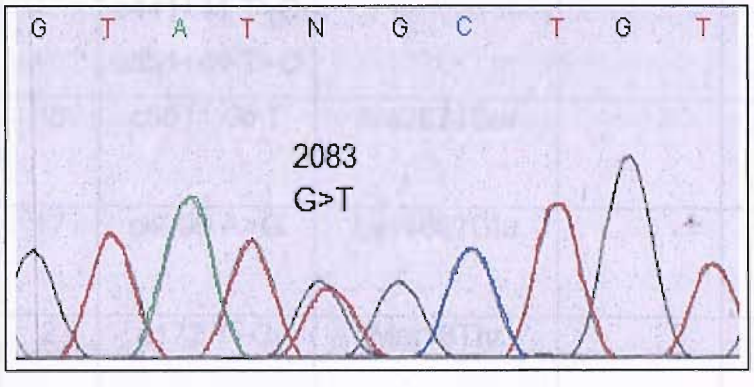
Subject	Mutation: ATGn	Mutation: Ref.seq no.	Predicted effect	No. of reports to BIC	RNA available
H12	c736 T>G	855 T>G	Leu > Val	62	N
C07	c2083 G>T	2202 G>T	Asp >Tyr	2	N
B04	c3119 G>A	3238G>A	Ser>Asn	23	Y
F02	c3119 G>A	3238G>A	Ser>Asn	23	N
F11	c3119 G>A	3238G>A	Ser>Asn	23	N

***Table 5.4**: Exon 11 missense mutations identified in 94 patients with negative routine BRCA1/2 screens using CSCE and sequencing*

a)



b)



c)

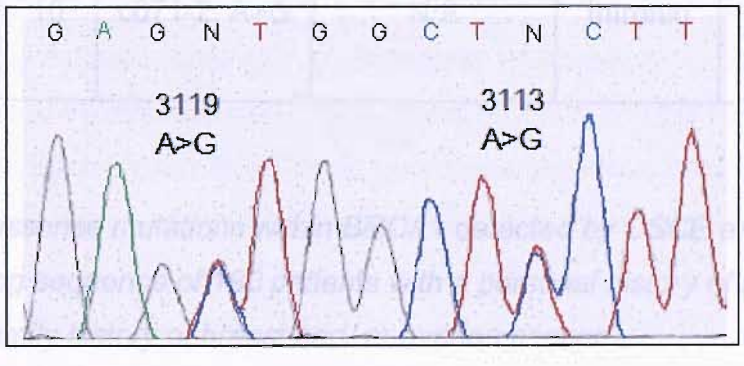


Figure 5.3: *Sequencing traces of patients a) H12, b) C07 and c) B04 (5.3c shows reverse complementary trace)*

Analysis by the WRGL of the entire BRCA1 coding sequence in 180 patients identified a further eight missense mutations, as detailed in table 5.5.

Subject	Exon	Mutation	Predicted effect on protein	Comment	RNA available
SAL03	3	c122 A>G	His41Arg		Y
SAL17	15	c4644 G>A	Thr1548Thr		Y
SAL34	7	c441+39 T>C c441+41 T>C c551+44 T>C	N/a	Intronic	Y
SAL39	18	c8011 G>T	Ala2671Ser		N
SAL71	17	c4999 A>G	Lys1667Glu		Y
SAL89	2	c172 T>C	Met18Thr		N
SAL188	11	c2060 A>C	Glyn 687Pro		N
J01	10	c671-2 A>G	N/A	Intronic	Y

Table 5.5: *Missense mutations within BRCA1 detected by CSCE analysis of the BRCA1 coding sequence of 180 patients with a personal history of breast cancer and strong family history of breast and/ or ovarian cancer*

All patients with missense mutations identified by CSCE and for whom RNA was available were subsequently analysed using a system of multiplex RT-PCRs. Additionally, RNA became available for SAL205, a relative of SAL 03 who proved to carry the same missense mutation (c122A>G) on predictive screening of exon 3 by CSCE.

5.4 Methods: Mutiplex PCRs of BRCA1 using overlapping primers

5.4.1 Subjects

Subjects were selected from the following categories:

a) Patients with a previously identified protein truncating mutation of BRCA1

b) Patients with a known DNA sequence variant predicted to disrupt splicing of BRCA1 due to its location within a splice donor/ splice acceptor site.

c) Patients with a DNA sequence variation of unknown significance of BRCA1 identified either on routine analysis or by CSCE

d) Patients with a personal history of breast cancer and strong family history of breast/ ovarian cancer who have previously had a negative routine BRCA1 and BRCA2 screen.

A mononuclear cell lysate preserved in Trizol had to be available for each subject and cDNA for each subject was synthesised using the random hexamer method as described in chapter 3. The majority of these subjects had previously been investigated by Pyrosequencing™, as described in chapter 4. A random group of control cDNAs were selected from the samples produced by the random hexamer method for the Pyrosequencing™ work, i.e. from anonymous patients referred to WRGL for investigation of non-neoplastic genetic disorders.

5.4.2 Primers

Primers were designed to amplify six overlapping sections of the BRCA1 gene, covering the entire coding region, namely exons 1-11, exons 11-13, exons 11-17, exons 15-24, exons 9-11 and exons 9-13, as detailed in the table below

Frag-ment	Primer	Sequence (5'-3')	Start (HUGO)	Finish (HUGO)	Size of amplicon
1-11	BRCA1 RTx1-11F	ctcgctgagacttcctggac	1-1273	1-1253	937bp
	BRCA1 RTx1-11R	atggctccacatgcaagttt	802	820	
11-13	BRCA1 RTx11-13F	tgcaggctttcctgtggttg	2777	2797	1520bp
	BRCA1 RTx11-13R	tgatggaagggtagctgttagaagg	4274	4297	
11-17	BRCA1 RTx11-17F	caagaagagcaagcatgga	4066	4085	951bp
	BRCA1 RTx11-17R	ggtgttttctggcaaacttgt	4997	5017	
15-24	BRCA1 RTx15-24F	ttgttgatgtggaggagcaa	4592	4611	1031bp
	BRCA1 RTx15-24R	cctgtggctctgtacctgtg	5604	5623	
9-11	BRCA1 RTx9-11F	acaaagcagcggatacaacc	499	518	2388bp
	BRCA1 RTx9-11R	tttcgttgcctctgaactga	2868	2887	
9-13	BRCA1 RTx9-11F	acaaagcagcggatacaacc	499	518	3798bp
	BRCA1 RTx11-13R	tgatggaagggtagctgttagaagg	2797	4297	

***Table 5.6**: Sequences of primers for multiplex RT-PCR of BRCA1 in 6 overlapping fragments*

5.4.3 PCRs

For fragments 1-11, 11-13, 11-17 and 15-24

2µl of the cDNA sample to be amplified was added to 5µl of 10x PCR buffer, 1.5µl 50mM $MgCl_2$, 1µl10mM dNTPs, 1µl 3' primer (10pmol/µl), 1µl 5' primer (10pmol/µl) and 0.4µl Platinum Taq Polymerase. The volume of the reaction was made up to 50µl with sterile water. The plate was then sealed with an adhesive PCR film before being placed in a PTC-0225 DNA Engine Tetrad thermal cycler (MJ research). PCR conditions were 94°C for 2 mins; 35 cycles with denaturation at 94°C for 30s, annealing at 60°C for 30s and elongation at 72°C for 30s; 1 cycle at 72°C for 7 min; and a final hold at 15°C

For fragments 9-11 and 9-13

2µl of the cDNA sample to be amplified was added to 5µl of 10x PCR buffer, 1.5µl 50mM $MgCl_2$, 1µl10mM dNTPs, 1µl 3' primer (10pmol/µl), 1µl 5' primer (10pmol/µl)

For fragments 9-11 and 9-13

2µl of the cDNA sample to be amplified was added to 5µl of 10x PCR buffer, 1.5µl 50mM $MgCl_2$, 1µl10mM dNTPs, 1µl 3' primer (10pmol/µl), 1µl 5' primer (10pmol/µl) and 0.5µl Platinum Taq Polymerase. The volume of the reaction was made up to 50µl with sterile water. The plate was then sealed with an adhesive PCR film before being placed in a PTC-0225 DNA Engine Tetrad thermal cycler (MJ research). PCR conditions for fragment 9-11 were 94°C for 2 mins; 35 cycles with denaturation at 94°C for 30s, annealing at 60°C for 30s and elongation at 72°C for 2 mins 30s; 1 cycle at 72°C for 7 min; and a final hold at 15°C. For fragment 9-13 the elongation period was extended to 4 minutes.

5.4.4 Analysis of transcripts

All PCR products were initially analysed by electrophoresis on a 2% agarose gel. Where agarose gel electrophoresis revealed more that one transcript, the PCR products were subsequently analysed using a 4.8% polyacrylamide gel. Any transcripts that were determined to be of a different size to the anticipated PCR product were excised from the polyacrylamide gel, extracted using GENECLEAN® (section 3.3.3-4) and then sequenced directly (3.3.6). Where direct sequencing was unsuccessful, the purified PCR products were cloned into pCR®4 or pGEM®-T Easy vectors, as described in section 3.3.8 and the resulting plasmids were sequenced.

5.5 Results

Good amplification of fragments 1-11, 11-13, 11-17 and 15-24 was obtained using the conditions detailed above. Amplification of fragments containing exons 9-11 and exons 9-13 (2388bp and 3798bp) proved technically difficult owing to the large sizes of these fragments. Altering PCR and thermal cycling conditions, (such as extending elongation time), and use of the high fidelity *PFU* DNA Polymerase (Invitrogen) under a variety of thermal cycling conditions did not improve product yield. Trials of specialist "long-range" PCR systems including "Dynazyme" PCR (Finnzymes) and Expand High Fidelity PCR System (Roche) were also

primer pairs successfully amplified the appropriate gene fragments using the PCR method and conditions described above for fragment 1-11.

Fragment	Primer	Sequence (5'-3')	Start (HUGO)	Finish (HUGO)	Size of amplicon
11a	BRCA1 RTx9-11AF	agctgagaggcatccagaaa	758	778	1284bp
	BRCA1 RTx9-11AR	ctcttcttggctccagttgc	2023	2042	
11b	BRCA1 RTx9-11bF	ggaactaacccaaacggagca	1603	1622	1357bp
	BRCA1 RTx9-11bR	tgatgggaaaaagtggtggt	2940	2959	

Table 5.7: *Details of additional two pairs of overlapping primers designed to amplify regions of BRCA1 gene between 1-11 and 11-13 amplicons*

Amplification of fragments 1-11, 11a, 11b, 11-13, 13-17 and 15-24 was initially performed on 10 control cDNA samples, 6 cDNA samples from subjects with known protein truncating mutations of BRCA1 or exonic deletions, 4 cDNAs from patients with a DNA mutation predicted to disrupt splicing, 5 cDNA samples from subjects with previously detected missense mutations and 21 subjects with strong family histories but negative routine BRCA1/2 screens, (see table 5.3). Seven other patients found to have BRCA1 missense mutations on CSCE analysis were subsequently also included.

Subject	Exon	Mutation	Comment	Detection method
T02	11	c1326 T>A	Nonsense mutation	PTT
T06	11	4bp del c3756	Frameshift	PTT
T07	11	c1141 A>T	Nonsense mutation	PTT
T11	10-12	Deletion of exons		MLPA
T12	11	4bp deletion c4065	Frameshift	PTT
T15	10-12	Deletion of exons		MLPA
S01	12	c4185 G>A	Predicted to affect splicing	Heteroduplex analysis
S02	12	c4185 G>A	Predicted to affect splicing (relative of SO1)	Heteroduplex analysis
S03	12	c4185 G>A	Predicted to affect splicing (relative of SO1)	Heteroduplex analysis
S04	12	c4185 G>A	Predicted to affect splicing (Not relative of SO1)	Heteroduplex analysis
U01	11	c1067 A>G	Missense mutation ?polymorphism	Heteroduplex analysis
U02	11	c2077 G>A	Missense mutation	Direct sequencing (predictive test)
U03	20	c5198 A>G	Missense mutation	Heteroduplex analysis
U04	2	c53 T>G	Missense mutation	Heteroduplex analysis
U05	11	c736 T>G	Missense mutation	Heteroduplex analysis
B04	11	c3119 G>A	Missense mutation	CSCE
SAL03	3	c122 A>G	Missense mutation	CSCE
SAL205	3	c122 A>G	Missense mutation	Direct sequencing (predictive test)
SAL17	15	c4644 G>A	Missense mutation	CSCE
SAL34	7	c441+39 T>C c441+41 T>C c551+44 T>C	Intronic	CSCE
SALL71	17	c4999 A>G	Missense mutation	CSCE
J01	10	c671-2 A>G	Intronic	CSCE
N01-21			No known BRCA1 mutation	Heteroduplex analysis and PTT of exon 11

Table 5.8: *Subjects analysed by multiplex RT-PCR of BRCA1*

Amplification of fragment 1-11

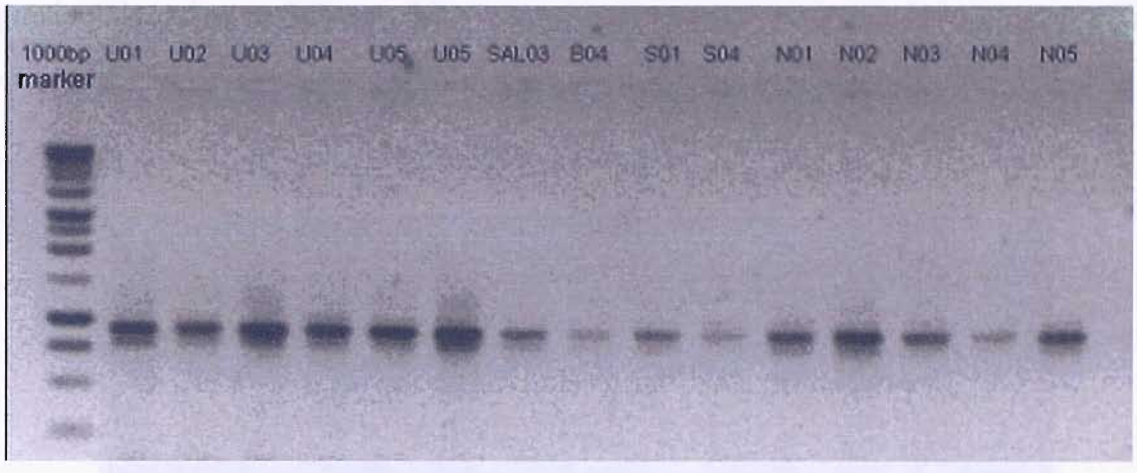


Figure 5.4: *Sample agarose gel electrophoresis of PCR products from amplification of BRCA1 fragment 1-11 for subjects U01-U05, SAL03, B04 and N01-N05. For each subject 2 cDNA bands are visible: a dominant 937bp band and a second very faint band of approximately 810bp*

In all controls and subjects, amplification of fragment 1-11 resulted in a dominant PCR product of the anticipated size of 937bp, but also an additional faint cDNA band approximately 120bp smaller than the presumed normal transcript (see figure 5.4). PCR products from this amplification were further analysed on a polyacrylamide gel, to permit improved separation of these two bands (see figure 5.5). Direct sequencing of the 937bp cDNA band in controls and a number of subjects confirmed that this represented the anticipated transcript of BRCA1 exons 1 to 11 extending between c1-1273 and c820 (the BRCA1 RT 1-11 primer positions). Direct sequencing of the ≈820bp cDNA band was unsuccessful but cloning of this cDNA band into pGEM T Easy® and sequencing of the resulting plasmids revealed that this band corresponded to the 1-11 BRCA1 transcript with complete deletion of exons 9 and 10 (total length = 814bp) in both controls and subjects. A third very faint band of approximately 790bp was also seen in all controls and subjects when large amounts of the PCR products were analysed.

Cloning and sequencing of this third transcript revealed that this represented a complete deletion of exons 9 and 10 with additional loss of the last (3') 22 nucleotides of exon 5 (figure 5.7).

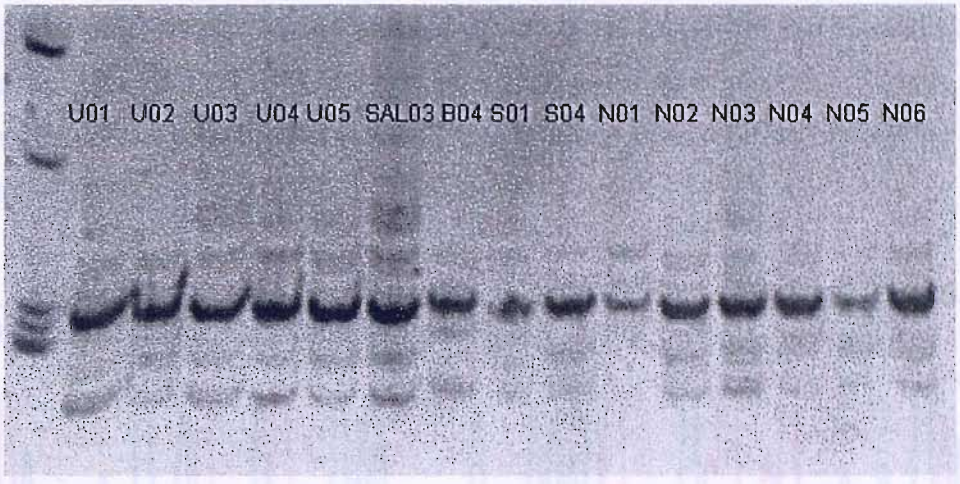


Figure 5.5: *Sample polyacrylamide gel electrophoresis of PCR products from amplification of BRCA1 fragment 1-11 for subjects U01-U05, SAL03, S01, S04 and N01-N06. For each subject a dominant 937bp band, a faint band of 814bp and a third band of approximately 792 are present. Although B04 appears in this picture to be producing an abnormal transcript pattern, subsequent repetitions of this analysis confirmed that no alternative splicing was taking place.*

The initial polyacrylamide gel electrophoresis analysis (PAGE) of the subjects detailed above suggested that amplification of SAL 03 and B04 cDNA with BRCA1 RT 1-11 primers resulted in transcripts of different sizes than other subjects and controls. Amplification of these samples was therefore repeated three times, and in the case of SAL 03, cDNA was also obtained from a relative, SAL205, carrying the same BRCA1 exon 3 missense mutation (c122 A>G). Subsequent PAGE analysis of these PCR products indicated that in fact, the initial result was not

the same BRCA1 exon 3 missense mutation (c122 A>G). Subsequent PAGE analysis of these PCR products indicated that in fact, the initial result was not reliable and that the transcripts obtained from these cDNA samples were of the same size as those from control samples. Each transcript obtained from the BRCA1 RT 1-11amplification of cDNA from BO4, SAL03 ad SAL205 was subsequently cloned into pGEM T Easy® and sequenced. This confirmed that all transcripts obtained from these samples represented either the 937bp wildtype 1-11 transcript, exons 1-11 with deletion of exons 9 and 10 or exons 1-11 with deletion of exons 9 and 10 and the 3' 22 nucleotides of exon 5 as previously observed in the controls.

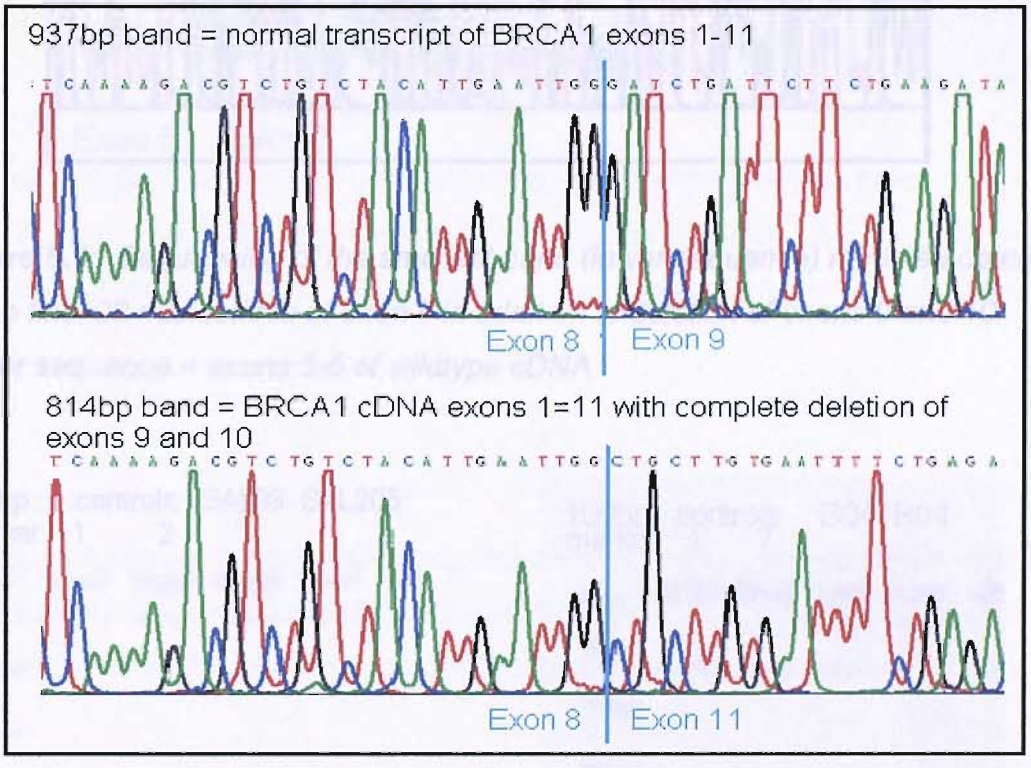


Figure 5.6: *Results of sequencing 937bp band and 814bp from amplification of BRCA1 fragment 1-11 in controls. The 937bp band contains the complete expected transcript of BRCA1 exons 1-11. The 815bp band contains a complete deletion of exons 9 and 10.*

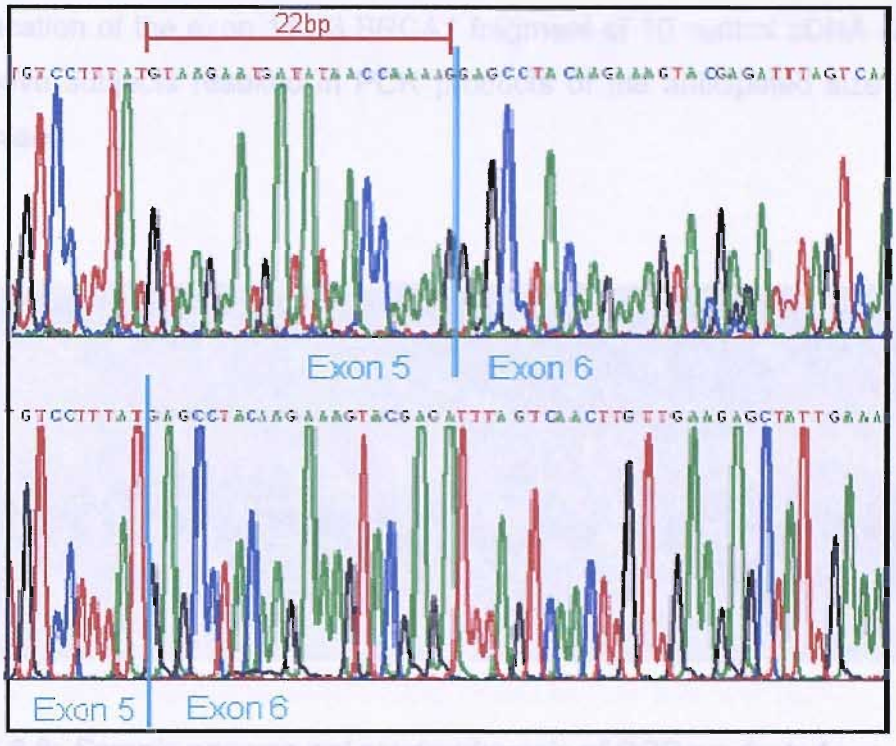


Figure 5.7: *Sequencing of the smallest band (lower sequence) revealed deletion of the final 22 nucleotides of exon 5 in addition to deletion of exons 9 and 10. Upper sequence = exons 5-6 of wildtype cDNA*

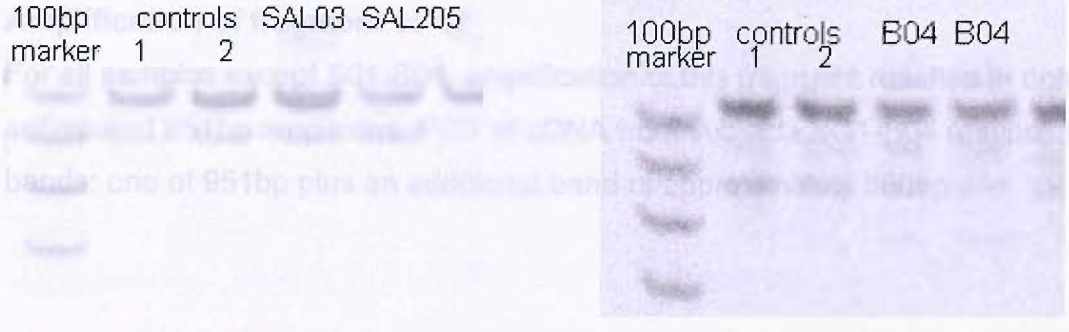


Figure 5.8: *Repeat amplification of fragment 1-11 using cDNA from subjects B04, SAL03 and SAL205 confirmed that the pattern of transcripts obtained was identical to that seen with control cDNA samples*

Amplification of fragment 11-13

Amplification of the exon 11-13 BRCA1 fragment of 10 control cDNA samples and the above subjects resulted in PCR products of the anticipated size (1520bp) in each case.

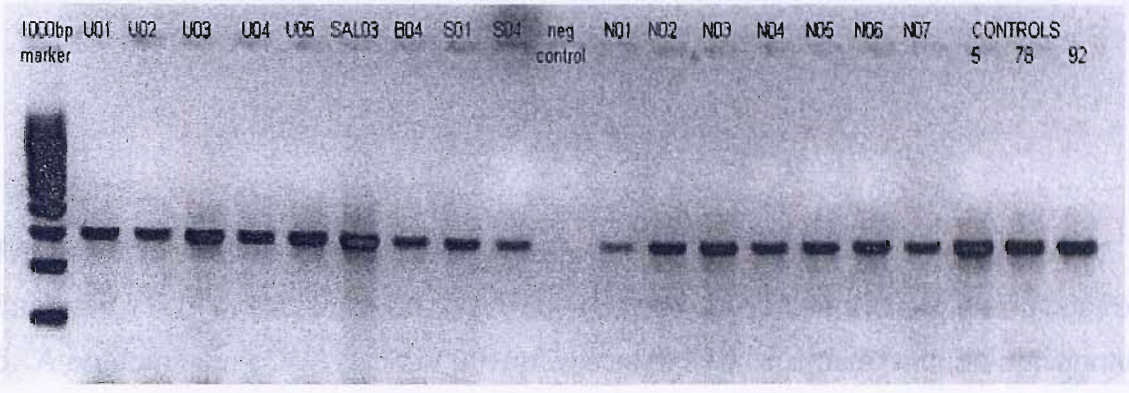


Figure 5.9: *Sample agarose gel electrophoresis of PCR products from amplification of BRCA1 fragment 11-13 for subjects U01-U05, SAL03, B04, N01-N07 and controls 5, 78 and 92. In each case a single cDNA band of 1520bp is seen.*

Amplification of fragment 11-17

For all samples except S01-S04, amplification of this fragment resulted in only the anticipated 951bp amplicons. PCR of cDNA from subjects SO1-S04 resulted in 2 bands: one of 951bp plus an additional band of approximately 860bp.

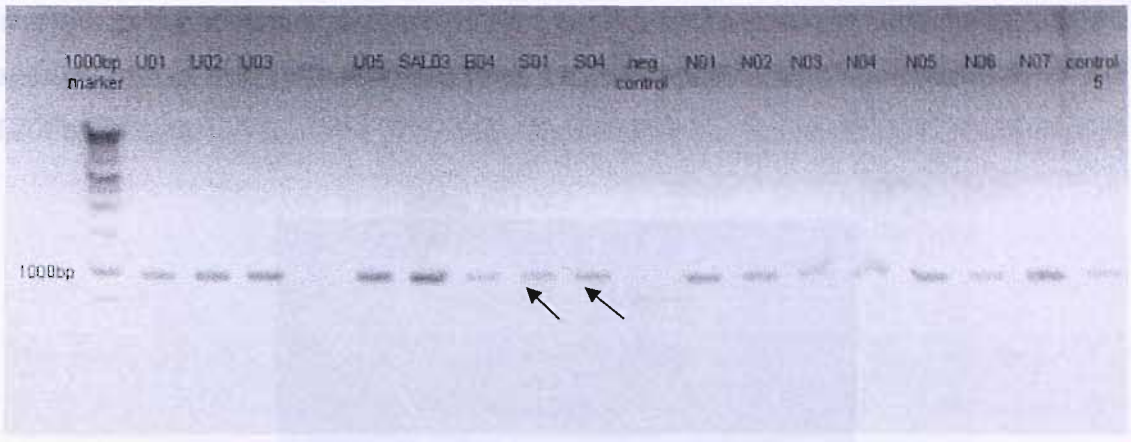


Figure: 5.10: *Agarose gel electrophoresis of PCR products from amplification of fragment 11-17 in subjects U01,2,3,5, SAL03, B04, S01, S04, N01-N07 and control 5. A predominant cDNA band of 951bp was seen in all subjects, with an additional very faint band of approximately 860bp visible in only S01 and S04 (indicated by ↖)*

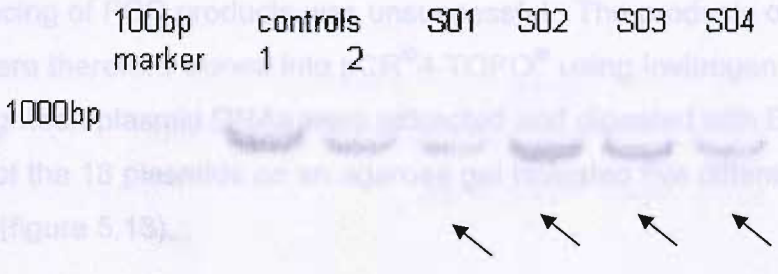


Figure 5.11: *Polyacrylamide gel analysis of the BRCA1 11-17 amplicons shows the additional 860bp cDNA band clearly present in S01, S02, S03 and S04 but not in controls 1 and 2 (indicated by ↖)*

In addition, amplification of the BRCA1 exon 11-17 fragment in subjects S01-S04 and in some controls resulted in a third very faint cDNA band approximately 1000bp in size.

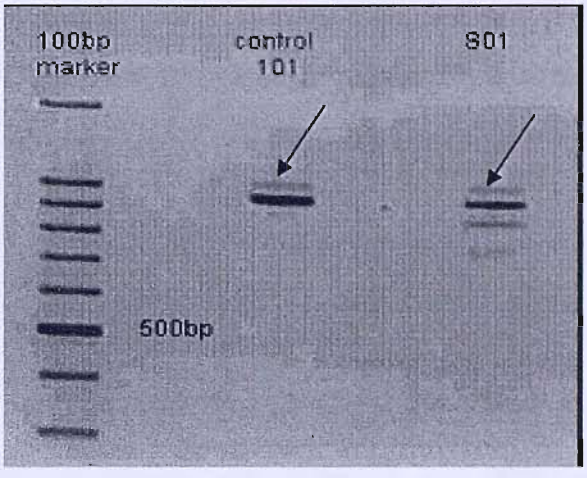


Figure 5.12: *Agarose gel analysis of BRCA1 11-17 amplicons from control 101 and S01. A third cDNA band of approximately 1000bp is visible for both subjects on this gel (indicated by ↙)*

Direct sequencing of PCR products was unsuccessful. The products of the S01 11-17 PCR were therefore cloned into pCR®4-TOPO® using Invitrogen TOPO TA® cloning kit. Eighteen plasmid DNAs were extracted and digested with EcoRI. Visualisation of the 18 plasmids on an agarose gel revealed five different sized cDNA inserts (figure 5.13).

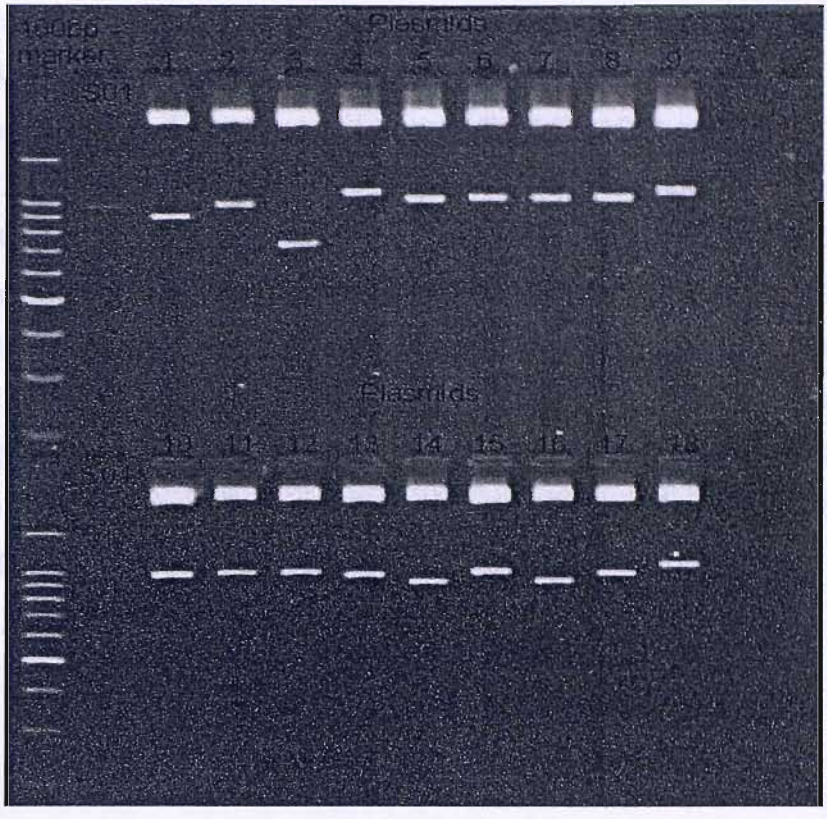
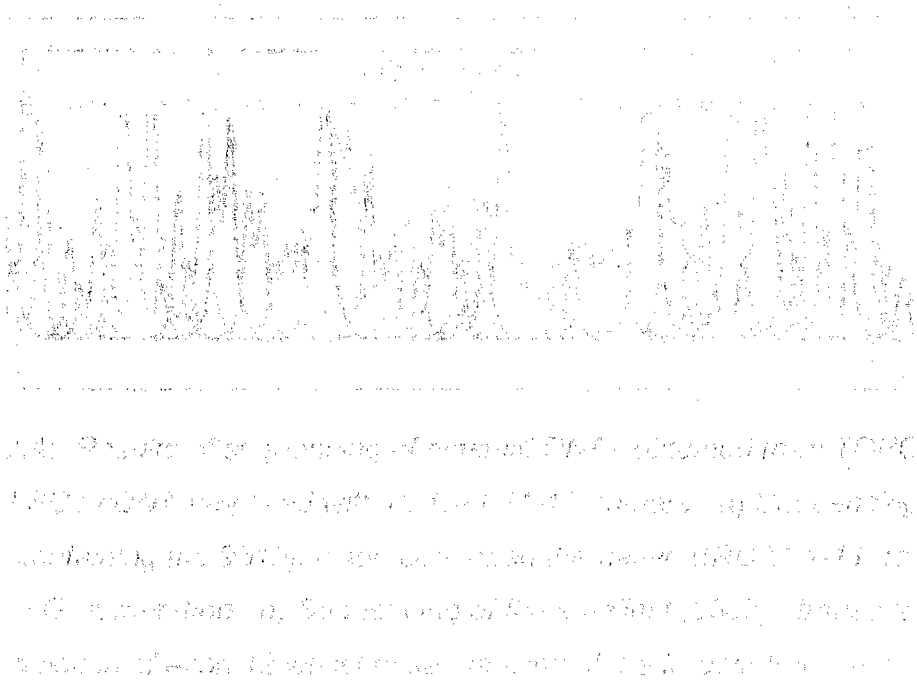


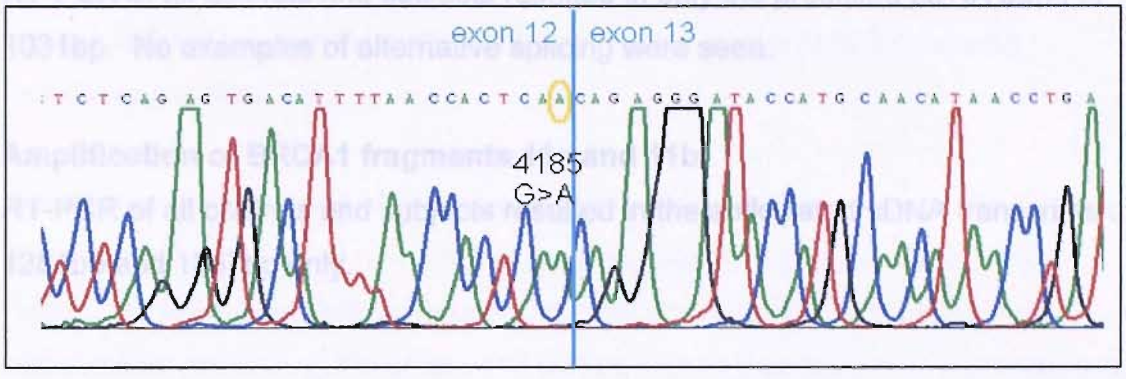
Figure 5.13: *Agarose gel electrophoresis of Eco RI digests of 18 plasmid DNAs obtained from TOPO TA cloning of SO1 cDNA amplified with BRCA1 11-17 primers. Five different size inserts are seen. Plasmids 6,7,10, 11, 12 and 15 contain a 951bp insert consistent with the anticipated exon 11-17 transcript. Plasmids 1, 14 and 16 contain a smaller 860bp transcript and plasmids 4,9,15 and 18 contain a 1000bp insert*

Sequencing of all plasmids except 8 and 13 was successful. The sequence of the 951bp inserts (plasmids 6,7,10,11,12 and 15) was identical in each case to the BRCA1 cDNA reference sequence between the BRCA1 11F and 17R primers with the exception of one which contained a G>A mutation at c4185 as expected for subject S01. Sequencing of the plasmids which corresponded to the smallest transcript (plasmids 1, 14 and 16) revealed deletion of the entire exon 12 whilst sequencing of the plasmids containing the largest insert of approximately 1000bp

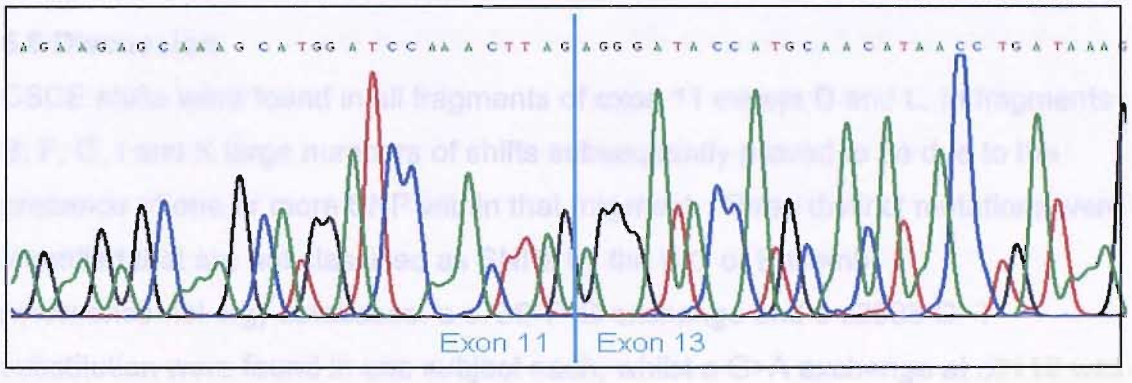
(4, 9, 15 and18) identified a 66 nucleotide in-frame insertion at the beginning of exon 14. Analysis of the inserted sequence using National Centre for Biotechnology Basic Local Alignment Search Tool (BLAST) software (www.ncbi.nlm.nih.gov/Education/BLAST) indicated that this corresponded to an in-frame intronic insertion from intron 13 (nucleotides 49332-49397 according to GenBank sequence L78833) between exons 13 and 14. Plasmid (17) proved to have both the exon 12 deletion and the intronic insertion, whilst plasmid 3 had a deletion of exon 15 in addition to the exon 12 deletion. This latter fragment did not correspond in size to any of the transcripts produced by PCR amplification of BRCA1 cDNA using the 11F and 17R primers and is therefore presumed to be an aberrant sequence induced by cloning.



a)



b)



c)

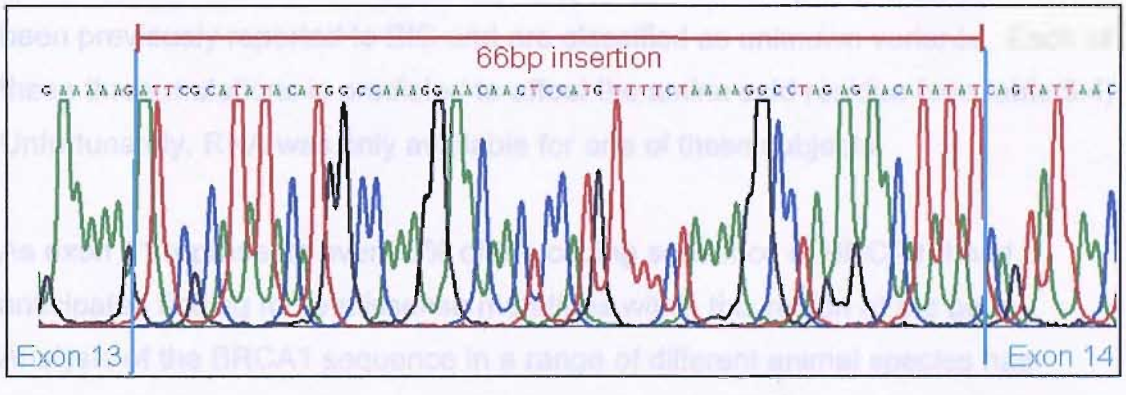


Figure 5.14: *Results of sequencing of plasmid DNAs obtained from TOPO TA cloning of SO1 cDNA amplified with BRCA1 11-17 primers.* a) *Sequencing of plasmids containing the 951bp insert confirmed the native BRCA1 1-11 transcript with c4185 G>A mutation. b) Sequencing of the smallest (860bp) transcript revealed deletion of exon 12 whilst c) sequencing of the largest transcript identified a 66 nucleotide insertion between exons 13 and 14.*

Amplification of BRCA1 fragment 15-24

RT-PCR of all controls and subjects resulted in only the predicted cDNA band of 1031bp. No examples of alternative splicing were seen.

Amplification of BRCA1 fragments 11a and 11b

RT-PCR of all controls and subjects resulted in the anticipated cDNA transcripts of 1284bp and 1357bp only.

5.6 Discussion

CSCE shifts were found in all fragments of exon 11 except D and L. In fragments B, F, G, I and K large numbers of shifts subsequently proved to be due to the presence of one or more SNP within that fragment. Three distinct mutations were identified that are not classified as SNPs by the BIC or Ensembl (www.ensembl.org) databases: a c736 T>G exchange and a c2083 G>T substitution were found in one subject each, whilst a G>A exchange at c3119 was found in three unrelated individuals. All three of these sequence variations have been previously reported to BIC and are classified as unknown variants. Each of these three mutations is predicted to affect the amino acid residue (see table 5.4). Unfortunately, RNA was only available for one of these subjects.

As exon 11 represents over 60% of the coding sequence of BRCA1, I had anticipated finding more missense mutations within this region of the gene. Analysis of the BRCA1 sequence in a range of different animal species has indicated that regions of exon 11 (notably codons 200-300 and around 900) are particularly highly conserved [105, 107]. Sequence alterations in these regions would therefore be expected to have especially deleterious effects. Examination of the distribution of missense mutations reported to BIC reveals that exon 11 has 0.073 distinct missense mutations per base pair compared with the overall average of 0.107 per base pair for all BRCA1 exons (see table 5.9). The distribution of

missense mutations within these conserved regions (0.063 missense mutations per base pair for codons 200-300 and 0.066 missense mutations per base pair for codons 850-950) is similar to that for exon 11 as a whole (0.073 missense mutations per base pair).

Whilst most observations of increased amino acid selection pressure (Ka/Ks where Ka represents the rate of non-synonymous mutations at a given site and Ks represents the rate of synonymous mutations) are due to an increase in non-synonymous mutations, Hurst and Pal demonstrated that the increased Ka/Ks ratio between exon 11 codons 200-300 was due to a reduced number of synonymous mutations [107]. Orban *et al.* have subsequently suggested that this reduction in Ks is due to the presence of splicing enhancer sequences which are under selection pressure to maintain the correct alternative splicing of exon 11 [149]. This data adds to the weight of evidence suggesting that all missense mutations in this region need to be carefully investigated at the RNA level to examine their effect on splicing.

Exon	Size (bp)	No. of distinct missense mutations reported	No. of distinct missense mutations /bp
1	378		
2	99	13	0.131
3	54	13	0.241
5	78	15	0.192
6	89	7	0.079
7	140	8	0.057
8	106	8	0.075
9	46	3	0.065
10	77	4	0.052
11	3426	250	0.073
12	89	5	0.056
13	172	14	0.008
14	127	8	0.063
15	191	15	0.079
16	311	33	0.106
17	88	14	0.159
18	78	15	0.192
19	41	5	0.122
20	84	15	0.179
21	55	8	0.145
22	74	8	0.108
23	61	11	0.180
24	1502	15	0.010

Table 5.9: *Distribution of missense mutations in BRCA1 exons*

Between April 2005 and January 2006, 180 subjects with a personal history of breast cancer and strong family history of breast/ovarian cancer but previous negative routine screens of BRCA1/2 had these genes screened by CSCE performed by the WRGL. A total of eight further missense mutations were found: eight exonic and two intronic. The exonic missense mutations were located in exons 2, 3,11,15,17 and 18 whilst the intronic missense mutations were located in introns 7 and 10. RNA was available for five of these subjects.

At the time of writing the WRGL has completed CSCE analysis of the entire BRCA1 coding sequence of 471 patients. Eight per cent of all CSCE fragments have required further analysis by direct sequencing. Summary data from all BRCA1 and BRCA2 screens indicates that 15.8% of the patients analysed by

WRGL carry a BRCA1 or BRCA2 UV. Currently, the failure rate of CSCE analysis of BRCA1/2 by the WRGL is 1.7%. These figures confirm that CSCE is an extremely efficient method of detecting BRCA1/2 exonic missense mutations.

On commencing this piece of work, this was, to the best of our knowledge, the first time that analysis of the entire BRCA1 transcript using multiplex RT-PCR had been designed as the basis of a rapid screening system to detect underlying mutations. Recently, Bonatti *et al.* reported the successful use of six overlapping BRCA1 cDNA amplicons to examine the effect of known splice site variants and also described the use of this system to examine the BRCA1 transcripts of 41probands belonging to Italian families with breast and/ or ovarian cancer but no known BRCA1/2 mutation [202]. Although their publication describes the use of an exon 10-12 amplicon, no technical details are given as to how this was produced. One previous publication details the use of four overlapping amplicons covering the entire BRCA1 transcript, but a two-stage PCR process was required to achieve this and no transcripts containing the whole of exon 11 were produced [206].

Despite trials of several different long-range PCR systems, I was unable to successfully amplify either the 2388bp BRCA1 9-11 fragment or the 3798bp 9-13 fragment. It proved technically impossible to design a single alternative pair of primers that would theoretically amplify the entire exon 11 and flanking exon boundaries in a single PCR. As the original aim of this aspect of my project was to develop a quick and simple set of PCRs that could be used as a method of screening BRCA1 transcripts for evidence of aberrant splicing patterns, two further pairs of primers were designed to amplify the BRCA1 segment falling in between the 1-11 and 11-13 amplicons. These primers amplified the 11a and 11b amplicons readily under standard PCR conditions. However, at present this system of overlapping primers lacks sensitivity in detecting abnormal splicing in the exon 11 region. This is demonstrated by the fact that although the common BRCA1 alternative transcript Δ9,10 was detected, as an abnormal 814bp transcript resulting from the 1-11 BRCA 1 amplification, I saw no evidence of the often

reported Δ11q and Δ9,10,11q transcripts. The Δ9, 10 alternative transcript was seen in all controls and subjects tested here. Failure to produce the Δ11q and Δ9,10,11q transcripts on amplifying the 11-13 and 11-17 fragments can be explained by the fact that the forward primers for these amplifications are sited within the c788-c4096 deleted region of exon 11 (at c2777 and c4066 respectively). Use of the Berkely Drosophilia Genome Project (BDGP, http://www.fruifly.org) Splice Site Prediction using Neural Networks (SSPNN) software calculates a splicing efficiency of only 45% for the site of exon 11 which may explain the tendency of this region to exhibit a variety of splicing patterns. As discussed in the introduction, the precise function of these common alternative transcripts remains unclear.

A second common alternative transcript was also identified on amplification of the 1-11 fragment. Deletion of the last (3') 22 nucleotides of exon 5 in addition to the deletion of exons 9 and 10 was seen in all controls and subjects as a third very faint cDNA band of 792bp. This transcript was only visible on PAGE analysis using at least 10μl of the 50μl PCR mix. A transcript consisting of this 22bp exon 5 deletion but with intact exons 9 and 10 (known as BRCA1-Δ22deltantex5) has previously been reported as occurring at low levels in controls [207]. Levels of this transcript are increased by the exonic mutation c190T>G and the intronic mutation IVS5+3A>G [172, 207], both of which apparently cause preferential use of the exon 5 cryptic donor TAT/gtaaga. The report by Velasco *et al.* that the exon 5 c211 A>G (330A>G) missense mutation causes "disruption of the natural 5' splice site and use of a cryptic donor site" resulting in the 22 nucleotide deletion described above is however potentially misleading as there does not appear to have been any attempt here to compare the level of this transcript produced by the mutant allele with that produced by wildtype BRCA1 [208]. Deletion of this sequence results in a frame shift alteration after codon 63 and is predicted to create a premature termination at codon 80 in exon 6. The shortened transcript would however still include the BRCA1 RING motif which is central to many protein-protein

interactions. This is the first report of the Δ22deltantex5 cryptic splicing in combination with deletion of exons 9 and 10.

The existence of these alternative transcripts resulted in some difficulty in analysing the agarose gel electrophoresis pattern created by the BRCA1 1-11 amplicons. Analysis of cDNA transcript sizes using polyacrylamide gel electrophoresis (PAGE) is much more sensitive than agarose gel electrophoresis, but is more time consuming and can frequently be complicated by distortion of PAGE sample wells resulting in unequal migration of samples, as seen with the original analysis of samples BO4 and SAL03. It is therefore not an ideal method to use within a rapid screening system.

A third example of a common alternative transcript was identified on amplification of the 11-17 fragment. This revealed an additional cDNA transcript in subjects and controls which corresponded to an in-frame 66bp intronic insertion between exons 13 and 14. This sequence variation would have the effect of adding 22 amino-acids after amino-acid 1452 and altering the trans–activation domain 1 (AD1) and the protein interacting domain of BRCA1 with BRCA2 and MSH2. It is interesting to note that whilst the exon 13 splice donor site matches the consensus sequence extremely well (SSPNN score 0.99), the exon 14 acceptor site deviates considerably from the consensus acceptor sequence (no SSPNN score). The 13A acceptor score is 0.66 potentially facilitating the use of this cryptic splice site.

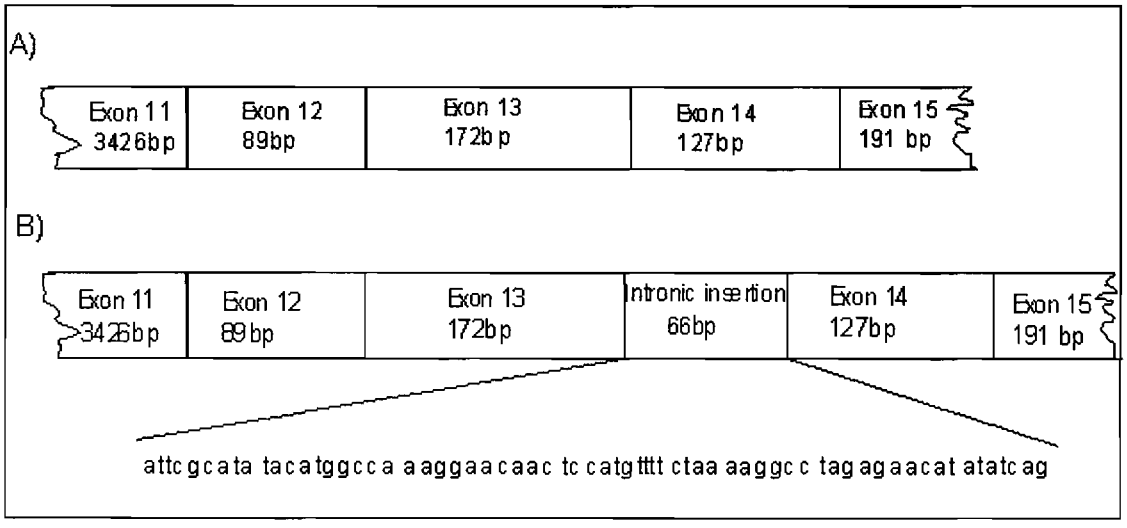


***Figure 5.15**: Illustration of A) wildtype transcript and B)) alternative transcript containing 66bp intronic insertion*

Review of previous amplifications of this portion of BRCA1 in 20 anonymous controls confirmed that a transcript of this size was present, albeit very faintly, in a large number of the controls. Claes *et al.* also observed an identical in-frame insertion of 66 basepairs from intron 13 between exons 13 and 14 in both the Czech patient and one control [209]. In order to establish the prevalence of this alternative transcript, I used BRCA1 12del forward and reverse primers (see appendix F) to amplify a 350bp segment of BRCA1 exons 13-14 from cDNA of 29 controls. A second transcript approximately 410 bp in size was visible in 26 cases (89.7%), consistent with the insertion of a 66bp intronic sequence between the normal exon boundaries of exons 13 and 14.

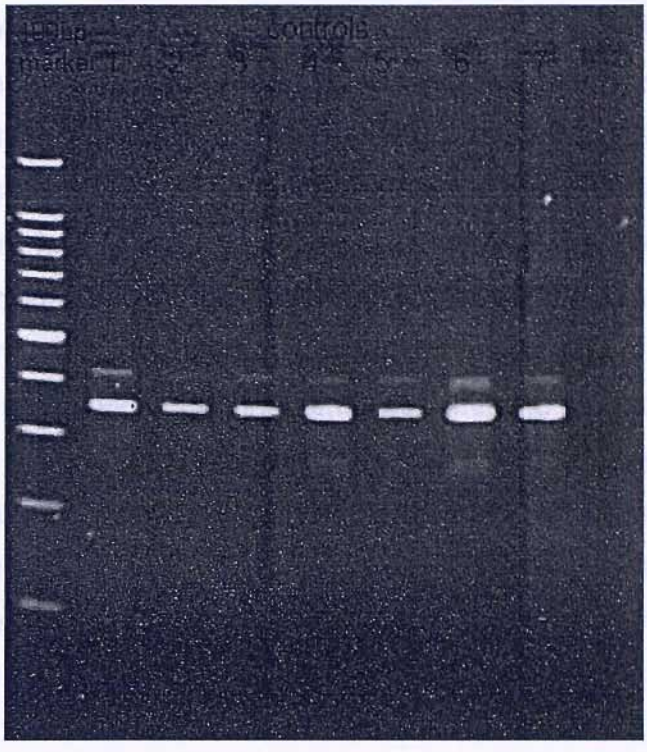


***Figure 5.16**: Demonstration of alternative transcript (413bp) containing 66bp intronic insertion between BRCA1 exons 13 and 14 and wildtype transcript (350bp) in seven control cDNAs amplified with BRCA1 12delF and BRCA1 12delR primers*

Attempts to quantify the ratio of wildtype transcript to alternative transcript using pyrosequencing technology were unsuccessful. However, in August 2005 Fortin *et al.* published an article describing their discovery of the same 66bp intronic insertion between BRCA1 exons 13 and 14 (designated exon13A) in total RNA samples from 12 normal tissues and several breast and other cancer lines [210]. Real-time PCR revealed that this transcript corresponded to 20-25% of the total BRCA1 mRNA expression levels in leucocytes, brain and cerebellum tissues but only 5% of total mRNA in other tissues and cancer cell lines. The high prevalence of this alternative transcript suggests that it is unlikely to carry significant pathogenic consequences.

Amplification of cDNA fragments spanning exons 1-11, 11a, 11b, 11-13, 11-17 and 15-24 did not reveal any evidence of additional alternative mRNA transcripts in

subjects with known truncating mutations or exonic deletions. This confirms the hypothesis that the normal allele in each case is still producing a normal mRNA transcript. Truncating mutations should either be producing a PTC that condemns the abnormal transcript to degradation via the NMD pathway, or will produce a normal length transcript that results in a truncated protein. Neither situation would be detectable by this technique, thus this method is unsuitable as a screen for truncating mutations. Exonic deletions will only be detected if the primers are still able to bind to the remaining sequence; otherwise only the products of the normal allele will be seen. This explains the normal sized fragments observed with T11 and T15 despite their heterozygous deletion of exons 10-12. It was predicted that the abnormal, shortened BRCA1 allele of these patients should be successfully amplified using the BRCA1 RT 9-13 primers as a 206 bp amplicon but this could not be achieved in practice.

Use of this multiplex PCR system identified one example of abnormal splicing in association with previously identified point mutations, in subjects S01-4. These subjects were previously found to have a heterozygous G to A substitution at c4185 on predictive screening involving heteroduplex analysis. Subjects SO1-3 all come from the same family, but S04 is not known to be related to these individuals. Cloning of the PCR products of the 11-17 amplification of subject S01 clearly indicated that the smallest (850bp) alternative transcript represented deletion of exon 12. Only one of the ten plasmids containing the complete 11-17 transcript with or without the additional intronic insertion carried the 4185 G>A mutation indicating that this mutation generally co-segregates with the exon 12 deletion.

Position 4185 represents the final nucleotide of exon 12. It is therefore an obligatory component of the 3" splice site and the G nucleotide is conserved in 78-86% of mammals [160]. Alteration of this nucleotide is therefore likely to result in an alternative splicing pattern and use of the BDGP splice site predictor software shows that the 4185 A>G mutation reduces splicing efficiency of the 5'donor site from 95% (wildtype) to 38%. Although only the PCR products of S01 were cloned,

the aberrant smaller transcript of subjects S02-S04 was identical in size to that of S01 supporting the hypothesis that the 4185 mutation is associated with deletion of exon 12 in related and non-related individuals. These observations are consistent with the findings of Claes *et al.* who also identified an aberrant transcript lacking exon 12 on amplification of BRCA1 exons 11-15 using cDNA from a single Czech patient carrying the c4185 (4304) silent mutation [209]. Deletion of the 89bp exon 12 results in a frame shift and is predicted to produce a stop codon at 1373, supporting the hypothesis that this mutation is pathogenic.

Although the mutation 122A>G, (located 12 base pairs form the exon 3 5' donor site), carried by both SAL03 and SAL205 is predicted by ESEfinder to affect a SRp55 ESE motif, repeat analysis of the PCR products produced by amplification of exons 1-11 failed to demonstrate any examples of aberrant splicing of exon 3 to exon 5. Analysis of the transcripts produced by subject U05, a carrier of the exon 11 missense mutation c736T>G, was also unremarkable, despite the previous report of possible constitutive skipping of exon 11 in this subject [171]. No examples of aberrant splicing were found in subjects in whom no mutation had been previously identified by routine screening (N01-21). This result is similar to that of Bonatti et al. who found no cDNA alterations in 41 breast/ ovarian cancer patients with no known BRCA1/2 mutations [202]. At the time of completing this project, all patients assessed using the multiplex RT-PCR method had also undergone full screening of BRCA1 and BRCA2 screens using CSCE and no exonic BRCA1 missense mutations were identified in any of these patients. One patient (N02) was however found to have a BRCA2 missense mutation (8567A>C).

It should be noted that care must be taken to correctly interpret the result of sequencing derived from cloning of cDNA amplicons. On three separate occasions, cloning of total PCR products resulted in one or more plasmids containing an insert which demonstrated the deletion of a whole exon (exon 13 on cloning of SO1 11-17, and exon 5 on cloning control 1 1-11 and SAL03 exon 5) on sequencing. However in each of these cases, the size of the plasmid insert did not

match any of the transcripts seen on gel electrophoresis and in the case of the SAL03 plasmid, sequencing of the insert showed absence of the missense mutation carried by this patient. These findings therefore cannot be interpreted as true alternative transcripts and are presumed to be the result of DNA damage during the cloning process.

5.7 Conclusions

CSCE is a rapid and effective method of screening the entire exonic sequences of BRCA1 and BRCA2 for sequence variations. My screen of exon 11 in 94 subjects with a strong family history of breast/ ovarian cancer but no known truncating BRCA1/2 mutation was limited by technical problems but did identify three missense mutations. None of these fall within the highly conserved regions of BRCA1. Subsequent screening of 180 patients by the WRGL identified a further seven BRCA1 rnissense mutations.

Analysis of transcripts produced from the amplification of BRCA1 cDNA using a series of overlapping primers did effectively demonstrate the common alternative transcripts BRCA1Δ9,10, BRCA1-Δ22deltantex5 and BRCA1 13A and confirmed that the c4185A>G silent mutation is associated with deletion of exon 12. This method however failed to identify any underlying BRCA1 mutations not picked up by other screening methods and is potentially most useful as a method of investigating patients with known missense or splice site mutations.

6.0 Use of Minigenes to Investigate the Effect of Missense Mutations on Splicing Regulation

6.1 Introduction

Experimental evidence indicates that the use of a minigene construct containing the genomic region of interest is an effective method of identifying the effects of *cis*-exonic or intronic mutations on splicing regulation (see review [211]). Current knowledge suggests that the primary splicing regulatory elements are typically sited within 300 nucleotides up or downstream of the regulated exon and it is extremely rare for a regulatory element to be sited upstream or downstream of the exons immediately adjacent to the regulated exon [212, 213]. A minigene construct containing the exon of interest as its central component flanked by complete introns and portions of the flanking exons is therefore considered to be very likely to include all the relevant elements required for splicing regulation. Transfection of minigenes into cells with high transfection efficiency such as HeLa, COS and HEK293 has been used extensively to identify the factors involved in non-constitutive splicing.

Three research groups have recently used the technique of site directed mutagenesis in minigenes to investigate the effect of mutations on splicing of cancer predisposition genes. McVety *et al.* used this technique to alter a series of bases within a purine rich region of MLH1 exon 3, following the observation that a DNA 3 base pair deletion in this locality was associated with cDNA skipping of exon 3 in a family with HNPCC and fell within an ESEfinder predicted ESE [173]. Transfection of the mutant minigenes into Cos-1 cells showed that silent, missense and nonsense mutations introduced into codon 461 all had the effect of deleting exon 3 in some or all of the resulting transcripts. Similarly Lastella *et al.* used separate hMLH1 minigene constructs containing one naturally occurring missense mutation K461X and 15 different novel mutations created by site-directed mutagenesis to attempt to abolish all ESEfinder predicted SF2/ASF motifs within exon 12 of this gene [175]. Transfection of these minigenes into Cos-1 cells showed that four of the sixteen mutations tested resulted in skipping of exon 12.

To date there have been no published examples of this approach in the investigation of BRCA1 missense mutations. I therefore decided to commence a systematic analysis of the effect of BRCA1 exonic missense mutations on the regulation of splicing by constructing a number of BRCA1 minigenes and introducing previously reported missense mutations into the central exon of these. The aim was to transfect wildtype and mutant minigenes into cell lines with high transfection efficiency and harvest RNA after 24 hours. Following RT-PCR, the resultant transcripts would be analysed by gel electrophoresis and direct sequencing. The results obtained would be compared with the predicted effect of the sequence variants on potential ESE motifs as determined by the algorithm ESE finder [163]. It was anticipated that RNA from patients carrying the tested missense mutations might become available as the project progressed to provide comparative native transcripts.

6.2 Methods

6.21 Design of minigenes and amplification primers

An ideal minigene construct contains a central exon of no more than 200bp and flanking introns of up to 1500bp each. Analysis of BRCA1 exon and intron sizes confirmed that no native section of wildtype BRCA1 DNA sequence conforms to these specifications. Exon 11 is a particularly poor candidate given its very large size (3426bp). For initial trials of this technique, I selected exons 5,6,10 and 18 which are all of moderate size (78, 89, 77 and 78 bp in size respectively) and which all have high frequencies of reported missense mutations (see table 6.1).

	Size of	Size of	Size of	Distinct missense mutations	
Exon	exon (bp)	5'Intron (bp)	3' Intron (bp)	no. /exon	no. /base pair
1	378		621		
2	99	621	8237	13	0.131
3	54	8237	9192	13	0.241
5	78	9192	1499	15	0.192
6	89	1499	606	7	0.079
7	140	606	4241	8	0.057
8	106	4241	2485	8	0.075
9	46	2485	1321	3	0.065
10	77	1321	985	4	0.052
11	3426	985	402	250	0.073
12	89	402	8638	5	0.056
13	172	8638	5789	14	0.008
14	127	5789	1966	8	0.063
15	191	1966	3092	15	0.079
16	311	3092	3232	33	0.106
17	88	3232	3656	14	0.159
18	78	3656	500	15	0.192
19	41	500	6197	5	0.122
20	84	6197	5934	15	0.179
21	55	5934	1868	8	0.145
22	74	1868	1417	8	0.108
23	61	1417	1840	11	0.180
24	1502	1840		15	0.010

Table 6.1: *Size of BRCA1 exons and introns, and distribution of exonic missense mutations*

As shown in figure 6.1, constructs for exon 6 and 10 were designed to incorporate the entire flanking intron sequence whilst the other 2 constructs required truncation of the proximal intron in order to make insertion of the minigene into a plasmid technically viable.

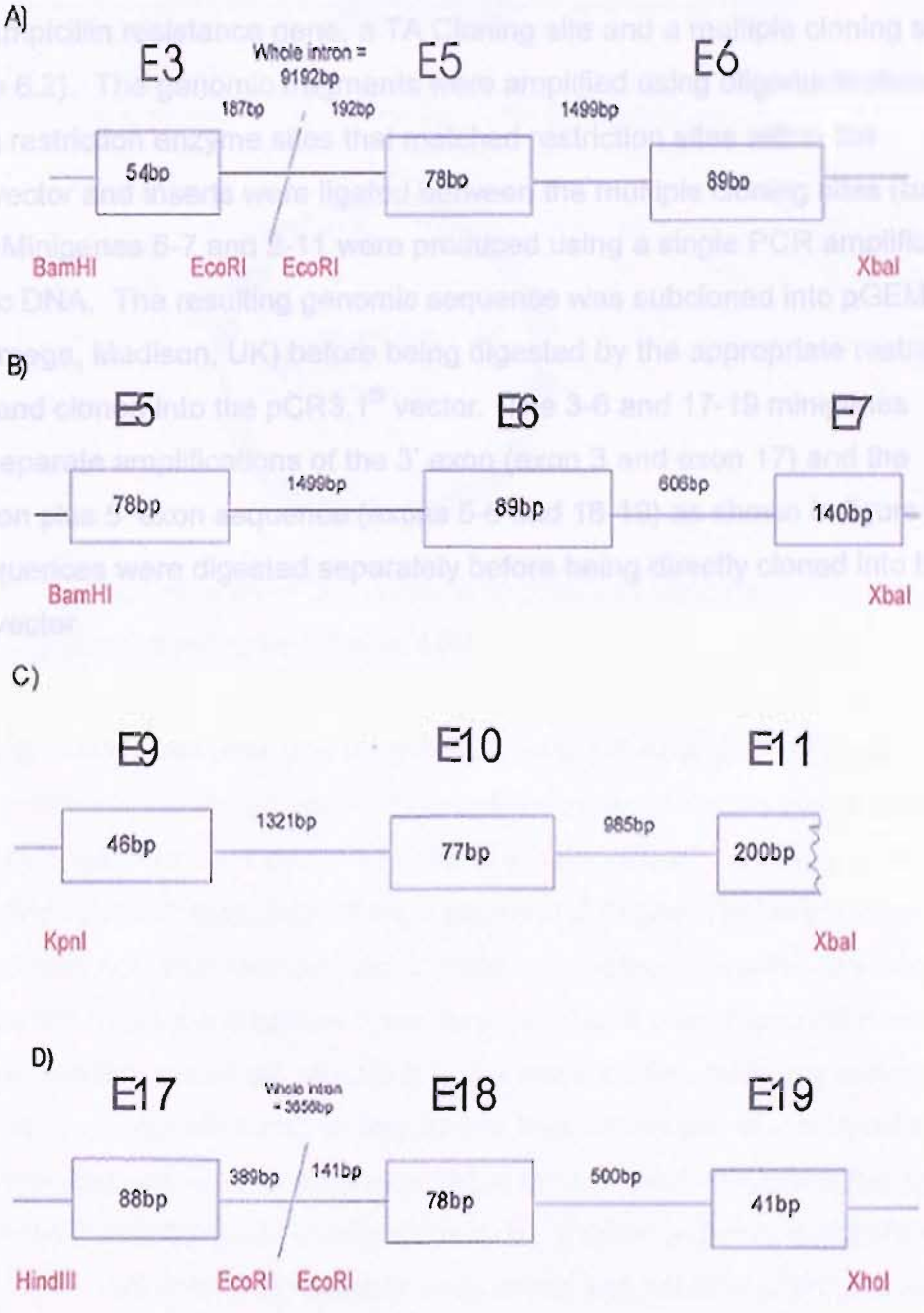


Figure 6.1: *the four BRCA1 minigenes consisting of exons 3-6 (A), 5-7 (B), 9-11 (C) and 17-19 (D). Minigenes 5-7 and 9-11 contain entire introns whilst the proximal intron in minigene 3-6 and 17-19 is truncated for technical reasons. Restriction enzyme sites are indicated in red.*

In order to permit transfection into HEK 293 cells, the pCR3.1® vector (Invitrogen, Paisley, UK) was chosen as an appropriate vector. This contains a CMV promoter for transcription in mammalian cells, a SV40 promoter for replication in mammalian cells, an Ampicillin resistance gene, a TA Cloning site and a multiple cloning site (see figure 6.2). The genomic fragments were amplified using oligonucleotides containing restriction enzyme sites that matched restriction sites within the pCR3.1® vector and inserts were ligated between the multiple cloning sites (bases 670-801). Minigenes 5-7 and 9-11 were produced using a single PCR amplification of genomic DNA. The resulting genomic sequence was subcloned into pGEM®-T Easy (Promega, Madison, UK) before being digested by the appropriate restriction enzymes and cloned into the pCR3.1® vector. The 3-6 and 17-19 minigenes required separate amplifications of the 3' exon (exon 3 and exon 17) and the central exon plus 5' exon sequence (exons 5-6 and 18-19) as shown in figure 6.1. These sequences were digested separately before being directly cloned into the pCR®3.1 vector.

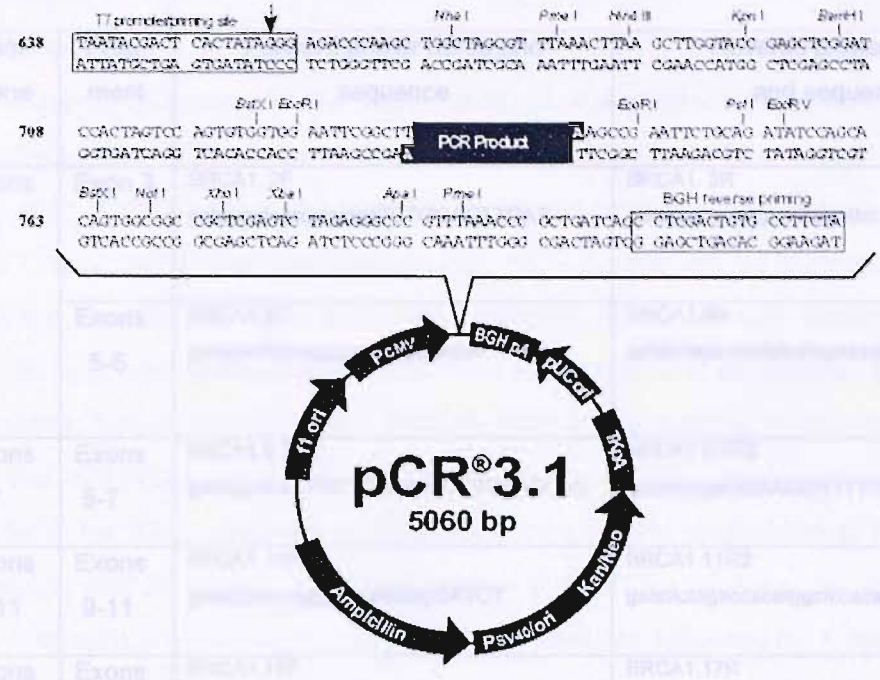


Figure 6.2: *Diagram of pCR®3.1 vector with details of sequences flanking cloning region (copyright Invitrogen, Paisley, UK)*

Twenty to twenty-five base pair amplification primers were designed using OLIGO® online software (http://oligo.net), with restriction enzyme linkers being added to the 3' aspect of each forward primer and the 5' aspect of each reverse primer. NEBcutter V2.0 software, (http://tools.neb.com/NEBcutter2/index.php), was used to identify restriction enzymes that would not cut any other site within the minigene and for each minigene fragment it was ensured that a pair of restriction enzymes was selected that would cut efficiently in the same buffer. As many restriction enzymes do not cut efficiently at sequences sited at the end of a molecule, four nucleotides (usually GATA) were inserted in front of the restriction linker to form the first few nucleotides of the oligonucleotide. Primer sequences are shown in table 6.2. For full details of minigene sequences and position of primers see Appendices G-J.

Mini-gene	Frag-ment	Forward primer name and sequence	Reverse primer name and sequence	Amplicon size (bp)
Exons 3-6	Exon 3	BRCA1. 3F gataggatccccctgctagTCTGGAGTTGAT	BRCA1. 3R gatagaattccaggaactatgattacaacca	250
	Exons 5-6	BRCA1.5F gatagaattcaagggcagttgtgagattatc	BRCA1.6R gatatctagaccacgtcatagaaagtaatt	1856
Exons 5-7	Exons 5-7	BRCA1.5.7F2 gataggatccATGCTGAAACTTCTCAACCAG	BRCA1.5-7R2 gatatctagaGGAAGGATTTTCGGGTTCACT	2402
Exons 9-11	Exons 9-11	BRCA1.11F2 gatatggtaccggggaaatttttagGATCT	BRCA1.11R2 gatattctagaccacatggctccacatgcaagt	2601
Exons 17-19	Exons 17	BRCA1.17F gataaagcttttaatttcagATGCTCGTGTA	BRCA1.17R tagaattctcatccatgctatgctcaaca	497
	Exons 18-19	BRCA1-18F gatagaattcaccaggggttttagaatcata	BRCA1-19R gatactcgagatctctggttagtttgtaaca	791

Table 6.2: *BRCA1 minigene construct primer sequences*
(Small case indicates intronic sequence, capital letters indicate exonic sequences)

6.22: Production of wildtype minigenes

All PCRs were performed using a high fidelity polymerase in order to minimise the risk of introducing random nucleotide substitutions during the amplification process. "Hot-start" techniques were also used to reduce background amplification.

a) Primary PCR

To amplify each wildtype minigene fragment, 5 µl of anonymous control genomic DNA (25 ng/µl) sample was added to 20 µl of a master mix containing 59.5 µl sterile water, 10 µl of 10x PCR Buffer II, 8 µl 10mM dNTPs, 1 µl 3' primer (10 pmol/µl), 1 µl 5' primer (10 pmol/µl), 0.25 µl AmpliTaq Gold polymerase and 0.25 µl *Pfu* DNA polymerase prepared on ice. PCR conditions for fragments 3 and 17 (260bp and 500bp) were: 94°C for 3 minutes; 30 cycles with denaturation at 94°C for 30 seconds, annealing at 54°C for 45 seconds and elongation at 72°C for 4 minutes; 1 cycle at 72°C for 7 minutes; and a final hold at 15°C. PCR conditions for the remaining fragments were: 94°C for 3 minutes; 15 cycles with denaturation at 94°C for 30s, annealing at 55°C for 1 minute and elongation at 72°C for 5 minutes and 20 cycles with denaturation at 94°C for 30 seconds, annealing at 63°C for 1 minute and elongation at 72°C for 5 minutes followed by 1 cycle at 72°C for 7 minute; and a final hold at 15°C.

All PCR products were visualised on a 2% agarose gel and excised. DNA was extracted from the agarose gel using GENECLEAN®, (section 3.3.2) and eluted into volumes of either 10 µl for sub-cloning into pGEM®-T Easy (minigenes 5-7 and 9-11) or 20 µl for digestion by restriction enzymes at flanking sites and direct cloning into pCR 3.1® vector (minigenes portions 3, 5-6 and 17, 18-19).

b) Cloning of 5-7 and 9-11 minigenes

The amplified 5-7 and 9-11 DNA sequences were cloned into pGEM®-T Easy as described in section 3.3.8.2 and plasmid DNA was extracted using Wizard® Plus SV Minipreps DNA Purification System. A diagnostic digest was performed of each plasmid to check that the DNA insert was of the anticipated size. The pGEM®-T Easy plasmid was then digested with the appropriate enzymes (see table 6.4) and cloned into vector pCR®3.1 as described in 3.3.8.2. JM 109 cells were transformed and colonies from each plate were grown up and plasmid DNA was extracted as previously detailed.

Minigene	Fragment	Size of fragment	Restriction linker inserted in primers	Restriction enzyme digestion	Buffer
3-6	3F-3R	250	3' G/GATCC 5' G/AATTC	BamHI EcoRI	E
	5F-6R	1856	3' G/AATTC 5' T/CTAGA	EcoRI Xba I	H
	3F-6R	2106	3' G/GATCC 5' T/CTAGA	Bam HI Xba I	E
5-7	5F2-7R	2402	3' G/GATCC 5' T/CTAGA	BamHI Xba I	E
5-7 modified*	INT5-7 –7R	1096	N/a 5' T/CTAGA	Pst I Xba I	H
9-11	9F1-11R1	2601	3' GGTAC/C 5' T/CTAGA	Kpn I Xba I	Multicore
17-19	17F-17R	497	3' A/AGCTT 5' G/AATTC	Hind III EcoRI	E
	18F-19R	791	3' G/AATTC 5' C/TCGAG	EcoRI Xho I	H
	17F-19R	1288	3'A/AGCTT 5' C/TCGAG	Hind III Xho I	B

Table 6.3: *The size of each minigene fragment and details of the linkage sequences incorporated into the relevant primers with restriction enzyme pairs and buffers required for digests prior to cloning into pCR 3.1® vector, (*see section 6.24b).*

c) Cloning of 3-6 and 17-19 minigenes

The amplified proximal exons (exon 3 and exon 17) were first cloned directly into the pCR®3.1 vector using the method described in section 3.3.8.2. Details of the restriction enzymes required are given in table 6.4. Ligations, transformation of JM109 cells and extraction of plasmid DNA were performed as previously described. Diagnostic digests using the appropriate restriction enzymes were performed to ensure that the plasmid contained a DNA insert of the anticipated size. The purified PCR products from the arnplification of exons 5-6 and 18-19 were then cloned into the pCR®3.1 vector containing exon 3 or 17 using the method described in section 3.3.8.2. with restriction enzymes as detailed in table 6.4. JM 109 cells were transformed with the resulting ligations and plasmid DNA was extracted as before.

d) Verification of minigene sequences

Diagnostic digests using the appropriate restriction enzymes were performed to ensure that the plasmids contained a DNA insert of the anticipated size. The entire minigene sequence was then analysed by direct sequencing to verify correct fragment orientation and ensure that no undesired mutations had been acquired during the amplification or cloning procedures. Sequencing primers are detailed below in table 6.4.

e) Storage of minigenes

A 10µl aliquot of each minigene was stored at –80°C. Care was taken to ensure that additional supplies of the wildtype minigenes were always produced by transformation of cells with this original plasmid to prevent propagation of mutations.

Minigene construct	Primer	Sequence of primer	Direction
Exons 3-6	PL3	GGG AGA CCC AAG CTG GCT A	Forward
	PL4	AGT CGA GGC TGA TCA GCG G	Reverse
	seq 3-7 F new	CGT TCC TAT AAA ACC ATT CAT C	Forward
	seq 3-7 R new	AAT ACT GCC CTA AAT CCA CA	Reverse
5-7	PL3	GGG AGA CCC AAG CTG GCT A	Forward
	PL4	AGT CGA GGC TGA TCA GCG G	Reverse
	seq 3-7 R new	AAT ACT GCC CTA AAT CCA CA	Reverse
	seq 5-7 R2	GAG ACC AGT GGG AGT AA	Reverse
9-11	PL3	GGG AGA CCC AAG CTG GCT A	Forward
	PL4	AGT CGA GGC TGA TCA GCG G	Reverse
	9-11 F new	GAT TGT TCT CTA AGT TCC TCA	Forward
	9-11 R1 new	TGT AAA GGT CCC AAA TGG T	Reverse
	9-11 R2 new	CTA CAG ACT TAC CAC TCC CTA	Reverse
17-19	PL3	GGG AGA CCC AAG CTG GCT A	Forward
	PL4	AGT CGA GGC TGA TCA GCG G	Reverse

Table 6.4: *Sequences of primers used for sequencing DNA / cDNA of minigene constructs*

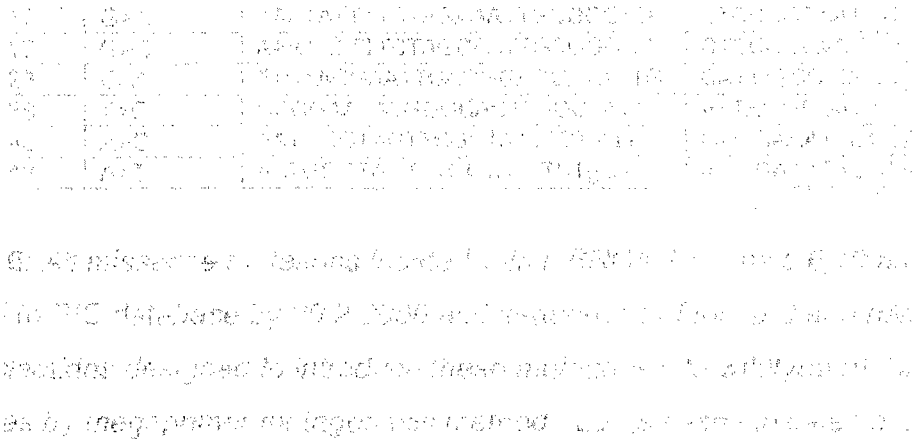
6.23 Identification of missense mutations and design of primers for mutagenesis

All exon 5,6,10 and 18 missense mutations reported to the BIC database (http://nhgri.nih.gov/intramural_research/lab_transfer/bic/) by 20.2.2006 were identified. For each missense mutation, 20-25bp forward or reverse oligonucleotides were designed that were identical to the wildtype sequence at all bases apart from the central base pair which was changed to match the reported mutation. Oligonucleotide sequences are provided in table 6.5.

Site specific mutagenesis was achieved using a megaprimer approach, as outlined in figure 6.2. Mutations of base pairs within the central exon of minigenes 5-7, 9-11 and 17-19 were achieved using forward megaprimers produced by using a primer located at the 5' end of the flanking plasmid sequence (PL4, see table 6.6 for sequences) in conjunction with the forward oligonucleotide. Reverse megaprimers produced using a primer located at the 3' end of the flanking plasmid sequence (PL3) with a reverse oligonucleotide, were used to mutate the minigenes.

Primer	Direction	Sequence
PL3	Forward	GGG AGA CCC AAG CTG GCT A
PL4	Reverse	AGT CGA GGC TGA TCA GCG G

Table 6.5: *Sequences of vector specific primers*



Exon	Position (ATG)	Nucleotide change	Forward primer (change from wildtype highlighted)	Reverse primer
5	135	A>T	ttataatttatagTTTTTGCATGCTG	CAGCATGCAAAAActataaattataa
5	139	T>G	AtttatagATTTGGCATGCTGAAAC	GTTTCAGCATGCCAAATctataaat
5	140	G>T	tttatagATTTTTCATGCTGAAAC	GTTTCAGCATGAAAAATctataaa
5	154	C>T	ATGCTGAAACTTTTCAACCAGAAG	CTTCTGGTTGAAAAGTTTCAGCAT
5	181	T>G	GGGCCTTCACAGGGTCCTTTATGTA	TACATAAAGGACCCTGTGAAGGCCC
5	181	T>C	GGGCCTTCACAGCGTCCTTTATGTA	TACATAAAGGACGCTGTGAAGGCCC
5	182	G>A	GCCTTCACAGTATCCTTTATGTA	TACATAAAGGATACTGTGAAGGC
5	189	A>T	AGTGTCCTTTTTGTAAGAATGA	TCATTCTTACAAAAAGGACACT
5	190	T>G	AGTGTCCTTTAGGTAAGAATGAT	ATCATTCTTACCTAAAGGACACT
5	190	T>C	AGTGTCCTTTACGTAAGAATGAT	ATCATTCTTACGTAAAGGACACT
5	191	G>A	AGTGTCCTTTATATAAGAATGATA	TATCATTCTTATATAAAGGACACT
5	199	G>T	TATGTAAGAATTATATAACCAAAAG	CTTTTGGTTATATAATTCTTACATA
5	201	T>G	ATGTAAGAATGAGATAACCAAAAG	CTTTTGGTTATCTCATTCTTACAT
5	203	T>A	TAAGAATGATAAAACCAAAAGgta	tacCTTTTGGTTTTATCATTCTTA
5	211	A>G	TATAACCAAAGGgtatataatttgg	ccaaattatatacCCTTTGGTTATA
6	216	C>A	TaatttcagGAGACTACAAGAAAGT	ACTTTCTTGTAGTCTCctgaaatta
6	230	C>T	ACAAGAAAGTATGAGATTTAGTCAA	TTGACTAAATCTCATACTTTCTTGT
6	231	G>T	TACAAGAAAGTACTAGATTTAGTCA	TGACTAAATCTAGTACTTTCTTGTA
6	259	T>G	GTTGAAGAGCTAGTGAAAATCATTT	AAATGATTTTCACTAGCTCTTCAAC
6	266	T>C	GCTATTGAAAACCATTTGTGCTTTT	AAAAGCACAAATGGTTTTCAATAGC
6	269	T>C	TATTGAAAATCACTTGTGCTTTTC	GAAAAGCACAAGTGATTTTCAATA
6	286	G>A	TGCTTTTCAGCTTAACACAGGTTTG	CAAACCTGTGTTAAGCTGAAAAGCA
6	292	G>C	CAGCTTGACACACGTTTGGAGTgta	tacACTCCAAACGTGTGTCAAGCTG
10	612	G>C	GAGATCAAGAATTCTTACAAATCAC	GTGATTTGTAAGAATTCTTGATCTC
10	637	A>G	CTCAAGGAACCGGGGATGAAATC	GATTTCATCCCCGGTTCCTTGAG
10	641	A>G	AAGGAACCAGGGGTGAAATCAGTT	AACTGATTTCACCCCTGGTTCCTT
10	661	G>T	GTTT GATTCTTCAAAAAAGG gtaa	TTACCCTTTTTTGAAGAATCAAAC
18	5085	T>A	agATGCTGAGTTAGTGTGTGAACGG	CCGTTCACACACTAACTCAGCATct
18	5086	G>C	gATGCTGAGTTTCTGTGTGAACGGA	TCCGTTCACACAGAAACTCAGCATg
18	5089	T>C	GCTGAGTTTGTGCGTGAACGGACAC	GTGTCCGTTCACGCACAAACTCAGC
18	5095	C>T	TTGTGTGTGAATGGACACTG AAAT	ATTTCAGTGTCCATTCACACACAA
18	5096	G>A	TTGTGTGTGAACAGACACTGAAATA	TATTTCAGTGTCTGTTCACACACAA
18	5096	G>T	TTGTGTGTGAACTGACACTG AAATA	TATTTCAGTGTCAGTTCACACACAA
18	5113	C>T	CTGAAATATTTTTTAGGAATTGCGG	CCGCAATTCCTAAAAAATATTTCAG
18	5117	G>A	AAATATTTTCTAGAAATTGCGGGAG	CTCCCGCAATTTCTAGAAAATATTT
18	5117	G>C	AAATATTTTCTAGCAATTGCGGGAG	CTCCCGCAATTGCTAGAAAATATTT
18	5123	C>A	TTCTAGGAATTGAGGGAGGAAAATG	CATTTTCCTCCCTCAATTCCTAGAA
18	5138	T>C	GAGGAAAATGGGCAGTTAGCTATT	AATAGCTAACTGCCCATTTTCCTC
18	5143	A>C	AATGGGTAGTTCGCTATTTCTgtaa	ttacAGAAATAGCGAACTACCCAT
18	5143	A>T	AATGGGTAGTTTGCTATTTCTgtaa	ttacAGAAATAGCAAACTACCCATT

Table 6.6: *All missense mutations located within BRCA 1 exons 5,6,10 and 18 as reported to BIC database by 20.2.2006 and sequences of forward and reverse oligonucleotides designed to introduce these mutations into wildtype BRCA1 minigenes by megaprimer mutagenesis method. Large case indicates exonic sequences, small case indicates intronic sequences.*

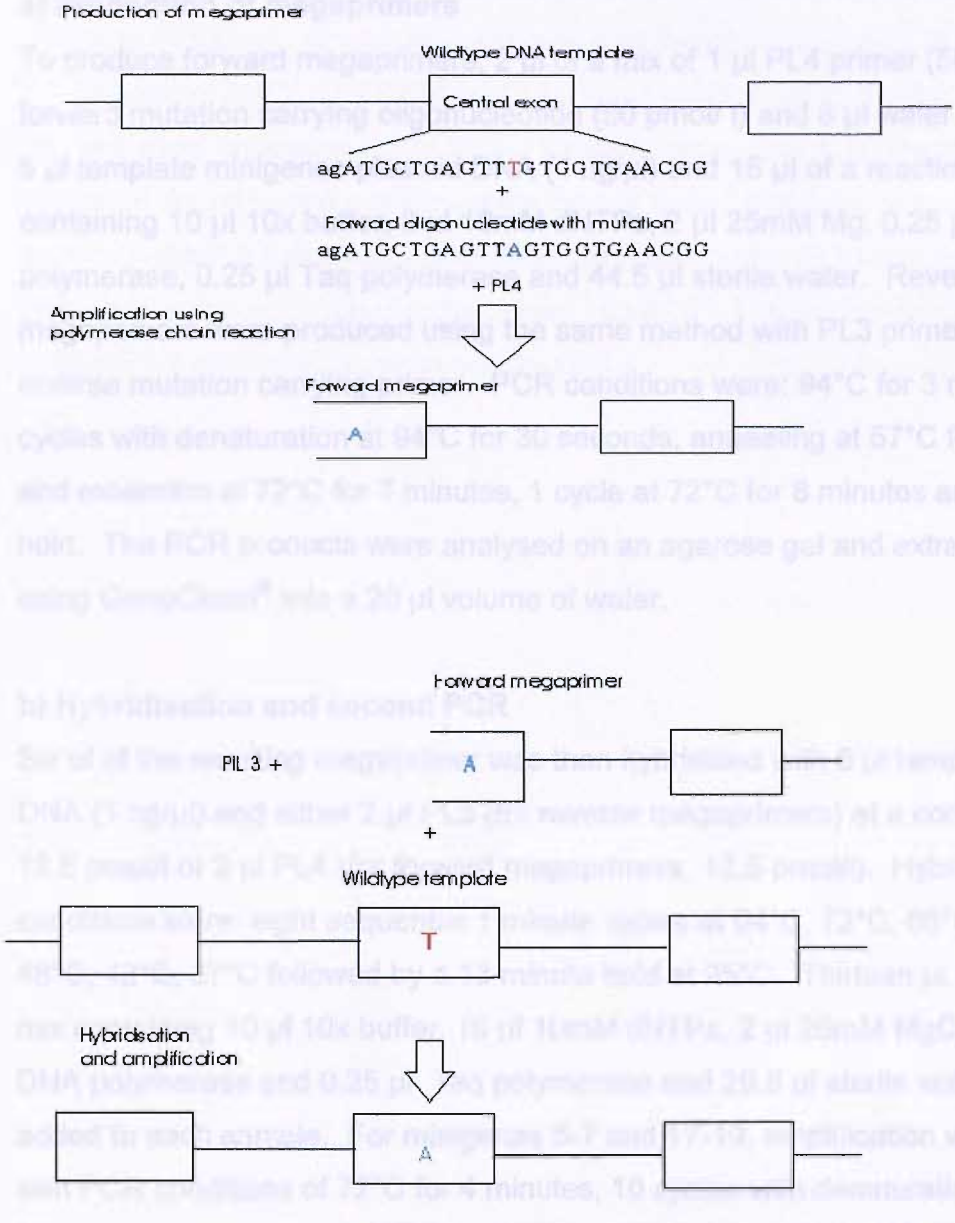


Figure 6.3: *Production of minigenes containing desired mutations.*
A) Amplification of template DNA using a forward primer containing a single specific exonic mutation produces a megaprimer containing a single desired base change. B) Hybridisation of the mutated megaprimer with additional template DNA and further amplification produces the minigene containing the desired mutation

6.24 Production of mutant minigenes

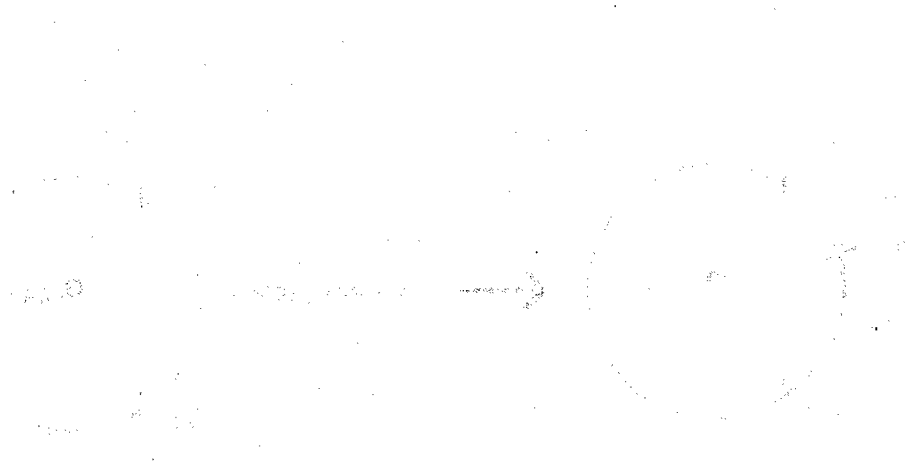
a) Production of megaprimers

To produce forward megaprimers, 2 µl of a mix of 1 µl PL4 primer (50 pmol/ l), 1 µl forward mutation carrying oligonucleotide (50 pmol/ l) and 8 µl water was added to 5 µl template minigenes plasmid DNA (1 ng/µl) and 15 µl of a reaction mix containing 10 µl 10x buffer, 8 µl 10mM dNTPs, 2 µl 25mM Mg, 0.25 µl *Pfu* DNA polymerase, 0.25 µl Taq polymerase and 44.5 µl sterile water. Reverse megaprimers were produced using the same method with PL3 primer and the reverse mutation carrying primer. PCR conditions were: 94°C for 3 minutes, 18 cycles with denaturation at 94°C for 30 seconds, annealing at 57°C for 1 minute and extension at 72°C for 7 minutes, 1 cycle at 72°C for 8 minutes and a 15°C hold. The PCR products were analysed on an agarose gel and extracted from this using GeneClean® into a 20 µl volume of water.

b) Hybridisation and second PCR

Six µl of the resulting megaprimer was then hybridised with 5 µl template minigene DNA (1 ng/µl) and either 2 µl PL3 (for reverse megaprimers) at a concentration of 12.5 pmol/l or 2 µl PL4 (for forward megaprimers, 12.5 pmol/l). Hybridisation conditions were: eight sequential 1 minute cycles at 94°C, 72°C, 65°C, 55°C, 50°C, 48°C, 42°C, 37°C followed by a 13 minute hold at 25°C. Thirteen µl of a reaction mix containing 10 µl 10x buffer, 10 µl 10mM dNTPs, 2 µl 25mM MgCl, 0.25 µl *Pfu* DNA polymerase and 0.25 µl Taq polymerase and 29.5 µl sterile water was then added to each sample. For minigenes 5-7 and 17-19, amplification was performed with PCR conditions of 72°C for 4 minutes, 10 cycles with denaturation at 94°C for 30 seconds, annealing at 55°C for 1 minute and extension at 72°C for 5 minutes, 15 cycles with denaturation at 94°C for 30 seconds, annealing at 63°C for 1 minute and extension at 72°C for 5 minutes, 1 cycle at 72°C for 8 minutes and a 15°C hold. For minigenes 3-6 and 9-11, amplification was performed with PCR conditions of 72°C for 4 minutes, 35 cycles with denaturation at 94°C for 30 seconds, annealing at 55°C for 1 minute and 45 seconds, and extension at 72°C for 7 minutes and followed by a 15°C hold.

Subsequent cloning of 5-7 mutant minigenes was hindered by the large size of the minigene (2500bp). An additional primer was therefore designed (BRCA-5-7INT-F7, sequence: AGCCTGGGCAACATAGTGA) located within the 5-6 intron and this was used instead of PL3 with 5-7 forward megaprimers to amplify a 1390 bp 5' section of the minigene containing the central exon. This was subsequently digested with Pst I and Xba I restriction enzymes (buffer H) to produce a 1196bp sequence which could be directly cloned into pCR3.1 vector containing the wildtype 5-7 minigene, as shown in figure 6.4.



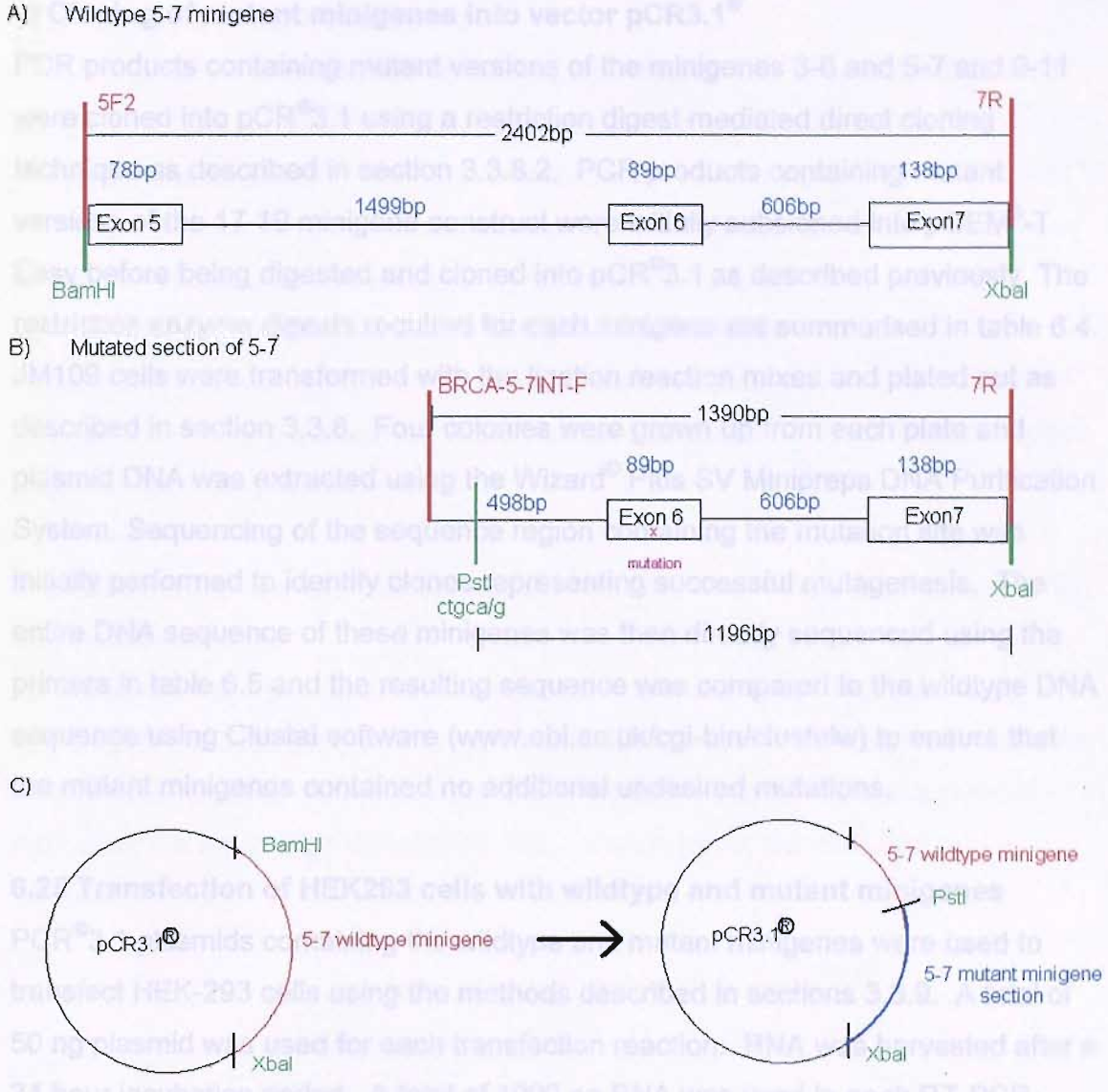


Figure 6.4: *Schematic representation of 5-7 minigene constructs. Diagram A) illustrates the 2402bp complete wildtype 5-7 minigene. Diagram B) illustrates the shortened mutant 5-7 minigenes created using BRCA-5-7INT-F as the 3' primer in the second PCR of the megaprimer mutagenesis method. C) Following digestion with Pst I and Xba I, the resultant 1196bp fragment can be directly cloned into pCR3.1® containing the wildtype 5-7 minigene.*

c) Cloning of mutant minigenes into vector pCR3.1®

PCR products containing mutant versions of the minigenes 3-6 and 5-7 and 9-11 were cloned into pCR®3.1 using a restriction digest mediated direct cloning technique as described in section 3.3.8.2. PCR products containing mutant versions of the 17-19 minigene construct were initially subcloned into pGEM®-T Easy before being digested and cloned into pCR®3.1 as described previously. The restriction enzyme digests required for each minigene are summarised in table 6.4. JM109 cells were transformed with the ligation reaction mixes and plated out as described in section 3.3.8. Four colonies were grown up from each plate and plasmid DNA was extracted using the Wizard® Plus SV Minipreps DNA Purification System. Sequencing of the sequence region containing the mutation site was initially performed to identify clones representing successful mutagenesis. The entire DNA sequence of these minigenes was then directly sequenced using the primers in table 6.5 and the resulting sequence was compared to the wildtype DNA sequence using Clustal software (www.ebi.ac.uk/cgi-bin/clustalw) to ensure that the mutant minigenes contained no additional undesired mutations.

6.25 Transfection of HEK293 cells with wildtype and mutant minigenes

PCR®3.1 plasmids containing the wildtype and mutant minigenes were used to transfect HEK-293 cells using the methods described in sections 3.3.9. A total of 50 ng plasmid was used for each transfection reaction. RNA was harvested after a 24 hour incubation period. A total of 1000 ng RNA was used in each RT-PCR reaction. The resultant cDNA was amplified by adding 21 µl of a reaction mix containing 10 µl 10x buffer, 8 µl 10mM dNTPs, 1 µl PL3 primer (50 pm/ l), 1 µl PL4 primer and 63.5 µl sterile water to 4 µl of each cDNA sample to be amplified. PCR conditions were: 94°C for 20 seconds, 20 cycles with denaturation at 94°C for 30 seconds, annealing at 56°C for 1 minute and extension at 72°C for 2 minutes, 1 cycle at 72°C for 8 minutes and a 15°C hold.

All transfections were repeated on two separate occasions. All amplifications of cDNA were repeated at least three times.

6.26 Analysis of cDNA products

CDNA products were analysed by running volumes of 5-12 µl on polyacrylamide gels, run at 100v for 90 minutes. Images of gels were produces using the DigiDoc-IT system and software, (Ultra-Violet Products Ltd., Cambridge, UK). *Phoretix*™ 1D Advance software (Nonlinear Dynamics Ltd., Newcastle upon Tyne) was used to analyse the ratio of concentration of different transcripts. cDNA bands of interest were excised and DNA was extracted as described in section 3.3.4. Further analysis was performed by either direct sequencing of the extracted DNA sample or by cloning the purified PCR product into pGEM®-T Easy for subsequent sequencing of the plasmid.

6.27 Prediction of effects of sequence variants on splicing regulation using ESEfinder software

All wildtype minigene sequences and sequence variants were analysed using the online ESE predictor programme ESEfinder [163]. The score for a given sequence was considered to be potentially significant if greater that the default threshold defined in the input page (SRp55 =2.676, SRp40=2.670, SC35=2.383 and SF2/ASF=1.956).

6.3 Results

6.31 Identification of missense mutations

Search of the BIC database for reported missense mutations within BRCA1 exons 5,6,10 and 18 identified 40 distinct mutations; 15 within exon 5, 8 within exon 6, 4 within exon 10 and 13 within 18. The number of separate reports for each missense mutation varied between 1 and 222 (updated October 2007). Sixteen of the missense mutations have been reported on only one occasion. Thirty-eight of the forty nucleotide alterations are non-synonomous. Five of these have recently been identified by the BIC congress as being "clinically important" as available data suggests that these sequence changes interfere with gene function and result in an increased risk of cancer. Only two of the reported DNA sequence alterations are synonomous changes: a G to T exchange at nucleotide 231 (exon 6) and a C to T exchange at 5113 (exon 18). Details of each missense mutations including predicted effect on amino-acid sequence and identity of the reporter(s) are shown in table 6.7.

Exon	Mutation site (HUGO)	Codon	Nucleotide Change	Predicted effect on a-a sequence	Name	No. of reports	Source	Reporter
5	135	45	A>T	Lys to Asn	K54N	3	BIC	Myriad
5	139	47	T>G	Cys to Gly	C47G	1	BIC	Casey
5	140	47	G>T	Cys to Phe	C47F	2	BIC	Myriad
5	154	52	C>T	Leu to Phe	L52F	5	BIC	Myriad, Casey
5	181	61	T>G	Cys to Gly*	C61G	222	BIC	Myriad, others
5	181	61	T>C	Cys to Arg	C61R	1	BIC	Bond
5	182	61	G>A	Cys to Tyr	C61Y	6	BIC	Myriad
5	189	63	A>T	Leu to Phe	L63F	1	BIC	Ostrow
5	190	64	T>G	Cys to Gly	C64G	4	BIC	Myriad, others
5	190	64	T>C	Cys to Arg	C64R	11	BIC	Myriad,Radice
5	191	64	G>A	Cys to Tyr*	C64Y	19	BIC	Myriad, others
5	199	67	G>T	Asp to Tyr	D67Y	8	BIC	Myriad
5	201	67	T>G	Asp to Glu	D67E	3	BIC	Patmisiriwat
5	203	68	T>A	Ile to Lys	I68K	1	BIC	Myriad
5	211	71	A>G	Arg to Gly*	R71G	35	BIC	Myriad, others
6	216	72	C>A	Ser to Arg	S72R	1	BIC	Myriad
6	230	77	C>T	Tyr to Met	T77M	2	BIC	Myriad,Radice
6	231	77	G>T	Thr to Thr		1	BIC	Casey
6	259	87	T>G	Leu to Val	L87V	2	BIC	Myriad
6	266	89	T>C	Ile to Thr	I89T	2	BIC	Myriad, Caldes
6	269	90	T>C	Ile to Thr	I90T	1	BIC	Weber
6	286	96	G>A	Asp to Asn	D96N	1	BIC	Myriad
6	292	98	G>C	Gly to Arg	G98R	1	BIC	Diez
10	612	204	G to C	Leu to Phe	L204F	6	BIC	Myriad, Lidereau
10	637	213	A>G	Arg to Gly	R213G	1	BIC	Myriad
10	641	214	A>G	Asp to Gly	D214G	11	BIC	Myriad, Cheetham
10	661	221	G>T	Ala to Ser	A221S	2	BIC	Myriad
18	5085	1695	T to A	Phe to Leu	F1695L	1	BIC	Myriad
18	5086	1696	G to C	Val to Leu	V1696L	1	BIC	Myriad
18	5089	1697	T toC	Cys to Arg	C1697R	2	BIC	Borg, Bergthorsson
18	5095	1699	C to T	Arg to trp*	R1699W	13	BIC	Myriad, others
18	5096	1699	G to A	Arg to Gln	R1699Q	11	BIC	Myriad, Abbs
18	5096	1699	G to T	Arg to Leu	R1699L	1	BIC	Myriad
18	5113	1705	C to T	Leu to Leu		1	BIC	Murphy, Voglino
18	5117	1706	G to A	Gly to Ala	G1706E	7	BIC	Myriad, others
18	5117	1706	G to C	Gly to Ala	G1706A	6	BIC	Myriad, Cheetham
18	5123	1708	C to A	Ala to Glu*	A1708E	45	BIC	Myriad, others
18	5138	1713	T to C	Val to Ala	V1713A	3	BIC	Myriad, Struewing
18	5143	1715	A to C	Ser to Arg	S1715R	1	BIC	Borg
18	5143	1715	A to T	Ser to Cys	S1715C	1	BIC	Myriad

Table 6.7: Details of all missense mutations within BRCA1 exons 5,6,10 and 18 reported to the BIC database by 20.2.2006. Mutations marked by * have been deemed to have a significant clinical effect by the BIC congress following review of all available data

6.32: Analysis of effect of missense mutations on splicing regulation motifs using ESEfinder [163]

The results of *in silico* analysis of the wildtype DNA sequences of BRCA1 exons 5,6,10 and 18 using the ESE predictor software ESEfinder are shown in table 6.8. Thirty-four potential ESE motifs were identified using the default thresholds (SRp55 =2.676, SRp40=2.670, SC35=2.383 and SF2/ASF=1.956): 12 in exon 5, 5 in exon 6, 7 in exon 10 and 10 in exon 18. Exons 5 and 18 were predicted to contain ESEs capable of binding all four SR protein groups (SRp55, Srp40, SC35 and SF2/ASF) whilst exons 10 and 6 were predicted to contain motifs complementary to SRp55, SC35, SF2/ASF and SRp55, SF2/ASF only respectively.

The predictions for the mutated exonic sequences are summarised in table 6.9. This analysis indicates that 33 of the 40, (82.5%), missense mutations are situated within potential ESE motifs. Seventeen of the pre-existing ESE motifs were destroyed by the mutations, five of the ESEs showed a reduced score, (still above threshold value), and four of the ESEs showed an increased score. Additionally, fourteen of the 40 missense mutations, (35.0%), were predicted to create new ESEs. Only 7 of the 40 missense mutations (18.5%) were predicted to have no effect on splicing regulators

Exon	SRp55 Threshold: 1.956			SC35 Threshold: 2.383			SF2/ASF Threshold: 2.67			Srp40 Threshold: 2.676		
	Position	Motif	Score	Position	Motif	Score	Position	Motif	Score	Position	Motif	Score
5	26	CAGAAGA	3.819131	36	GGCCTTCA	2.869066	19	TCTCAAC	3.866663	5	TGTCATG	3.763907
	42	CACAGTG	3.138162	48	GTCCTTTA	2.865828	25	CCAGAAG	3.675556	45	AGTGTA	2.697407
							30	AGAAAGG	2.884213	54	TATGTA	3.610283
							41	TCACAGT	3.942668			
							72	CCAAAAG	3.830855			
6	76	CACAGGT	5.631733				5	CTACAAG	5.087824			
							65	TTTCAGC	4.122436			
							73	TGACACA	2.815500			
							75	ACACAGG	5.274785			
10	28	CACCCCT	2.664563	61	GGATTCTG	4.450894	20	TTACAAA	3.062783			
	33	CTCAAGG	2.216055				27	TCACCCC	3.439169			
							32	CCTCAAG	4.352605			
							63	ATTCTGC	2.827091			
18	4	CTGAGTT	2.142950	22	GGACACTG	4.968561	36	TTTCTAG	4.151518	11	TGTGTG	2.820871
	23	GACACTG	2.164970	43	GAATTGAG	2.530293				13	TGTGTG	2.820871
	50	GGGAGGA	3.714214							47	TGCGGG	3.558695
										60	TGGGTA	2.708481

Table 6.8: *Exonic Splicing Enhancers (ESEs) predicted to exist within BRCA1 exons 5,6,10 and 18 as predicted by the in silico prediction programme ESEfinder [163]*

Exon	Position	Nucleotide change	ESE affected	Position in exon	Motif	Wildtype score	Mutation score	Predicted Effect
5	135	A.>T	Nil					
5	139	T>G	SRp55	5	TGCATG	3.763907	-	-
5	140	G>T	SC35	4	TTTCATG	-	2.942525	+
			SRp55	5	TGCATG	3.763907	-	-
5	154	C>T	SRp40	19	TTTCAAC	3.866663	3.554607	↓
5	181	T>G	SRp40	41	TCACAGG	3.942668	6.324404	↑
			SF2/ASF	42	CACAGGG	3.138162	5.258191	↑
			SRp55	45	AGTGTC	2.697407	-	-
5	182	G>A	SF2/ASF	42	CACAGTA	3.138162	3.864944	↑
			SRp55	45	AGTGTC	2.697407	-	-
5	189	A>T	SC35	48	GTCCTTTA	2.865828	-	-
			SRp55	54	TATGTA	3.610283	-	-
5	190	T>G	SRp55	54	TATGTA	3.610283	-	-
5	191	G>A	SRp55	54	TATATA	3.610283	3.001623	↓
5	199	G>T	SRp55	65	TATATA	-	3.610283	+
5	201	T>G	Nil					
5	203	T>A	Nil					
5	211	A>G	SRp40	72	CCAAAGG	3.830855	4.398684	↑
6	216	C>A	SF2/ASF	1	GAGACTA	-	2.645522	+
			SRp40	2	AGACTAC	-	2.797046	+
			SC35	3	GACTACAA	-	3.914543	+
6	230	C>T	Nil					
6	231	G>T	SRp40	15	GTACTAG	-	2.694455	+
6	259	T>G	SC35	42	AGCTAGTG	-	3.081021	+
6	266	T>C	Nil					
6	269	T>C	SRp40	54	TCACTTG	-	4.056130	+
6	286	G>A	SRPp40	73	TGACACA	2.815500	-	-
6	292	G>C	SF2/AS	74	GACACAA	-	2.434044	+
			SRp40	75	ACACAAG	5.274785	4.706956	↓
10	612	G>C	Nil					
10	637	A>G	Nil					
10	641	A>G	SF2/ASF	43	CAGGGGT	-	2.476925	+
10	661	G>T	SRp40	63	ATTCTGC	2.827091	-	-
18	5085	T>A	SC35	7	AGTTAGTG	-	2.512089	+
			SRp55	11	TGTGTG	2.820871	-	-
18	5086	G>C	SC35	7	AGTTTCTG	-	3.140572	+
			SRp55	11	TGTGTG	2.820871	-	-
18	5089	T>C	SRp55	11	TGTGTG	2.820871	-	-
18	5095	C>T	SRp40	17	TGAATGG	3.688106	+	+
18	5096	G>A	SC35	22	AGACACT	4.968561	3.216489	↓
18	5096	G>T	SC35	22	TGACATG	4.968561	2.9188919	↓
			SRp40	22	TGACACT	-	2.815500	+
18	5113	C>T	SRp40	36	TTTCTAG	4.151518	-	-
18	5117	G>A	SC35	43	GAATTGCG	2.530293	-	-
18	5117	G>C	SC35	43	GAATTGCG	2.530293	-	-
18	5123	C>A	SRp40	47	TGCGGG	3.558695	-	-
18	5138	T>C	SRp55	60	TGGGTA	2.708481	-	-
18	5143	A>C	SC35	66	GTTCGCTA	-	3.887473	+
18	5143	A>T	SC35	66	GTTTGCTA	-	3.656632	+

Table 6.9: *Predicted effect of missense mutations on ESEs within BRCA1 exons 5,6,10 and 18 according to analysis with ESEfinder algorithm (-= loss of ESE, + =gain of ESE, ↓= reduced score but still above threshold, ↑ = increased score).*

6.33 Results of transfection of wildtype minigenes

All four wildtype minigenes were successfully transfected into HEK 293 cells. Examination of the cell culture plates by light microscopy 24 hours after transfection confirmed that all cells remained viable at this time and RNA was subsequently harvested. Samples of the resultant cDNA were amplified using 20, 25 and 30 cycles of PCR and visualised on 6.8% polyacrylamide gels. The optimal appearance was achieved using 25 cycles of PCR (data not shown).

All wildtype minigenes produced transcripts of more that one size. The 17-19 minigene produced a dominant transcript of approximately 383 bp and a second extremely faint transcript of 883 bp. Direct sequencing confirmed that the 383bp transcript contained the anticipated wildtype cDNA sequence of exons 17-18-19 (plus vector flanking sequences) whilst the 883bp transcript consisted of the wildtype exons17-18-19 with inclusion of the 500bp intron 18-19. Transfection of the 9-11 minigene also resulted in 2 transcripts: a dominant 401bp transcript, confirmed by direct sequencing to represent the wildtype cDNA sequence of exons 9-10-11 (plus vector flanking sequences), and a slightly weaker 324bp transcript. The latter transcript contained wildtype exon 9 adjacent to exon 11 with deletion of the entire central 77bp exon 10.

Transfection of the 5-7 wildtype minigene produced four separate transcripts as shown in figure 6.5. Indirect sequencing (of plasmids produced from cloning individual transcripts into pGEM®-T Easy) confirmed that the dominant 433bp transcript contained the anticipated wildtype cDNA transcript of exons 5-6-7 (with flanking vector sequences), and the 344bp transcript consisted of wildtype exons 5-7 with complete deletion of the 89bp exon 6. Both of the 2 additional transcripts proved to be examples of cryptic splicing. The 411bp fragment contained exons 5-6-7 with deletion of the last 22 base pairs of exon 5 and the 619bp fragment represented exons 5-6-7 with inclusion of the first (5') 186bp of the 5-6 intron. In addition to these four transcripts, visualization of products of PCR amplification of the 5-7 transfection products on poyacrylamide gels revealed an apparent cDNA

149band of approximately 1100bp. Repeated attempts to sequence these bands by both direct sequencing and sequencing of plasmids produced from cloning these bands into pGEM®-T Easy vector confirmed that these large cDNA bands represented only polymerization of the wildtype transcript.

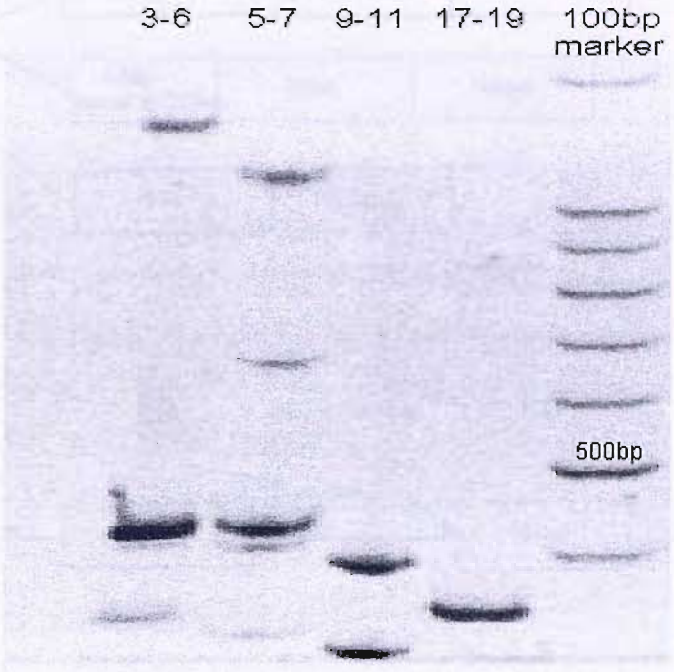


Figure 6.5: *Polyacrylamide gel showing cDNA transcripts resulting from transfection of control minigenes into HEK 293 cells; RNA was harvested after 24 hours*

Transfection of the 3-6 wildtype minigene resulted in 3 cDNA transcripts. Sequencing confirmed that the 450bp fragment represented the anticipated wildtype exon 4-5-6 cDNA (plus vector flanking sequences) and the 372bp transcript consisted of exons 3-6 with complete deletion of exon 5. The third fragment, 423bp, showed cryptic splicing, containing exons 3-5-6 with deletion of the last 22 nucleotide pairs of exon 5 as seen with transfection of the 5-7 minigene. Again, an additional large cDNA band (of approximately 1300bp) was seen on visualisation of polyacrylamide gels –this was confirmed to represent polymerization of the wildtype cDNA sequence.

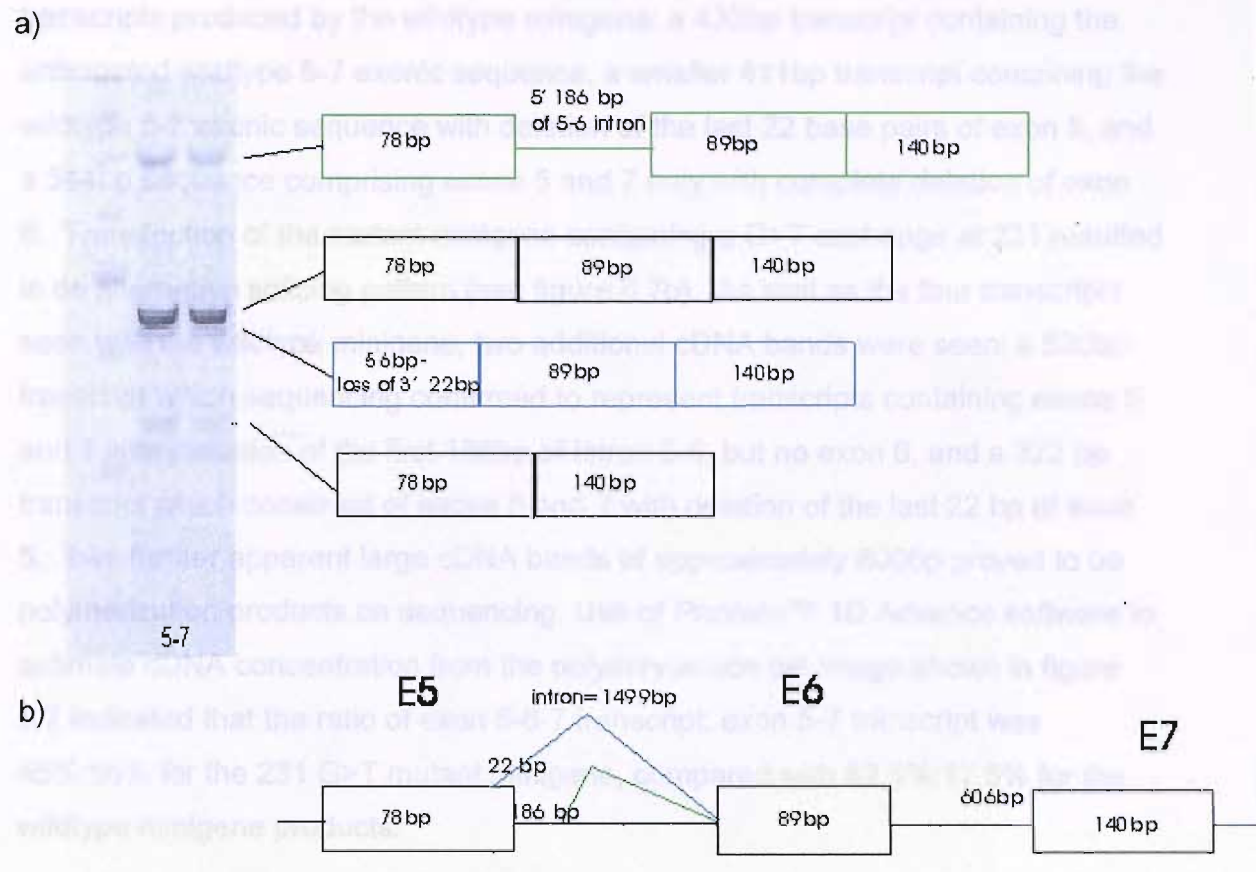


Figure 6.6: a) *Diagrammatic representation of the four separate transcripts produced by transfection of the BRCA1 exon 5-7 minigene into HEK 293 cells, b) Representation of the two examples of cryptic splicing: loss of the last 22 base pairs of exon 5 and inclusion of 186 base pairs of the 5-6 intron.*

6.34 Results of transfection of BRC1 exon 5-7 mutant minigenes

All eight mutant versions of the BRCA1 exon 5-7 minigene were successfully transfected into HEK 293 cells. Visualisation on a 6.4% polyacrylamide gel of the amplification products of cDNA made from RNA harvested 24 hours after transfection showed that all the mutant minigenes except the minigene containing a G>T nucleotide change at 231 produced the same sized cDNA fragments as the wildtype 5-7 minigene (see figure 6.7a). Indirect sequencing of the cDNA bands for all mutant minigenes confirmed that these were identical in sequence to the four

transcripts produced by the wildtype minigene: a 433bp transcript containing the anticipated wildtype 5-7 exonic sequence, a smaller 411bp transcript containing the wildtype 5-7 exonic sequence with deletion of the last 22 base pairs of exon 5, and a 344bp sequence comprising exons 5 and 7 only with complete deletion of exon 6. Transfection of the mutant minigene containing a G>T exchange at 231 resulted in an alternative splicing pattern (see figure 6.7b). As well as the four transcripts seen with the wildtype minigene, two additional cDNA bands were seen: a 530bp transcript which sequencing confirmed to represent transcripts containing exons 5 and 7 with inclusion of the first 186bp of intron 5-6, but no exon 6, and a 322 bp transcript which consisted of exons 5 and 7 with deletion of the last 22 bp of exon 5. Two further apparent large cDNA bands of approximately 800bp proved to be polymerization products on sequencing. Use of *Phoretix*™ 1D Advance software to estimate cDNA concentration from the polyacrylamide gel image shown in figure 6.7 indicated that the ratio of exon 5-6-7 transcript: exon 5-7 transcript was 45%:55% for the 231 G>T mutant minigene, compared with 82.5%:17.5% for the wildtype minigene products.

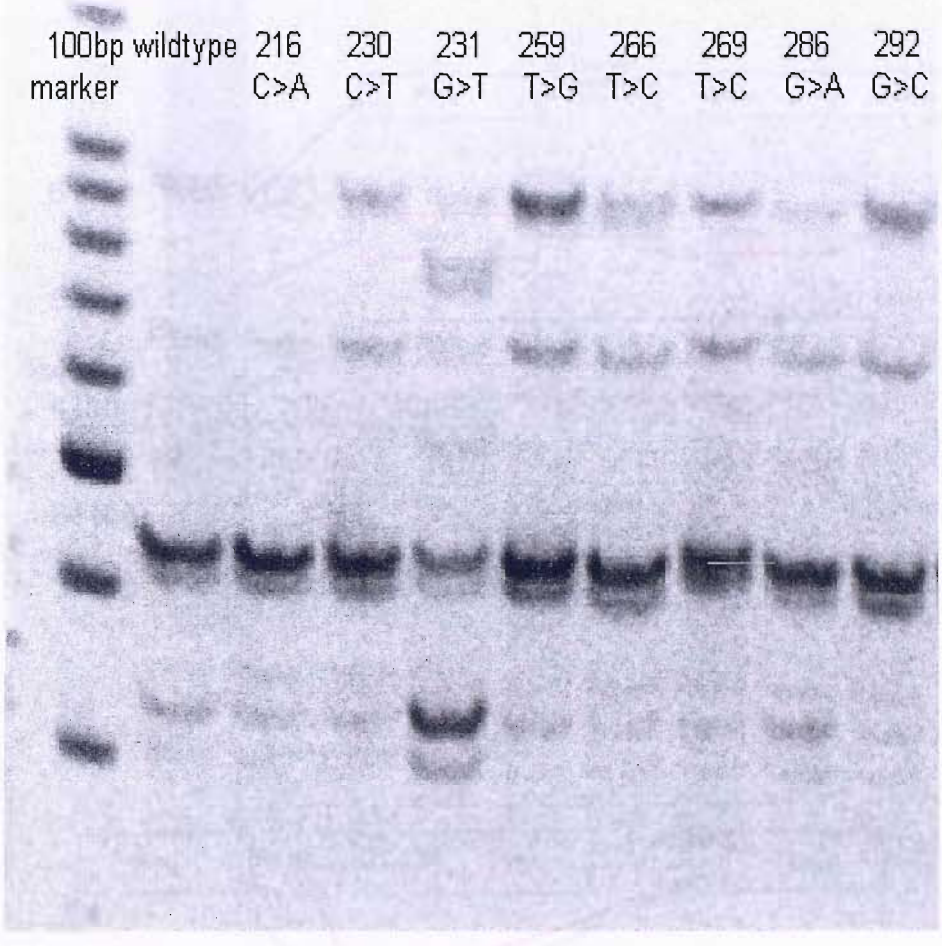


Figure 6.7: *Polyacrylamide gel electrophoresis of transcripts obtained from transfection into HEK293 cells of BRCA1 exon 5-7 minigenes containing either wildtype DNA sequence or one of eight different missense mutations. All missense mutations produced the same splicing pattern as the wildtype minigene except 231 G>T.*

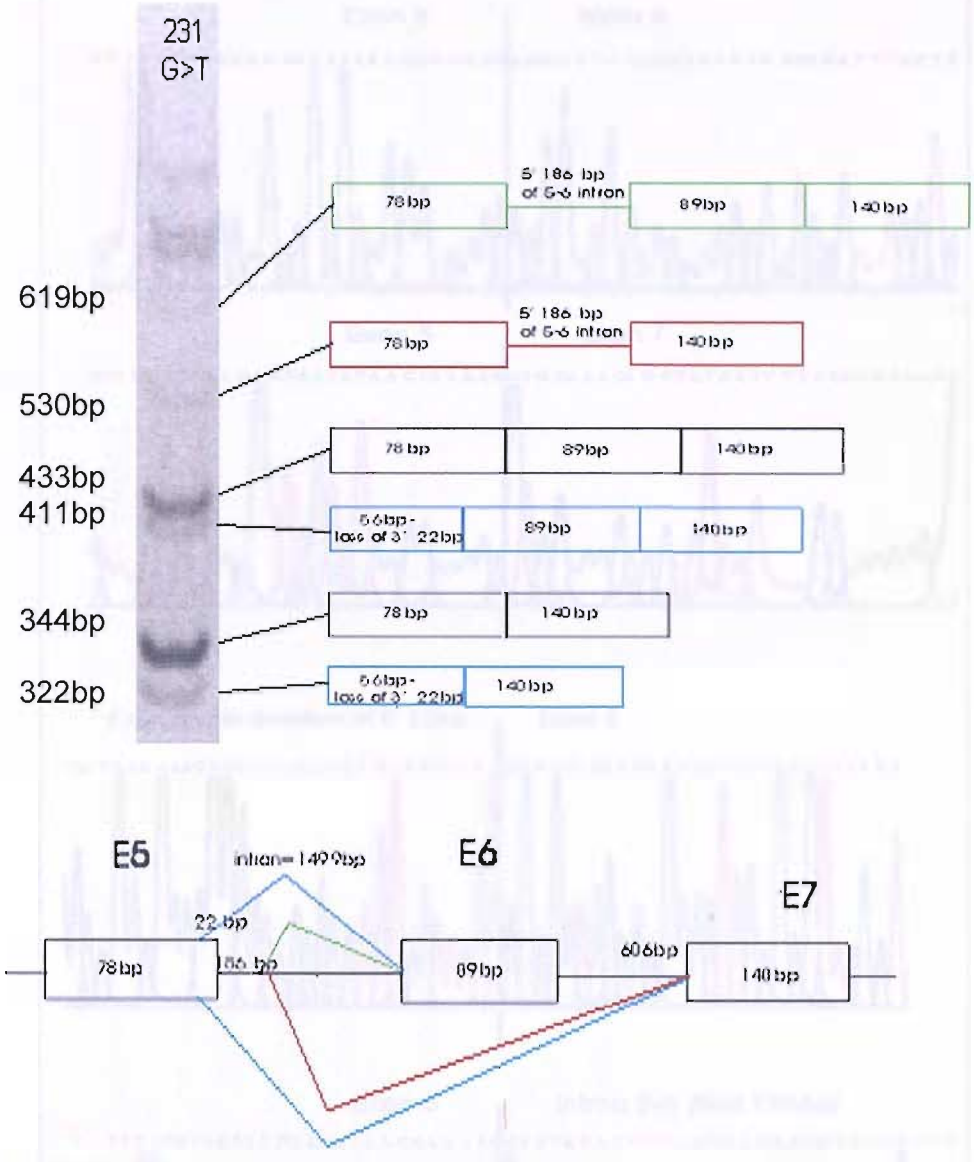


Figure 6.8: *Diagram illustrating transcripts produced by transfection of BRCA1 exon 5-7 minigene containing G>T mutation at 231: a 433bp transcript contains the anticipated exon 5-6-7cDNA band, the 411bp transcript contains this with deletion of the last 22bp of exon5, a 344bp band contains exons 5-7 with deletion of exon 6, the 322 band contains this transcript with deletion of the last 22bp of exon 5, and the 619bp band contains exons 5-6-7 with inclusion of 186bp of intron 5-6 and a 530bp containing this arrangement with deletion of exon 6.*

a)

b)

c)

d)

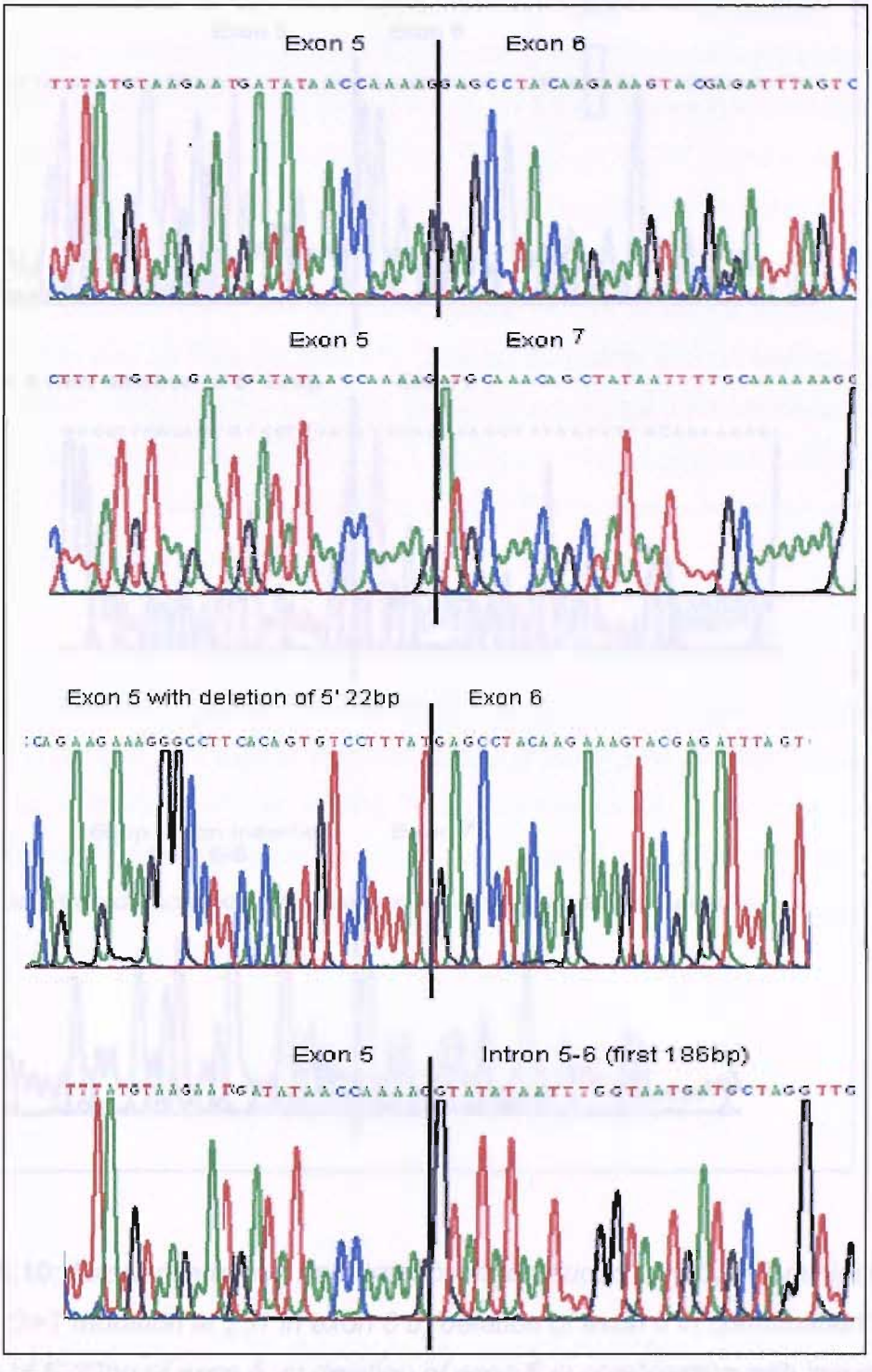


Figure 6.9: *Sequencing traces from transcripts produced by all 5-7 minigenes demonstrating a) wildtype exon 5-6 sequence, b) deletion of exon 6, c) insertion of 186bp portion of 5-6 intron after exon 5, d) deletion of 5'22bp of exon 5*

a)

b)

c)

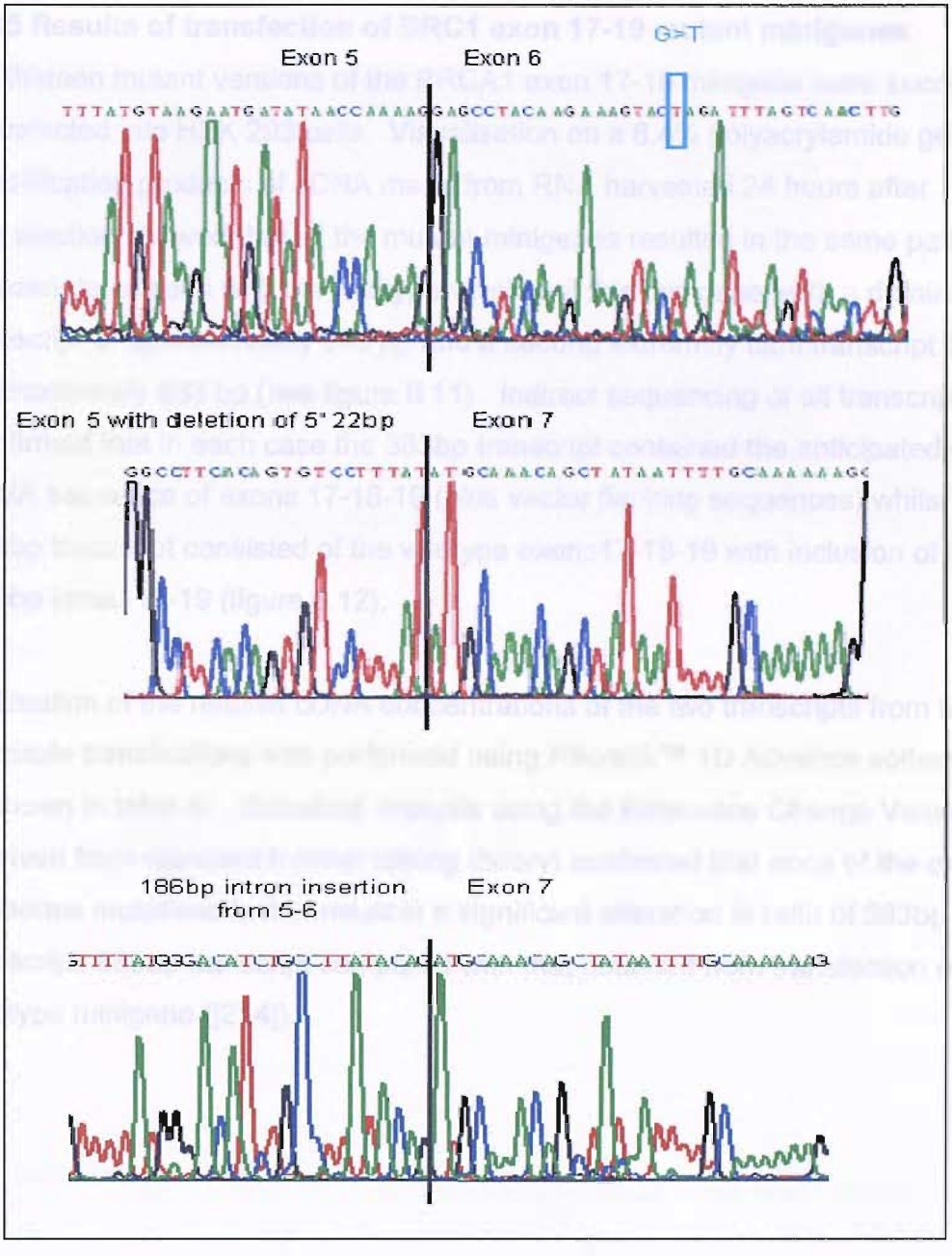


Figure 6.10: *Sequence traces from transcripts produced only by 5-7 minigene 231 G>T: a) G>T mutation at 231 in exon 6 b) deletion of exon 6 in combination with deletion of 5' 22bp of exon 5, c) deletion of exon 6 in combination with insertion of first 186bp of intron 5-6.*

6.35 Results of transfection of BRC1 exon 17-19 mutant minigenes

All thirteen mutant versions of the BRCA1 exon 17-19 minigene were successfully transfected into HEK 293 cells. Visualisation on a 6.4% polyacrylamide gel of the amplification products of cDNA made from RNA harvested 24 hours after transfection showed that all the mutant minigenes resulted in the same pattern of transcripts as seen with the wildtype version of this minigene, with a dominant transcript of approximately 383 bp and a second extremely faint transcript of approximately 883 bp (see figure 6.11). Indirect sequencing of all transcripts confirmed that in each case the 383bp transcript contained the anticipated wildtype cDNA sequence of exons 17-18-19 (plus vector flanking sequences) whilst the 883bp transcript consisted of the wildtype exons17-18-19 with inclusion of the 500bp intron 18-19 (figure 6.12).

Estimation of the relative cDNA concentrations of the two transcripts from two separate transfections was performed using *Phoretix*™ 1D Advance software and is shown in table 6. Statistical analysis using the Reference Change Value (derived from standard Normal testing theory) confirmed that none of the exon 18 missense mutations tested result in a significant alteration in ratio of 383bp transcript: 883bp transcript compared with that obtained from transfection of the wildtype minigene ([214]).

a)

b)

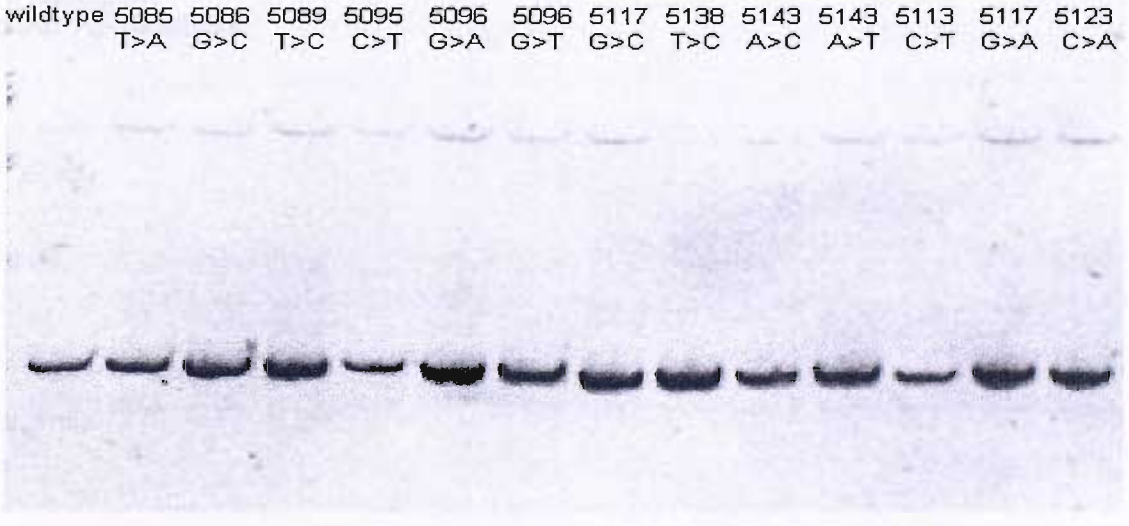


Figure 6.11: *Sample polyacrylamide gel electrophoreses* of *transcripts obtained from a) first and b) second transfections into HEK293 cells of BRCA1 exon 17-19 minigenes containing either wildtype DNA sequence or one of thirteen different missense mutations. All missense mutations produced the same splicing pattern as the wildtype minigene.*

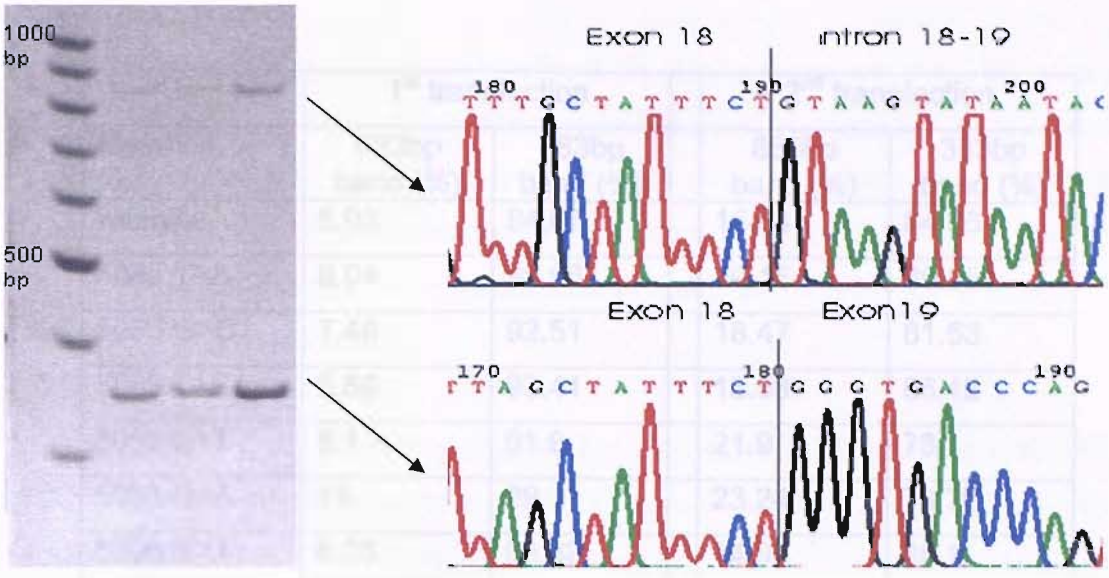


Figure 6.12: *Sequencing of the 383bp transcript for the wildtype and each mutant 17-19 minigene confirmed that this contained the anticipated BRCA1 exon 17-18-19 transcript. Sequencing of the 883bp transcript for the wildtype and each mutant 17-19 minigene indicated that this contained the 17-18-19 transcript with inclusion of the 18-19 intron.*

	1st transfection			2nd transfection	
Mutation	883bp band (%)	383bp band (%)		883bp band (%)	383bp band (%)
wildtype	5.93	94.07		15.64	84.36
5085 T>A	8.04	91.96		10.15	89.85
5086 G>C	7.49	92.51		18.47	81.53
5089 T>C	6.59	93.41		13.58	86.42
5095 C>T	8.1	91.9		21.9	78.1
5096 G>A	11	89		23.24	76.76
5096 G>T	6.38	93.62		14.9	85.1
5117 G>C	7.5	92.5		20.01	79.99
5138 T>C	1.53	98.47		5.26	94.74
5143 A>C	6.5	93.5		12.41	87.59
5143 A>T	7.31	92.69		20.39	79.61
5113 C>T	9.27	90.73		5.44	94.56
5117 G>A	8.06	91.94		7.78	92.22
5123 C>A	12.21	87.79		9.02	90.98

Table 6.10: *Relative proportions of 383bp and 585bp transcripts resulting from two separate transfections of BRCA1 exon 17-19 wildtype and mutant minigenes as estimated from polyacrylamide gel electrophoresis images using Phoretix™ 1D Advance software*

6.36: Results from transfection of 3-6 and 9-11 mutant minigenes

During the course of this project, all attempts to clone mutant versions of either the BRCA1 exon 3-6 or exon 9-11 minigene into pCR 3.1 vector were unsuccessful.

6.4: Discussion

In this part of my study I used the *in silico* ESE predictor programme ESEfinder to forecast the effect of BRCA1 missense mutations sited in exons 5,6,10 and 18 on splicing regulation of the BRCA1 gene. The aim was to compare these results with those achieved by *in vitro* studies using minigene constructs to study the effect of the same missense variations. At this point in time, *in vitro* work has been completed for exons 6 and 18.

BRCA1 is a large gene covering a span of approximately 100kb of genomic DNA [11]. The coding region contains some 5500bp nucleotides arranged in 22 exons (an additional 2 codons are non-coding) that are translated to create a protein of 1,863 amino acids [21]. Almost 60% of the coding DNA sequence lies within the large central exon 11. No BRCA1 exon conforms to the conventional pattern of the "ideal" minigene, with a central exon of no more than 200bp and flanking introns of less than 1500bp. The four BRCA1 exons selected for investigation in this project represented compromises between the optimal intron/ exon dimensions whilst still containing a significant number of reported missense mutations. In the case of exons 5 and 18 this necessitated production of minigenes with truncation of the proximal intron. Care was taken to ensure that an intronic sequence of at least 125 nucleotides was left intact immediately adjacent to the intron/ exon boundary as evidence largely suggests that this is the main region involved in splicing regulation [215]. However, there have been suggestions that motifs up to 300bp and even 1000bp away from an intron/exon boundary could potentially have an effect on splicing and therefore we cannot completely exclude the possibility that deletion of part of the flanking intron may have influenced the splicing patterns seen [212, 213, 216]. This point may be particularly relevant to the analysis using the 17-19 minigene, where technical considerations of primer design resulted in a 5' native intronic sequence of only 141bp. The same minigene structure was maintained for both wildtype and mutant transfections, ensuring that splicing patterns could be legitimately compared.

The missense mutations analysed in this project were identified by searching the information held by the Breast Information Core, an open access on-line breast cancer mutation data base, (http://nhgri.nih.gov/intramural_research/ lab_transfer/bic/). This facility acts as a central repository for information regarding mutations and polymorphisms of the inherited breast cancer genes and represents an international collaborative effort hosted by the National Genome Research Institute, USA. This method of identifying BRCA1 missense mutations has limitations. The reporting of DNA sequence variations to the BIC database is entirely voluntary and it is well recognised that the number of reports documented for each distinct mutation does not provide accurate information on the frequency of the mutation as there is a tendency for users to not enter their own finding if that sequence variant has already been reported. Historically there have also been problems due to the use of different nomenclature systems for BRCA mutations. To ensure that the correct mutation was identified at transcript level, I checked each documented exon 5,6,10 and 18 BRCA1 sequence variant for correspondence with the GenBank U14680 cDNA sequence and verified the predicted effect on amino-acid sequence with the translated protein sequence (GenBank AAA73985). Ongoing efforts by the BIC database curators to standardise all reports to HUGO nomenclature should minimise future problems with this. The vast majority of BIC database reports do not provide references to peer reviewed journals but where these were given the stated DNA sequence variation was verified with the original published data. Mutations defined as splice site mutations by BIC were not included in this analysis as they represent nucleotide changes in the first or last 3 base pairs of each exon which would clearly be expected to alter splicing patterns by directly affecting the splice donor or acceptor site; mutations that destroy natural donor sequences are usually observed to cause skipping of their associated exon [156] [166].

Forty distinct missense mutations were identified on searching the BIC database entries relevant to the exons of interest: 15 within exon 5, 8 within exon 6, 4 within exon 10 and 13 within 18. The incidence of reporting for each mutation varied

between 1 and 222 times; 16 of the mutations were unique reports and fourteen of these had been reported only by the American genetic screening company Myriad. All but two of the reported mutations were non-synonomous. Recent evaluations by the BIC congress have identified five of the forty sequence changes as being "clinically important" where available data suggests that these sequence changes interfere with gene function and result in an increased risk of cancer. These also represented the five most commonly reported missense mutations of those analysed. Two of the three "clinically important" exon 5 missense mutations, C61G and C64Y, have been defined as such because they occur within the conserved ring domain and functional studies have demonstrated that these mutations block BRCA1 E3 ubiquitin ligase activity [217]. The third clinically important exon 5 missense mutation, R71G, has been shown to disrupt splicing [218]. The two clinically significant exon 18 missense mutations, A1708E and R1699W, were both considered to be deleterious because of evidence of conservation of these residues within the BRCT domain, lack of co-occurrence with nonsense mutations and functional data [109,111,116, 219, 220].

In silico analysis of the effects of the identified BRCA1 exon 5,6,10 and 18 missense mutations on splicing regulation was initially performed using the web-based ESE predictor programme "ESEfinder" [163]. ESE sequences bind to proteins from the arginine/ serine rich family (SR). ESEfinder acts by searching DNA for consensus sequences that would bind the SR proteins SRp55, Srp40, SC35 and SF2/ASF, as determined using the systematic evolution of ligands by exponential enrichment technique (SELEX),[136] [163]. Input sequences are scored according to match with the consensus sequences and regions scoring more than a default threshold value are predicted to bind SR proteins and therefore act as functional ESEs.

Analysis of the wildtype BRCA1 cDNA sequence using ESEfinder set to the default threshold values (SRp55 =2.676, SRp40=2.670, SC35=2.383 and SF2/ASF=1.956) identifies a total of 669 potential ESE motifs within the coding

region of this gene, with 34 potential ESEs occurring within exons 5,6,10 and 18. Our analysis using ESEfinder indicated that all but 7 of the 40 missense mutations studied here could affect splicing regulation in some way by interfering with potential ESE motifs. Twenty of the mutations either destroyed seventeen of the identified ESE motifs and/ or altered the score (nine cases). Additionally, fourteen of the 40 missense mutations, (35.0%), were predicted to create new ESEs. In several cases, a single nucleotide alteration was predicted to have opposite effects on potential complementary sites for different SR proteins, such as a G to T exchange at 140 which apparently has the capacity to create a new SC35 motif and destroy a pre-existing SRp55 motif. The complex nature of the data provided by the ESEfinder algorithm make it impossible to use as a lone agent to assess the clinical significance of a mutation.

There have now been several attempts to compare the results gained from ESEfinder analysis with those achieved by *in vivo* studies of the effects of exonic mutations on splicing regulation and all have shown significant disparity between the predictions of the algorithm and the outcomes of functional splicing assays [171, 173,174,175]. Comparison of systematic screening for aberrant splicing in MLH1 and MSH2 genes in a total of 60 HNPCC patients who carried different exonic or intronic mutations using a RT-PCR based analysis with results of predictions of ESEs by ESEfinder revealed that his programme has a sensitivity of 80% and a specificity of 42% [174].

To the best of our knowledge this was the first attempt to develop a systematic *in vitro* method of assessing BRCA1 missense mutations for functional significance in splicing regulation. All four minigenes were successfully transfected into HEK 293 cells. Assessment of cell viability after an incubation period of 24 hours confirmed that transfection of the wildtype minigenes was not detrimental to cell survival. Subsequent analysis of the transcripts derived from RNA harvested at this timepoint indicated that all four control minigenes produced more than one transcript. In each case the dominant transcript was the anticipated wildtype cDNA

sequence containing three correctly spliced exons. In the case of the exon 17-19 minigene, the faint additional cDNA band seen consisted of the three wildtype exons with insertion of the 18-19 intron. Each of the other control minigenes produced transcripts exhibiting deletion of the central exon.

It is very difficult in the context of *in vitro studies* to hypothesise whether these examples of intron insertion and exon deletion occur *in vivo* and are of biological relevance or whether they simply represent artefacts of the minigene assay. There is little data available on the frequency of complete intron retention as an alternative splicing event; computational investigations of alternative splicing have tended to exclude this phenomenon from analysis because of the difficulty in distinguishing true intron retention from contamination of EST databases by pre-mRNA or genomic material [221]. One recent report has however estimated the frequency of "normal" intron retention events in the human genome to be in the region of 15% [222]. The biological significance of most of these events is unknown but there are examples in the literature of aberrant intron retention events in association with pathological splicing defects, including the association of phaeochromocytomas with intron retention in the Ret tyrosine kinase gene [223]. Retention of the BRCA1 18-19 intron has not been previously reported.

Isolated BRCA1 deletions of exons 5 and 6 have not been described in previous reviews of alternative splicing of BRCA1 [149]. Deletion of exon 5 alone is an in-frame event but deletion of exon 6 results in a frame shift with creation of a new stop codon at codon 76. These cDNA bands demonstrating exon 5 and exon 6 deletions are faint in comparison to that representing the exon 10 deletion. The deletion of BRCA1 exon 10 is a frequently reported phenomenon and this result is therefore less likely to represent pure technical artefact. It is likely to reflect the known "leakiness" of the exon 11 acceptor site. However, most reports describe deletion of exon 10 in association with a complete deletion of exon 9 (reviewed by [149]).

Two examples of cryptic splicing were demonstrated by the control. Deletion of the final 22 nucleotides of exon 5 was seen with both the 3-6 and 5-7 minigenes. This alternative transcript, now known as BRCA1-Δ22deltantex5 has been previously reported at low levels in controls [207] and was also found by myself during my investigation of alternative splicing using overlapping primers (chapter 5.0). Deletion of this sequence results in a frame shift alteration after codon 63 and is predicted to create a premature termination at codon 80 in exon 6. The shortened transcript would however still include the BRCA1 RING motif which is central to many protein-protein interactions.

It is interesting to note that one previous publication has reported the association of the 211 A>G missense mutation with occurrence of an alternative transcript showing deletion of the 5' 22 nucleotides of exon 5 in equal concentration to the wildtype transcript [218]. The 211 nucleotide coincides with the –2 position of the exon 5 donor splice site and Vega *et al*, suggested that disruption of the donor splice site by this sequence change activates a cryptic splice site 22 base pairs from the exon boundary. My work indicates that deletion of the 5' 22 base pairs of exon 5 is a common alternative transcript, found in controls and carriers of BRCA1 mutations. This type of report exemplifies the need to compare splicing patterns found in association with mutations with those produced by controls.

Analysis of the exon 5-6 boundary with programmes such as the Gil Ast Laboratory Analyser Site Score (http://www.tau.ac.il/~gilast/) indicates that the exon 5 donor sites is relatively weak (splice site score = 77.93) and therefore prone to alternative splicing arrangements [224]. It should be noted that *in vivo* one would expect a deletion of 22 base pairs to create an out-of frame transcript that would be likely to be degraded by the nonsense-mediated decay pathway. The deletion of 22 nucleotides from the end of exon 5 changes codon 64 from TGT (which encodes Cysteine) to the termination codon TGA (see figure 6.13).

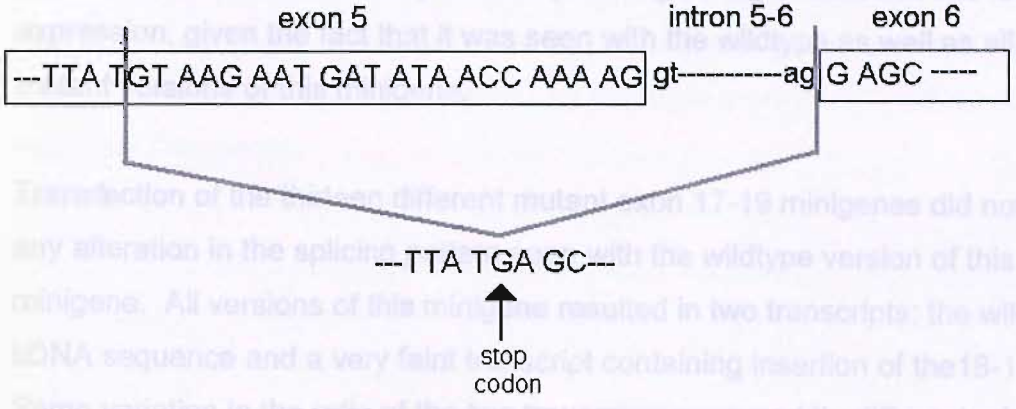


Figure 6.13: *Diagram illustrating effect of deletion 22 nucleotides from 5' end of exon 5 on cDNA sequence of BRCA1. A new termination codon, TGA, is created by the splicing of the exon 5 codon 64T to the GA base pairs at the 3' end of exon 6.*

The second example of alternative splicing produced by the control minigenes was the insertion of the first 186bp of the 5-6 intron between exons 5 and 6 following transfection of the 3-6 minigene. The human genome seems to display an abundance of potential "pseudoexons", intronic sequences that have the potential ability to encode for "false" exons, typically 50-200 nucleotides in length with viable 5' and 3' splice sites at either end [225]. The 5' region of the BRCA1 intron is a an *Alu* rich region, (a class of particularly abundant intronic sequences), and of interest is the recent proposal that the "exonisation" of *Alu* repeats may represent a major and frequent source of "new" exons and therefore play an important role in evolution [226]. One BRCA1 pseudoexon, an in-frame 66 base pair inclusion between exons 13 and 14, now designated 13A has previously been described here in detail (chapter 4) and in the published literature [210]. It is again worth noting that this pseudoexon had previously been described in association with a family of early onset breast/ ovarian cancer; again no controls appear to have been

analysed in this report [227]. The 186bp 5-6 intron inclusion seen here has not previously been reported and would be predicted to result in-frame insertion of 62 amino-acids. It seems unlikely to be of pathological significance at this level of expression, given the fact that it was seen with the wildtype as well as all eight mutant versions of this minigene.

.

Transfection of the thirteen different mutant exon 17-19 minigenes did not result in any alteration in the splicing pattern seen with the wildtype version of this minigene. All versions of this minigene resulted in two transcripts: the wildtype cDNA sequence and a very faint transcript containing insertion of the18-19 intron. Some variation in the ratio of the two transcripts produced by different minigenes was seen, but statistical analysis confirmed that none of the thirteen mutations tested altered the ratio significantly from that produced by the wildtype minigene. My results therefore suggest that the 13 BRCA1 exon 18 missense mutations tested here are unlikely to have any significant biological effects on splicing regulation. This is in contrast to the predictions of ESEfinder, which determined that all of the these mutations could affect ESEs, with 8 of the mutations predicted to destroy pre-existing ESEs and 6 predicted to create new motifs.

Transfection of one of the eight different mutant 3-6 minigenes did result in a different splicing pattern that that seen with the wildtype minigene. The minigene containing a G to T exchange at nucleotide 231 resulted in marked reduction in transcripts containing exon 6. Instead, a novel 530bp transcript was seen which sequencing confirmed to represent transcripts containing exons 5 and 7 with inclusion of the first 186bp of intron 5-6, but no exon 6, as well as a new 322 bp transcript which consisted of exons 5 and 7 with deletion of the last 22 bp of exon 5. Measurement of the concentration of the two predominant transcripts using *Phoretix*™ 1D Advance software confirmed confirmed that the splicing pattern produced by minigene 231 G>T clearly showed an excess of transcripts with deletion of exon 6 (greater that 55% of the total transcript for the 231 G>T mutation compared with approximately 17.5% for the wildtype minigene). Exon 5 of BRCA1

encodes the RING domain region of the BRCA1 protein. This is known to complex with BARD1 to produce E3 ubiquitin ligase activity which may influence the stability and activity of the BRCA1 multiprotein complexes [42]. A significant reduction in transcripts containing this part of the BRCA1 sequence could be postulated to reduce BRCA1 mediated DNA repair and therefore increase the risk of BRCA1 associated cancer.

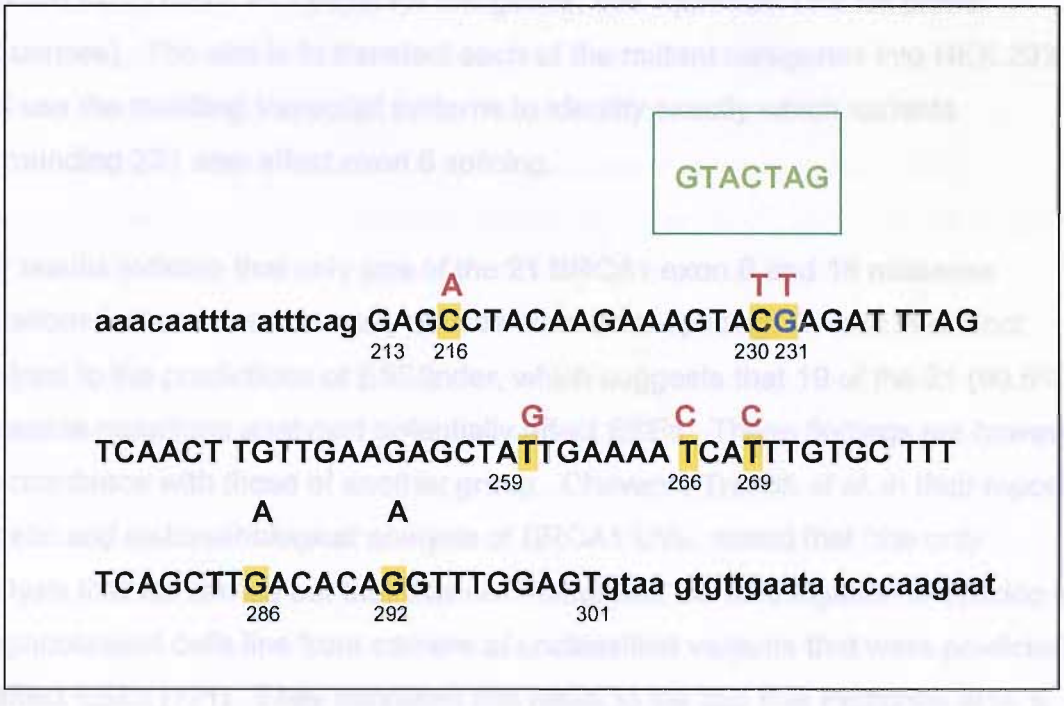


Figure 6.14: *Sequence of BRCA1 exon 6, with investigated missense mutations, indicating position of 231 G>T mutation and new ESE (in green), as predicted by ESE finder*

The G>T mutation at nucleotide 231 is sited at the 19th position of exon 6, and is therefore well away from the conventional splice donor and splice acceptor zones. ESEfinder predicts that this single base exchange creates a new SRp40 binding

ESE, starting at position 15, with a motif of GTACTAG. My functional studies suggest that this ESE promotes splicing of exon 5 to exon 7 with deletion of exon 6. In order to fully define *in vitro* the nature of the splicing regulatory motif influenced by the G>T mutation at 231, we have designed a series of additional oligonucleotides to create a series of exon 3-6 minigenes in which each nucleotide base between positions 224 and 238 is systematically and individually mutated into each of the alternative three nucleotide variants (eg. the A at 229 will become C, G, and T in 3 separate minigenes; see Appendix I for full primer sequences). The aim is to transfect each of the mutant minigenes into HEK 293 and use the resulting transcript patterns to identify exactly which variants surrounding 231 also affect exon 6 splicing.

Our results indicate that only one of the 21 BRCA1 exon 6 and 18 missense mutations tested *in vitro* actually causes alternative splicing. This is in distinct contrast to the predictions of ESEfinder, which suggests that 19 of the 21 (90.5%) missense mutations analysed potentially affect ESEs. These findings are however in accordance with those of another group. Chevenix-Trench *et al*, in their report of genetic and histopathological analysis of BRCA1 UVs, stated that "the only analysis that we carried out that was not fruitful was the investigation of splicing in lymphoblastoid cells line from carriers of unclassified variants that were predicted to affect ESEs [121]. They attributed this result to the fact that Pettigrew *et al*.'s work suggests that only a small proportion of the BRCA1 ESE motifs predicted by ESEfinder are likely to be of functional importance [165]. Pettigrew's group demonstrated that increasing the threshold values (to 3.0 for SF2/AFP and 3.0 for SC35, SRp40 and SRp55) reduces the total number of ESEfinder predicted BRCA1 ESEs to 464. By limiting analysis to the first and last 125 nucleotides of each exon (i.e. excluding analysis of the central portion of exon 11 and the 311bp exon 16 only) the number of putative ESEs is reduced to 211. Pettigrew hypothesised that BRCA1 ESEs that had been conserved through evolution were more likely to be functional than those found only in the human gene. Cross-species sequence analysis using the BRCA1 cDNA for the chimpanzee, rhesus

monkey, orang-utan gorilla, cow, mouse, opossum and dog indicated that 23 of the 211 predicted human BRCA1 ESE sequences (11%) were found in the same positions in all these species. Interestingly, limiting the number of putative ESEs in this way increased the co-localisation of ESEs with UVs predicted (by analysis of protein function) to be deleterious whilst reducing co-localisation with polymorphisms and UVs predicted to have neutral effects on protein function.

Our results are also in keeping with those reported by Wang *et al.* following their mathematical and experimental investigations of tumour-associated alternative RNA splicing in human cancer using the human EST database [160, 225]. A survey of 11014 mRNAs from the RefSeq database and some 3.47 million expressed sequence tags resulted in the identification of 26,258 alternative splice variants, only 3.2% of which were significantly associated with cancer [160]

An alternative method of predicting ESE sites is provided by the on-line programme RESCUE-ESE [164]. Candidate ESEs are identified by comparison with a list of 238 specific hexanucleotide sequences that occur with a significantly higher frequency in exons than in introns, and with a significantly higher frequency in exons with weak splice sites than strong splice sites, derived from analysis of 30,000 human genes. Use of this predictor method identifies 14 ESEs within exon 5, 11 within exon 6, 16 within exon 10 and 7 within exon 18. Of particular interest is the fact that only two of the ESEs found by RESCUE-ESE are identical in sequence and position to those found by ESEfinder, with a further 13 showing an overlap of at least 4 nucleotides with ESEfinder defined ESEs. The remaining 32 potential ESE motifs are entirely distinct from those recognised by ESEfinder. The RESCUE-ESE programme predicts that the G>T 231 BRCA1 mutation has no effect on ESE motifs.

Exon	Position	Motif	?match to ESEfinder		Exon	Position	Motif	?match to ESEfinder
5	11	CTGAAA	No		10	5	GGAGAT	No
5	12	TGAAAC	No		10	8	GATCAA	No
5	13	GAAACT	No		10	9	ATCAAG	No
5	15	AACTTC	No		10	11	CAAGAA	No
5	23	AACCAG	Yes*		10	12	AAGAAT	No
5	26	CAGAAG	Yes		10	36	AAGGAA	Yes*
5	27	AGAAGA	Yes*		10	37	AGGAAC	No
5	28	GAAGAA	Yes*		10	39	GAACCA	No
5	29	AAGAAA	No		10	40	AACCAG	No
5	30	AGAAAG	Yes		10	47	GATGAA	No
5	59	AAGAAT	No		10	48	ATGAAA	No
5	60	AGAATG	No		10	52	AATCAG	No
5	70	AACCAA	No		10	58	TTTGGA	No
5	73	CAAAAG	Yes*		10	59	TTGGAT	Yes*
					10	68	GCAAAA	No
6	6	TACAAG	Yes*		10	71	AAAAAG	No
6	7	ACAAGA	Yes*					
6	8	CAAGAA	Yes*		18	26	ACTGAA	Yes*
6	9	AAGAAA	No		18	27	CTGAAA	No
6	10	AGAAAG			18	51	GGAGGA	Yes*
6	29	CAACTT	No		18	52	GAGGAA	Yes*
6	36	TTGAAG	No		18	53	AGGAAA	Yes*
6	37	TGAAGA	No		18	54	GGAAAA	No
6	38	GAAGAG	No		18	55	GAAAAT	No
6	49	GAAAAT	No					
6	82	TTTGGA	No					

Table 6.11: *Position and motifs of ESEs within BRCA1 exons 5,6,10 and 18 predicted by RESCUE-ESE [164], with identification of ESEs also identified by ESEfinder*

(Yes= not perfect match but at least 4 nucleotides overlap with potential ESE motif as identified by ESEfinder algorithm).*

Similar significant disparity of results from the use of ESEfinder and RESCUE-ESE to predict ESEs in a variety of different genes is now common in the published literature [165, 173, 175]. A third on-line algorythm is now available to assess DNA sequences for splicing regulators. ExonScan (www.genes.mit.edu/exonscan/) identifies splice sites from a primary transcript using a maximum entropy model ESEs are determined using RESCUE-ESE and ESSs are identified using the FAS-hex3 set of hexanucleotides, a set of 103 hexamers identified by Wang *et al*, as inhibiting the use of local splice sites in an *in vivo* splicing reported system [229] [230]. Analysis of the BRCA1 5-7 minigene sequence by ExonScan also assessed the G>T 231 mutation to have no effect on splicing regulation. Thus, identification of potential splicing regulators by any *in silico* based program should be treated with caution. Recent reports of exonic sequences containing composite elements with overlapping enhancer and silencer properties (CERES; composite exonic regulatory element of splicing) further underline the complex nature of splicing regulatory motifs and the difficulties involved in approaching them from a purely theoretical or mathematical angle [231].

The finding of an alternative splicing pattern in association with only one of the 21 BRCA1 missense mutations tested in vivo may partially reflect the location of the mutations investigated. Exons 5,6,10 and 18 were selected as the central exons for minigenes mainly for technical reasons; they represented the four best available compromises between optimum exon and intron sizes, and frequency of previously reported missense mutations. Exon 6 encodes part of the RING domain which is involved in interaction with BARD1, whilst exon18 encode the proximal part of the BRCT domain (involved in transcription control) and the region of interaction with BRCA2. One would therefore expect errors in splicing regulation in these regions to have functional effects. It is feasible that more striking effects of missense mutations on splicing regulation might be seen using minigenes containing the regions of BRCA1 most highly conserved from an evolutionary point of view. Unfortunately, the best example of evolutionary conservation in this gene includes the proximal end of exon 11 [105, 107] With an exonic length of 3426 nucleotides

and flanking introns of 985 and 402 base pairs the development of a minigene encompassing BRCA1 exon 11 would be extremely difficult to achieve, and most likely would prove impossible without removal of a significant central portion of this exon. My results may also reflect the limitations of the minigene approach to investigation of splicing regulators in general. As in the majority of published examples of work with minigenes, our minigenes comprised a central exon containing the sequence variants of interest flanked by native introns with portions of the adjacent exons at the 3' and 5' end of the DNA sequences. Cases have been identified in which cell-specific regulatory motifs have been identified more that 1 kilobase away from the regulated splice site [216], with more than one intervening intron. Thus, the standard design of minigenes may simply not be sufficient to identify cis-regulatory factors influencing the splicing of exons of the large BRCA1 gene.

It should be emphasised that failure to affect splicing regulation does not imply that a BRCA1 missense mutation is not pathogenic. Studies such as that of Quaresima *et al.* have clearly identified potential pathogenicity of missense mutations where the base change results in a amino acid alteration that directly affects BRCA1 protein interaction with the tumour suppressor p53 [118]. Other groups have demonstrated examples of single nucleotide exchanges that clearly alter the BRCA1 protein surface resulting in destabilisation and unfolding of the protein at physiological temperatures [232]. It is however of particular interest that the one missense mutation tested that resulted in an abnormal splicing pattern was a synonomous change.

Synonomous, or silent, mutations have been largely overlooked by researchers when investigating the effect of BRCA1 DNA sequence variations. Even the recently reported genome wide screen for low-risk breast cancer susceptibility genes excluded synonomous mutations from analysis [10]. My results indicate that a synonomous mutation can affect splicing regulation of BRCA1 with potentially pathogenic effects. Similar findings have been described by Pagani *et al.* in their

studies of the cystic fibrosis transmembrane regulator (CFTR) [159, 231]. Six of 19 (32%) synonomous sequence changes introduced into exon 12 of this gene induced exon skipping [151], leading him to suggest that synonomous changes are not neutral in evolution. This observation is in keeping with other recent reports, including a comparative study of chimpanzee and human transcripts which has shown that a significant proportion of silent sites in protein coding regions are under purifying selection, and the observation that the average substitution rate of silent sites is higher in human pseudogenes than in functional genes [233], [234]. Such information suggests that the long-held assumption that the ratio of rate of nonsynonomous (Ka) and synonomous (Ks) substitutions can be used to identify areas of purifiying selection (Ka/Ks = <1), neutral evolution (Ka/Ks = 1) and positive selection (Ka/Ks>1) may need to be re-evaluated as this relies on the presumed neutrality of synonomous variations. Of the 477 SNPs present in BRCA1, 398 are synonymous [234.] Exclusion of these types of sequence variants from searches for cancer susceptibility genotypes may result in failure to identify sequence variants with significant effects on cancer predisposition.

The development of the minigenes used in this investigation provided a considerable technical challenge. Significant problems were encountered on cloning wildtype and mutant minigenes into the pCR® 3.1 vector. This was most probably due to the large size of the minigenes. Cloning of the 17-19 mutant minigenes, (the smallest constructs with a total length of 1288 bp) was achieved following subcloning into pGEM®-T Easy. This was tried without success for each of the other set of mutant minigenes and again failure was probably due to the large size of inserts. Adaptations of the cloning technique, including variations in conditions for transformation and ligation were all unsuccessful (data not shown). Successful cloning of the mutant versions of the exon 5-7 minigene required re-design of primers for the 2nd mutagenesis PCR such that only a 1096bp portion was re-amplified and required cloning into the wildtype minigene. Cloning of the 3-6 and 9-11 mutant minigenes has yet to be successfully achieved. All further attempts will involve similar re-design of the mutagenesis protocol to permit

amplification of a smaller piece of the minigene with the hope that this can be more readily ligated into the pCR® 3.1 vector.

We initially attempted to achieve site-specific mutagenesis through the technique of minigene template DNA amplification with 2 overlapping primers (one containing the mutation, one wildtype). This approach was uniformly unsuccessful, again presumably due to the large size of the minigene templates. We therefore moved on to the use of a megaprimer approach, as described in section 6.24. Small yields from each second PCR, despite optimisation of PCR parameters meant that all PCRs had to be performed in triplicate, with combination of PCR products prior to preparation for cloning. The number of PCR cycles was maintained as low as possible, (less that 30 in total) to reduce the risk of non-specific mutations developing. New, undesired mutations were found in 2 of the mutant minigenes: these were discarded and new mutant minigenes were prepared in their place. Commercial mutagenesis kits such as QuikChange® Site-Directed Mutagenesis Kits (Stratagene, California, USA) potentially offer a more rapid method of creating mutant minigenes and would be trialled in future work on the remaining 3-6 and 9-11 minigenes.

The transfection experiments described were performed using the renal cell line HEK 293. This cell line was chosen for these initial experiments as it is easy to transfect and sufficient yields of RNA for cDNA analysis can be harvested after 24-48 hours. It must be noted that there are a number of reports in the literature documenting the variability of splicing regulation in different cell lines. By analysing the splicing patterns of genomically aligned expressed sequence tags derived from libraries of cDNAs from different human tissues, Yeo *et al.* have demonstrated that levels of alternative splicing events are significantly higher in brain, liver and testes tissue (≥40% of genes exhibiting at least one AS event) than the ovary, breast, muscle, uterus or pancreas (less than 25% of genes) [221]. These observations applied to frequencies of complete exon deletion and alternative 3' and 5' splice site usage.

Ideally, all our experiments should be repeated using breast cell lines. It is interesting to note that the position of mutations within the BRCA2 gene seem to determine the relative risk of carriers developing ovarian rather than breast cancer. This suggests that breast and ovarian tissues differ in their requirements for BRCA gene function. It would be interesting to see whether transfection of ovarian cell lines result in the same pattern of splicing as seen in breast cells. Evidence that the frequency of alternative splicing is significantly lower in breast and ovarian tissue than other sites could perhaps suggest that these tissues are more susceptible to detrimental results from this phenomenon. It has been observed that distribution of the two different (and mutually exclusive) 5' BRCA1 promotors (which account for the two different exon 1 sequences) may be tissue specific, with mammary tissue expressing only exon 1A [149]. This could partially explain the variation in other transcript levels between tissue types.

Work on the splicing of the neurofibromatosis gene NF1 has raised the possibility that RNA extracted from blood stored at non-physiological temperatures can increase abundance of some alternative transcripts [236, 237]. The results I obtained from the two separate transfections of the exon 17-19 minigene series (table 6.10) clearly show the degree of variation in ratio of one transcript to another that can occur on repeat transfection of the same minigenes on different occasions, even under identical active experimental conditions. The differences seen here are most likely to have arisen as a result of a variation in storage time of the RNA prior to cDNA production and highlight the instability of RNA and difficulties inherent in these types of analyses. The potential for dimerisation of dominant transcripts during amplification of cDNA, resulting in apparent large alternative transcripts, should also be noted as this can complicate gel electrophoresis assessment of transcription products and emphasises the need to sequence all visible cDNA bands.

The rapid progression of new technologies means that future years should provide vital further insights into the significance of alternative splicing of the BRCA1 gene.

Existing EST libraries provide a potential source of valuable information regarding BRCA1 alternative splicing events. It should however be noted that a significant limitation of AS analyses using transcript sequence data is that EST coverage is often biased towards the 3' or 5' end of each transcript and that there are generally insufficient sequenced transcripts to calculate the frequency with which any one particular splicing event occurs in any given cell or tissue [238]. Microarray techniques offer a potential means for large-scale profiling of BRCA1 alternative splicing events in breast and other tissues, and investigation of specific splice factor usage [221, 238].

6.5 Conclusions

My work has shown that the use of minigene constructs is a feasible method of systematically investigating the effect of missense mutations on splicing regulation of BRCA1. The results from transfection of mutant BRCA1 minigene constructs into the renal cell line HEK 293 confirms previous reports that indicate *in silico* ESE predictor programmes do not accurately identify *in vivo* splicing regulatory motifs. Only 1 of the 21 (4.8%) BRCA1 exon 6 and exon 18 missense mutations tested here, a G>T mutation at 231, resulted in a different splicing pattern compared with that seen on transfection of the wildtype version of this minigene. This compares with ESEfinder predictions that 19 (90.5%) of these mutations potentially affect ESEs. Work is currently underway to completely define the splicing regulatory motif surrounding 231. To gain further insight into the effect of these missense mutations on splicing regulation, these experiments need to be repeated in a variety of cell lines, including breast and ovarian cell lines.

7.0. Investigation of Influence of MDM2 SNP309 on Tumour Development in BRCA1 mutation carriers

7.1 Introduction

MDM2 encodes a protein which directly binds to and inhibits the tumour suppressor p53 [189] but also has p53 independent negative effects on DNA double strand break repair [190, 191]. MDM2 gene amplification is found in a number of tumour types, including sporadic breast cancers, [192] and overexpression of MDM2 is associated with accelerated tumour development and failure to respond to treatment [193].

Bond *et al.* [194] recently described a common single nucleotide polymorphism (SNP) within the promoter region of MDM2 (a T to G exchange at nucleotide 309) which increases the affinity of the transcriptional activator Sp1 and results in higher levels of MDM2 protein with subsequent attenuation of the p53 tumour suppression pathway. Their study of 88 individuals with inherited p53 mutations demonstrated that carriers of this SNP in either its' homozygous or heterozygous form developed cancer on average 7 years before their wildtype counterparts. Similar results have subsequently been reported in an independent study of 61 patients with germline p53 mutations [239]. Bond *et al.* additionally found that individuals homozygous for SNP309 but with no known hereditary cancer predisposition showed a significantly earlier age of onset of sporadic soft tissue sarcoma than those with the T/T genotype [194].

The age of onset of first malignancy in patients with BRCA1 mutations of established pathogenic potential varies considerably as does (albeit to a lesser extent) the lifetime risk of developing breast or ovarian cancer (see introduction). This may partially be explained by differential exposure to external risk factors such as the contraceptive pill, and variations in individual hormonal profiles as influenced by age of menarche, parity, breast feeding and body mass index [240]. However, the clinical course of malignant disease in a carrier of a pathogenic

BRCA1 mutation may also be influenced by additional genes, which interact with the BRCA1 cellular pathway and modify the response to DNA damage. Identification of putative genetic disease modifiers would facilitate clinical risk management of patients with diagnosed BRCA1 mutations.

A higher frequency of p53 mutations in BRCA1 associated breast tumours than sporadic breast cancer has been observed in a number of studies [56, 60, 67]. It is feasible that the MDM2 SNP309 influences tumour development in BRCA1 mutation carriers via modulation of the p53 tumour suppressor pathway. I have therefore investigated the effect of the MDM2 SNP309 on clinical outcome in a cohort of patients with germline mutations of BRCA1 with established pathogenic potential.

7.2 Methods

7.2.1 Subjects

Potential study participants with previously diagnosed truncating mutations of BRCA1 were identified from the database of the Wessex Clinical Genetics Service (WCGS), Southampton. Genomic DNA was obtained for 116 BRCA1 mutation carriers (age range 29–57 years) and the clinical histories of all subjects were reviewed to ascertain age at first and subsequent malignancies. Fifty-nine gene carriers had developed breast cancer and a further 16 had ovarian cancer. Ten patients had undergone prophylactic bilateral oophorectomy (five after developing breast cancer) and 3 patients had undergone bilateral risk reducing mastectomy. One hundred and two anonymous genomic DNA samples, referred by WCGS for genetic screening of non-neoplastic conditions, were obtained to provide a control group.

7.2.2 Genotyping

Genotyping was performed using Pyrosequencing™ technology. Amplicons were generated in a 50µl reaction volume with 15pmol each of MDM2F (5' GTCTCCGCGGGAGTTCA 3') and MDM2R (5' Biotin – GACTACGCGCAGCGT TCAC 3'), 0.2mM dNTPs, 1.5mM $MgCl_2$, 1X Buffer II, 1U AmpliTaq Gold using 10ng genomic DNA. PCR conditions were 94°C for 7 minutes; 50 cycles with denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds and elongation at 72°C for 30 seconds; 1 cycle at 72°C for 7 minutes; and a final hold at 15°C. Thermocycling was performed using a PTC-0225 DNA Engine Tetrad.

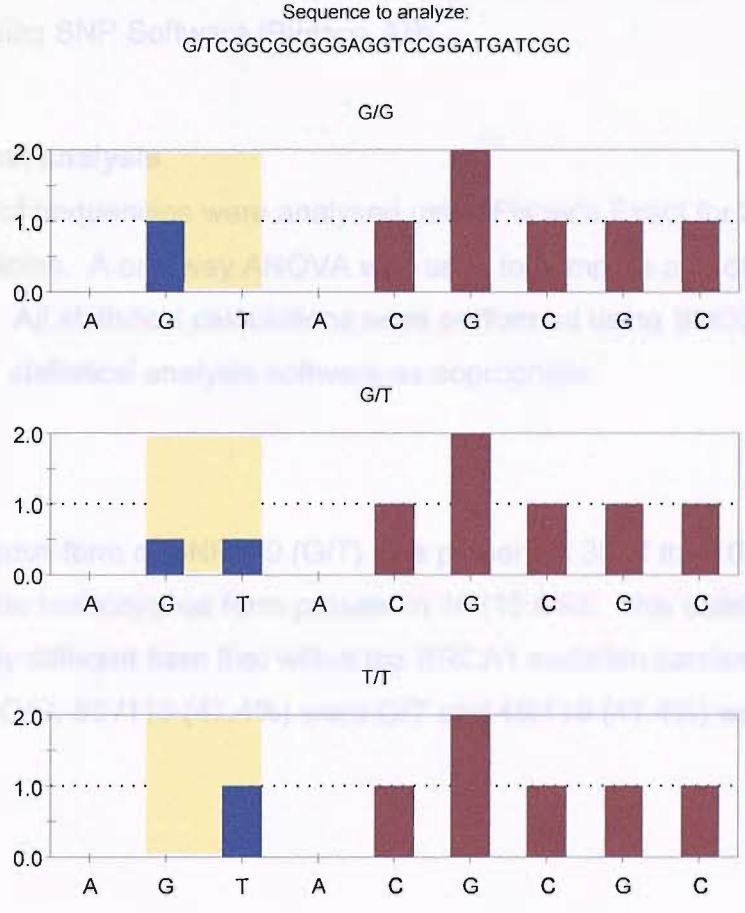


Figure 7.1*: Histograms demonstrating selected theoretical outcomes for pyrosequencing™ simplex entry MDM2*

Single-stranded biotinylated PCR products were prepared for Pyrosequencing as described in section 4.2. The beads were released into a PSQ 96 well plate containing 45µl annealing buffer (20 mM Tris-Acetate, 2mM $MgAc_2$ pH 7.6), 0.3µM MDM2 sequencing primer (5'GGGCTGCGGGGCCGCT 3'). The samples were heated to 80°C for 2 minutes and then allowed to cool to room temperature.

Pyrosequencing reactions were performed according to the manufacturer's instructions using the PSQ 96 SNP Reagent Kit (Biotage AB) which contained the enzyme and substrate mixture and nucleotides. Assays were performed using the nucleotide dispensation order AGTACGCGC. The sample genotype was determined using SNP Software (Biotage AB).

7.2.3 Statistical analysis

Comparisons of frequencies were analysed using Fisher's Exact for 3 x 2 contingency tables. A one way ANOVA was used to compare age of onset of malignancies. All statistical calculations were performed using StatXact™ version 5 or Minitab™ statistical analysis software as appropriate.

7.3 Results

The heterozygous form of SNP309 (G/T) was present in 38 of the 102 controls (37.3%) with the homozygous form present in 16 (15.6%). This distribution was not significantly different from that within the BRCA1 mutation carriers: 13 /116 (11.2%) were G/G, 55 /116 (47.4%) were G/T and 48/116 (41.4%) were T/T, (p=0.286).

Genotype (female subjects only)	**Total** n=96	**G/G** n=12	**G/T** n=43	**T/T** n=41	
Number of subjects with ≥ 1 malignancy (%)	71 (74.0%)	10 (83.3%)	30 (69.8%)	31 (75.6%)	P=0.61
No. of subjects with breast cancer (%)	59 (61.5%)	5 (41.7%)	27 (62.8%)	27 (65.8%)	P=0.34
Mean age of onset of 1st breast cancer/ years (range)	41.6 (26-57)	41.2 (29-49)	38.6 (29-52)	39.0 (26-57)	P=0.80
No. of subjects with ovarian cancer (%)	14 (14.6%)	4 (33.3%)	4 (9.3%)	6 (14.6%)	P=0.32
Mean age of onset of 1st ovarian cancer/ years (range)	58.0 (44-74)	62.3 (51-74)	60.3 (51-68)	53.7 (44-66)	P=0.32
Other malignancies	3	1 plasmacytoma	1 cervical cancer 1 uterine cancer		
No. of patients undergoing prophylactic bilateral oophorectomy	10	0	5	5	
No. of patients undergoing bilateral risk reducing mastectomy	3	1	1	1	

Table 7.1: *Incidence and age of onset of all malignancies, breast and ovarian cancer in carriers of pathogenic BRCA1 mutations genotyped for MDM2 SNP309 (p values refer to comparisons of frequencies between the G/G, G/T and T/T subgroups)*

Seventy-one of the 116 BRCA1 mutation carriers had been diagnosed with at least one malignancy. The incidence of malignancy was highest in G/G genotype carriers with these patients developing tumours in 10/13 cases (76.9%). G/T carriers developed tumours in 30/55 cases (54.5%) and wildtype (T/T) individuals developed tumours in 31/48 cases (64.6%), but statistical analysis revealed no significant association between incidence of malignancy and genotype (p=0.28).

All documented malignancies occurred in female BRCA1 mutation carriers and no significant association was found between frequency of malignancy and genotype when limiting analysis to female subjects only (n=96, incidence of malignancy =83.3 % [10/12] in G/G carriers, 69.8 % [30/43] in G/T carriers and 75.6% [31/41] in wildtype patients, p=0.61).

Breast cancer had developed in 59 of the females with BRCA1 mutations: 5/12 (41.7%) of those with the G/G genotype, 27/43 (62.8%) of those with the G/T genotype and 27/ 41 (65.8%) of those with the T/T genotype. Mean age of diagnosis of first breast cancer did not vary significantly with genotype, being 41.2 years (range 29-49) in the SNP309 G/G carriers, 38.6 years (range 29-52) in SNP G/T carriers, and 39.0 years (range 26-57) in subjects with the wildtype genotype (p=0.80).

Ovarian cancers occurred in 14 females: 4 with G/G genotype (33.3%), 4 with G/T (9.3%) and 6 with T/T (14.6%, p=0.32). Mean age of onset was similar in each genotype group being 62.3 years in the G/G group (range 51-74), 60.3 years in the G/T group (range 51-68 years) and 53.7 years in the wildtype group (range 44-66), p=0.32).

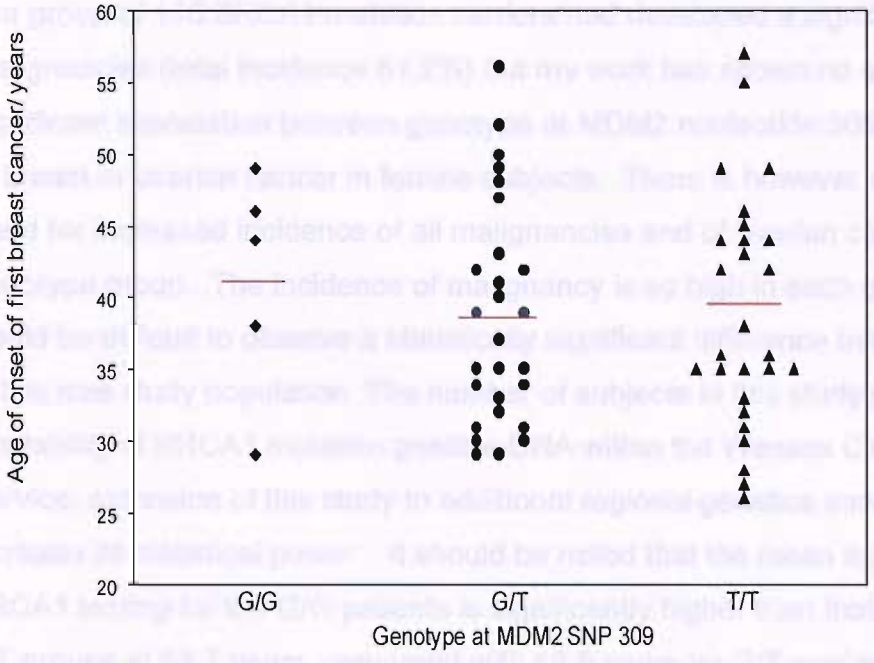


Figure 7.2: *Scattergram of age of onset of first breast cancer in BRCA1 mutation carriers categorised according to genotype at MDM2 SNP 309. Mean age of diease onset in each group is indicated by red bar.*

7.4 Discussion

Analysis of the control group of 102 confirmed that the MDM2 SNP309 genetic variation is relatively common within the general population with a frequency of 37.3% for the heterozygous state (G/T) and 15.6% for the homozygous state (G/G). This figure for the homozygous frequency is slightly higher than that reported by Bond *et al.* [194] in their analysis of 50 healthy volunteers (40% G/T, 12% G/G) but is similar to that found by Campbell *et al.* in another control group from the UK [241]. My results suggest that distribution of this polymorphism within carriers of pathogenic BRCA1 mutations (11.2%G/G, 55 47.4%G/T and 41.4% T/T) is not significantly different from that within the general population. This finding is

supported by the very similar results of Wasielewski *et al.* in their study of 87 BRCA1 mutation carrying families (13% G/G, 45% G/T, 42% T/T) [242].

Our group of 116 BRCA1 mutation carriers had developed a significant number of malignancies (total incidence 61.2%) but my work has shown no evidence of a significant association between genotype at MDM2 nucleotide 309 and incidence of breast or ovarian cancer in female subjects. There is however a non-significant trend for increased incidence of all malignancies and of ovarian cancer in the G/G genotype group. The incidence of malignancy is so high in each group that it would be difficult to observe a statistically significant difference between subgroups in this size study population. The number of subjects in this study was limited by availability of BRCA1 mutation positive DNA within the Wessex Clinical Genetics Service; extension of this study to additional regional genetics services would increase its statistical power. It should be noted that the mean age at time of BRCA1 testing for the G/G patients is significantly higher than that of the G/T and T/T groups at 58.7 years, compared with 49.9 years for G/T carriers and 48.9 years for wildtype patients ($p=0.034$). The G/G group therefore has had longer to develop malignancies than the other subjects. In addition, 5 patients in each of the G/T and T/T groups had undergone prophylactic oophorectomy whilst no patients with the G/G genotype had had this procedure. This may have affected the incidence of ovarian cancers in the G/T and T/T groups and could also have reduced the frequency of breast cancers in the 5 subjects (4 G/T, 1 T/T) who had not already developed this disease. The trend for increased incidence of ovarian cancers in the G/G group does perhaps warrant further study in a larger cohort of patients who have not undergone prophylactic surgery.

The findings of this study are, however, in agreement with the results of Wasielewski *et* al.'s report of MDM309 genotypes in 271 early-onset breast cancer cases [242]. In 87 BRCA1 mutation carrying cases, they observed no difference in MDM2 SNP309 GG frequency between early-onset (defined as first breast cancer diagnosed at or before age 51) and late-onset (breast cancer diagnosed after age

51). A study by Campbell *et al.* also found no significant association between the G/G genotype and breast cancer in a group of 218 early onset breast cancer patients (age at onset of less than 40 years) or in a group of 248 patients with familial breast cancer (defined as two or more cases of breast cancer reported in first or second-degree relatives or bilateral breast cancer) [241]. The BRCA1 mutation status of Campbell's subjects was not reported but obviously, a significant proportion is anticipated to carry BRCA1 mutations.

The fact that both I and Wasielskewski *et al.* found no association between SNP309 genotype and age of onset of breast cancer in female subjects with pathogenic BRCA1 mutations is of biological interest. Pre-clinical studies have shown clear evidence of an interaction between the p53 tumour suppressor and BRCA1 cellular pathways. It could therefore be expected that partial attenuation of the p53 pathway associated with the MDM2 SNP309, as described by Bond *et al.* would have exacerbated the effect of the BRCA1 mutation in promoting early onset breast cancer [194]. These results may indicate that the increased MDM2 levels associated with the SNP309 are insufficient to attenuate response in the BRCA1 mutant breast cells, whereas in the Li Fraumeni cohort studied by Bond *et al,* there is already loss of one p53 allele. Alternatively, the transcription factors influenced by SNP309 could be cell type specific so that this polymorphism does not affect MDM2 expression in breast and ovarian tissue. MDM2 amplification has in fact been reported to be rare in ovarian cancers regardless of p53 mutation status, suggesting that this gene may not be significantly involved in the development of these tumours [243]. Overexpression of MDM2 has however been found in up to 73% of breast cancers [192] and breast cancers comprised 17 out of 66 of the tumours reported by Bond's cohort of Li Fraumeni individuals [194].

P53 mutations are the most commonly found genetic alteration in human tumours and a number of studies have reported a higher frequency of p53 mutations in BRCA1 associated breast tumours and ovarian cancers than sporadic breast/ ovarian cancers [56, 67, 183, 244]. A study of 248 breast cancer patients has

revealed a statistically significant interaction between SNP309 status and tumour p53 expression for breast cancer survival [240]. In carriers of the wildtype genotype (T/T), presence of p53 mutations and aberrant p53 protein expression in breast tumours were associated with poor survival. The association between p53 tumour status and breast cancer survival was not found in carriers of the variant (G) allele. The p53 status of breast cancers occurring in our subjects is currently unknown and it is possible that an association between accelerated tumour development and MDM2 SNP309 would be found if analyses were restricted to BRCA1 mutation carriers with evidence of associated p53 protein abnormalities in the breast tumour. Interestingly, Menin et al. have recently reported that the SNP309 G allele accelerates age of colorectal cancer diagnosis in patients with a p53 wild-type carcinoma [245].

Bond *et al,* [194] found clear evidence of an association between the SNP309 and accelerated tumour development in 88 individuals carrying germline mutations of p53 and in 105 soft tissue sarcoma patients with no known familial tendency. Although there have been a few subsequent reports describing a similar association between this polymorphism and breast tumour development, a number of subsequent studies have found no relationship between MDM2 SNP309 genotype and incidence of sporadic or familial (non-BRCA) breast cancer [240, 241, 246, 247]. Alhopuro *et al.* observed no statistically significant association between SNP309 genotype and patient age at disease onset in cohorts of early uterine leiomyosarcoma (68 subjects), colorectal cancer (1042 subjects) or squamous cell head and neck cancer (162 subjects) [244] although there have been reports of a positive association with incidence of lung, oesophageal and gastric cancer [249-251]. These results together with my own findings suggest that the influence of the MDM2 SNP309 on tumour development is specific to certain patient subgroups.

Following their original publication, Bond *et al.* have commented on the disparity in the difference in mean age of disease onset between the G/G and T/T groups of

sporadic sarcomas (12 years) and the difference in median age of disease onset (21 years). These figures arise as a result of a few G/G individuals developing cancer at a later age comparable with the wildtype individuals and have been attributed to the existence of additional genetic modifiers [252]. More recently, this group has published data indicating that the G/G genotype (compared with the T/T genotype) accelerates the onset of breast tumours expressing high levels of oestrogen receptors, but not oestrogen negative breast tumours [246]. This biological effect is plausible as the oestrogen receptor binds to the region of the MDM2 promoter containing SNP309 to activate MDM2 transcription [253]. BRCA1 tumours tend to show reduced expression of oestrogen receptors and so could be protected from an oestrogen mediated MDM2 effect. Wasielseski et al. however found no evidence for an influence of oestrogen signalling on cancer acceleration in their report of a positive association between SNP 309 G/G genotype and familial breast cancer [242].

7.5 Conclusions

I have found no evidence that the presence of the MDM2 SNP 309 accelerates tumour development in carriers of known pathogenic germline mutations of BRCA1 to date [254]. These findings are in agreement with two other recent publications [241, 242] and this polymorphism therefore seems unlikely to act as a significant disease modifier in this population. However, the p53 mutation status of the breast tumours developed by the patients is currently unknown. It is still possible that an association between the MDM2 SNP 309 and accelerated cancer development does exist in BRCA1 mutation carriers with p53 overexpressing breast tumours but that this effect has been masked by the contribution of non-p53 expressing breast tumours to the results of this study.

In order to investigate further the hypothesis that presence of the MDM2 SNP309 effects the rate of breast tumour development in BRCA1 mutation carriers, I have sought and obtained approval from the Southampton LREC to retrieve breast tumour pathological specimens for all of the above subjects who have previously

had a histologically proven breast carcinoma resected, with the aim of submitting these specimens to a thorough histological review. The p53, p21, oestrogen receptor, progesterone receptor and erB2 receptor status of each tumour specimen will be established using immunohistochemical and tissue microarray techniques and the morphological characteristics will also be reviewed. This profile will then be correlated with the MDM2 SNP309 haplotype and clinical outcome. Additionally, this data will be analysed in conjunction with the underlying BRCA1 mutation to determine whether the site or type of mutation affects the tumour phenotype. In order to increase the statistical power, this project will be extended to include patients from other regional genetics services.

This work has confirmed that pyrosequencing is a rapid, reliable and cost effective technique for genotyping large numbers of subjects at the locus of a known SNP. The technique described above can easily be adapted to investigate other SNPs that may influence the development of malignancy, as proved by work I have recently performed on the DAPK-1 SNP1348 (data not shown here but see reference [255]).

Chapter 8: Summary and Plans for Future Work

It is currently estimated that 5% of all breast cancers are due to an inherited mutation of the tumour suppressor genes BRCA1 or BRCA2 [18]. Approximately 17% of genetically screened families with a history of breast and/ or ovarian cancer typical of autosomal dominant inheritance pattern are found to carry nonsense mutations of BRCA1/2 which would be expected to produce a truncated protein with pathological implications [256]. However, the Breast Cancer Information Core database (http://research.nhgri.nih.gov/bic/, contains over 3000 reports of BRCA1 missense mutations which would not be predicted to directly effect protein function. The phenotypical consequences of such DNA sequence variations are not yet fully understood. Identification of such mutations is complicated by the very large size of both the coding and intronic regions of the BRCA1 gene. Screening of the entire coding sequence has provided a technical and financial challenge and has only been achieved on a routine basis in the UK within the last 2 years [257,258].

There have now been a number of published examples of missense mutations resulting in aberrant gene transcripts through interference with splicing regulators such as ESEs and ISEs [173-175]. Investigation of missense mutations in genes such as the cystic fibrosis transmembrane receptor suggests that such genetic variants can have pathological effects via reduction of the normal amount of "healthy" transcript, but only a very small number of BRCA1 missense mutations have been analysed in this respect [159, 231]. The aim of this project therefore was to investigate further whether missense mutations in the BRCA1 gene can affect splicing by disrupting splicing regulatory motifs. This has been achieved using several different experimental techniques.

In recent years there have been two published reports of putative monoallelic expression of BRCA1 in association with missense mutations [171,202]. I therefore initially used the technique of Pyrosequencing™, a rapid real time sequencing method to look for examples of complete allele dropout, in subjects

with known truncating mutations, known BRCA1 missense mutations and patients with strong family histories of breast/ ovarian cancer but no known DNA sequence variation. Common SNPs in exons 13 and 16 were used as allelic markers. This technique identified subtle changes in the cDNA allelic ratios when compared with DNA allelic ratios of the patients with known truncating mutations but similar findings were also found in controls. No significant variations were seen in missense patients or those with no known mutations and no examples of complete allelic dropout were seen in any of the subjects analysed. Further investigations indicated that allelic ratios were affected by RNA quality, in particular time between venepuncture and RNA extraction. This technique was therefore judged to be too unreliable to develop further as a screening tool.

This work did reveal that, although Sharp *et al.* had correctly identified loss of heterozygosity of two exon 11 SNPs in cDNA compared with genomic DNA in two separate individuals, his report of additional loss of heterozygosity of an exon 16 SNP in one subject was erroneous [171]. The conclusion of monoallelic BRCA1 expression in this situation was therefore inappropriate. It is notable that both of the two published reports of loss of heterozygosity refer to SNPs at sites within the region of BRCA1 deleted in three common alternative transcripts (Δ (9, 10) with deletion of exons 9 and 10, Δ (11q) with partial deletion of exon 11, and Δ (9, 10, 11q). This observation therefore may be due to selective loss of alleles within alternative transcripts. This hypothesis should be tested by using Pyrosequencing™ to genotype BRCA1 cDNA sequences at a series of SNPs 3' and 5' to the BRCA1 exon 10-exon 11 boundary in individuals proven to be heterozygous at these loci.

I went on to search for alternative transcripts of BRCA1 by using multiplex PCRs to amplify four overlapping sections of the BRCA1 cDNA sequence in a series of subjects with known BRCA1 truncating or missense mutations. Additional subjects with BRCA1 missense mutations were subsequently identified using the technique of conformational sensitive capillary electrophoresis recently developed by NGRL

(Wessex) which permits rapid screening of the entire coding sequence of the BRCA1 gene. BRCA1 transcripts from 6 patients with BRCA1 truncating mutations, 12 with known missense mutations, 4 with splice site mutations and 21 individuals with strong breast /ovarian cancer family histories but negative routine BRCA1/2 screens were analysed. This confirmed that the c4185A>G splice site mutation is associated with deletion of exon 12 and effectively demonstrated the common alternative transcripts BRCA1Δ9, 10, BRCA1-Δ22deltantex5 and BRCA1 13A but failed to identify any underlying BRCA1 mutations not picked up by other screening methods. The mutiplex RT-PCR technique is therefore potentially most useful as a method of investigating patients with known missense or splice site mutations.

This experimental process was deliberately designed as a simple, fast and cheap procedure suitable for screening purposes, involving only four PCR reactions for each cDNA sample. However, I was unable to successfully amplify either the 2388bp BRCA1 9-11 fragment or the 3798bp 9-13 fragment and it proved technically impossible to design a single alternative pair of primers that would theoretically amplify the entire exon 11 and flanking exon boundaries in a single PCR. The current overlapping amplicon series therefore lacks sensitivity in the exon 11 region. To improve upon this situation I would like to develop a series of much smaller overlapping amplicons designed to amplify 1-2 exon long sections, of BRCA1 with the aim of detecting any exonic deletions which may be being missed here due to failure of a primer to bind to the aberrant transcript. The use of nested PCR techniques may be required to provide sufficient yields of fragments including the large exon 11 and a number of different exon 11 primers should be included, sited at 3' and 5' positions of the cryptic splice sites used in the known exon 11 alternative transcripts.

At this point in my investigation it became apparent that, to perform a systematic investigation on of the effect of BRCA1 missense mutations on splicing regulation, it was not feasible to rely on the availability of RNA from local mutation carriers. Instead, I decided to artificially recreate previously reported BRCA1 missense

mutations and analyse their effect on transcription using a minigene technique. I successfully developed four separate minigenes centred on exons 5,6,10 and 18. Transfection of the wildtype versions of these minigenes into HEK 293 cells resulted in several transcripts for each minigene. The alternative transcripts included two examples of cryptic splicing: the previously described deletion of the 3' 22bp of exon 5 and a novel finding of insertion of the 5' 186bp of intron 6-7 [203].

Missense mutations previously reported to the BIC database were inserted into the wildtype minigenes using a megaprimer approach and the thirteen exon 18 mutants and eight exon 6 mutants were successfully transfected into HEK 293 cells. Only one of these mutations, a G>T exchange at nucleotide in exon 6, resulted in an alteration in transcript pattern compared to that seen with the wildtype version of this minigene, with an increase in transcripts exhibiting deletion of exon 6. The online algorithm, ESEfinder [163], predicts that this mutation creates a new ESE and work is underway, using a further series minigenes carrying systematic single base changes between bases 224 and 238, to completely define the splicing regulatory motif involved here.

The fact that only one of the twenty-one (4.8%) mutations tested so far has produced an aberrant transcript pattern is interesting because software such as ESEfinder [163] predicts that nineteen (90.5%) of these would be expected to affect splicing. My results are further evidence, to add to the published work of Auclair *et al.*, Chevenix-Trench *et al.*, Lastella *et al.* and McVety *et al.*, that suggests that ESEfinder significantly overestimates the number of exonic SR protein binding motifs with functionally significant ESE activity [121,173-175]. My work also indicates that investigation of potential splicing regulatory sites as predicted by algorythms such as RESCUE-ESE will miss regulatory sites with potentially important clinical consequences. As Lastella states, "functional in vivo splicing assays should be mandatory for proper genetic analysis" [175].

Direct cloning of the mutant minigenes into the desired pCR®3.1 vector was extremely challenging from a technical point of view, presumably due to the large size of the minigenes. Cloning the exon 18 mutants was achieved by initially subcloning the mutants into pGEM®-T Easy prior to restriction site cloning into pCR®3.1. Cloning of the exon 6 mutants required re-design of the mutation primers such that only a 1096bp portion of the minigene was submitted into mutagenesis and then ligated directly into pCR®3.1 vector containing the remaining portion of the minigene. All attempts to clone the exon 10 and exon 5 mutations either directly into pCR®3.1, or indirectly via subcloning into pGEM®-T Easy have been unsuccessful. I am therefore now planning to design alternative mutation primers which will amplify selective portions of these minigenes and hopefully permit direct cloning into pCR®3.1 containing the remaining portion of the minigene. I would also consider trialling a commercial site-directed mutagenesis system as an alternative to using the megaprimer approach for mutagenesis. This work will be developed further by repeating transfections in a number of different cell culture systems, including breast, ovarian and breast and ovarian cancer cell lines. We also intend to repeat these experiments using Puromycin to inhibit NMD to see whether this affects the levels of transcripts seen.

The development of these minigenes and subsequent mutagenesis was challenging from a technical point of view and very consuming of both time and resources. Such techniques do however offer the opportunity to study BRCA1 (and other gene) missense mutations in a systematic fashion. This report is, to the best of our knowledge, the first time such a systematic enquiry has been performed on BRCA1, or any of the other cancer predisposition genes. It is intended that this work will be extended in the future to study the remainder of the BRCA1 gene in a similar fashion, to provide a significant database of information of the effects of BRCA1 missense mutations on the cDNA level. This information could then be used in conjunction with existing information on the outcome of other types of functional studies to further inform physicians advising patients with BRCA1 missense mutations of their potential risk of breast cancer. Ideally, a single, fully

referenced database should be developed to hold all information on BRCA1 missense mutations in one location for ease of use. The difficulty of investigating exon 11 missense mutations should not however be underestimated, owing to the very large size of this exon. Realistically, minigenes centred on exon 11 would probably have to be designed as two distinct entities, one including the first 200 nucleotides of this exon, with the other incorporating the last 200 bp of the exon.

To further extend this work, one should go on to examine protein expression of BRCA1 products using either western blotting or micoarray technology. Lixia *et al.* have recently reported that western blot analysis can successfully be performed on lysates of the breast cancer cell line ZR-75-30, with variable expression of BRCA1 splice isoforms being demonstrated [259]. One recent publication reports the loss of nuclear BRCA1 protein staining in normal tissue cells derived from BRCA1 and BRCA2 mutation carriers [260]. This finding, together with the progress made in recent years identifying the basal cell phenotype, suggests that a future method of identifying pathogenic underlying BRCA mutations in patients already affected by malignancy may involve the use of a detailed histopathological review to select patients for genetic investigation.

Finally, I also used the Pyrosequencing™ technique to investigate the association between genotype at the MDM2 promoter SNP 309 and age of onset of breast cancer in carriers of protein truncating BRCA1 mutations. This work was performed in the wake of the pivotal report by Bond *et al.* which identified an association between this allele and accelerated tumour development in Li Fraumeni patients [194]. Despite the known preponderance of BRCA1 tumours to show abnormal p53 expression, I found no association between this SNP and accelerated tumour development in 116 BRCA1 mutation carriers. This finding is in accordance with the subsequently published studies of Campbell *et al.* and Wasielewski *et al.* [241, 242]. To investigate this further, I have obtained ethical permission to submit breast tumour samples from my subjects for a full histopathological review, to include analysis of p53 status. There is also considerable overlap between my

subjects and participants of the EMBRACE, (http://www.srl.cam.ac.uk/genepi/embrace/embrace), and POSH, (http://www.som.soton.ac.uk/research/cancersciences/Research_Groups/Cancer_genetics/POSH), studies, which include microarray analysis of p53 mutation status and it is hoped that in the future, data from this study will become available for analysis in the context of this particular issue.

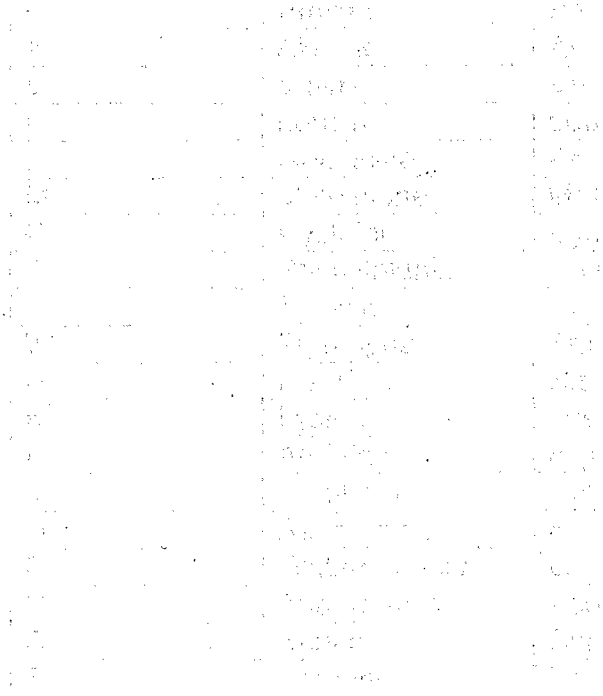
The Pyrosequencing™ technique proved to a very efficient and inexpensive method of assessing genotype at a known SNP in large number of subjects and would be a suitable method of screening subjects for genotype at any known SNP that was thought to contribute to cancer risk. Assays could for example be easily developed to identify genotype at any of the loci recently identified as increasing the risk of breast cancer [10]. This work has already led to a successful collaboration with Professor Hupps' team, University of Edinburgh, who are working on the effect of a SNP in the death domain in the DAPK gene and its' possible association with oesophageal carcinoma [255].

Since 1993, when BRCA1 was first identified, a wealth of research has been directed at identifying and managing patients with an elevated risk of hereditary breast and ovarian due to clearly pathogenic mutations in this gene. For most patients with BRCA1 missense mutations, undergoing genetic screening currently provides more questions than answers. Despite increasing evidence in the literature that missense mutations can disrupt regulatory motifs, resulting in alternative transcripts with potentially pathogenic effects, my research suggests that only a small percentage of the reported BRCA1 missense mutations will actually result in aberrant splicing. Identifying which BRCA1 missense mutations do affect splicing remains a huge technical challenge.

Appendix A: Reference cDNA sequence of BRCA1 (ascension no. U14680)

```
   1 agctcgctga gacttcctgg accccgcacc aggctgtggg gtttctcaga taactgggcc
  61 cctgcgctca ggaggccttc accctctgct ctgggtaaag ttcattggaa cagaaagaaa
 121 tggatttatc tgctcttcgc gttgaagaag tacaaaatgt cattaatgct atgcagaaaa
 181 tcttagagtg tcccatctgt ctggagttga tcaaggaacc tgtctccaca aagtgtgacc
 241 acatattttg caaattttgc atgctgaaac ttctcaacca gaagaaaggg ccttcacagt
 301 gtcctttatg taagaatgat ataaccaaaa ggagcctaca agaaagtacg agatttagtc
 361 aacttgttga agagctattg aaaatcattt gtgcttttca gcttgacaca ggtttggagt
 421 atgcaaacag ctataatttt gcaaaaaagg aaaataactc tcctgaacat ctaaaagatg
 481 aagtttctat catccaaagt atgggctaca gaaaccgtgc caaagactt ctacagagtg
 541 aacccgaaaa tccttccttg caggaaacca gtctcagtgt ccaactctct aaccttggaa
 601 ctgtgagaac tctgaggaca aagcagcgga tacaacctca aaagacgtct gtctacattg
 661 aattgggatc tgattcttct gaagataccg ttaataaggc aacttattgc agtgtgggag
 721 atcaagaatt gttacaaatc acccctcaag gaaccaggga tgaaatcagt ttggattctg
 781 caaaaaaggc tgcttgtgaa ttttctgaga cggatgtaac aaatactgaa catcatcaac
 841 ccagtaataa tgatttgaac accactgaga agcgtgcagc tgagaggcat ccagaaaagt
 901 atcagggtag ttctgtttca aacttgcatg tggagccatg tggcacaaat actcatgcca
 961 gctcattaca gcatgagaac agcagtttat tactcactaa agacagaatg aatgtagaaa
1021 aggctgaatt ctgtaataaa agcaaacagc ctggcttagc aaggagccaa cataacagat
1081 gggctggaag taaggaaaca tgtaatgata ggcggactcc cagcacagaa aaaaaggtag
1141 atctgaatgc tgatcccctg tgtgagagaa aagaatggaa taagcagaaa ctgccatgct
1201 cagagaatcc tagagatact gaagatgttc cttggataac actaaatagc agcattcaga
1261 aagttaatga gtggttttcc agaagtgatg aactgttagg ttctgatgac tcacatgatg
1321 gggagtctga atcaaatgcc aaagtagctg atgtattgga cgttctaaat gaggtagatg
1381 aatattctgg ttcttcagag aaaatagact tactggccag tgatcctcat gaggctttaa
1441 tatgtaaaag tgaaagagtt cactccaaat cagtagagag taatattgaa gacaaaatat
1501 ttgggaaaac ctatcggaag aaggcaagcc tccccaactt aagccatgta actgaaaatc
1561 taattatagg agcatttgtt actgagccac agataataca agagcgtccc ctcacaaata
1621 aattaaagcg taaaaggaga cctacatcag gccttcatcc tgaggatttt atcaagaaag
1681 cagatttggc agttcaaaag actcctgaaa tgataaatca gggaactaac caaacggagc
1741 agaatggtca agtgatgaat attactaata gtggtcatga gaataaaaca aaaggtgatt
1801 ctattcagaa tgagaaaaat cctaacccaa tagaatcact cgaaaaagaa tctgctttca
1861 aaacgaaagc tgaacctata agcagcagta taagcaatat ggaactcgaa ttaaatatcc
1921 acaattcaaa agcacctaaa aagaataggc tgaggaggaa gtcttctacc aggcatattc
1981 atgcgcttga actagtagtc agtagaaatc taagcccacc taattgtact gaattgcaaa
2041 ttgatagttg ttctagcagt gaagagataa agaaaaaaaa gtacaaccaa atgccagtca
2101 ggcacagcag aaacctacaa ctcatggaag gtaaagaacc tgcaactgga gccaagaaga
2161 gtaacaagcc aaatgaacag acaagtaaaa gacatgacag cgatactttc ccagagctga
2221 agttaacaaa tgcacctggt tcttttacta agtgttcaaa taccagtgaa cttaaagaat
2281 ttgtcaatcc tagccttcca agagaagaaa aagaagagaa actagaaaca gttaaagtgt
2341 ctaataatgc tgaagacccc aaagatctca tgttaagtgg agaaagggtt ttgcaaactg
2401 aaagatctgt agagagtagc agtatttcat tggtacctgg tactgattat ggcactcagg
2461 aaagtatctc gttactggaa gttagcactc tagggaaggc aaaaacagaa ccaaataaat
2521 gtgtgagtca gtgtgcagca tttgaaaacc ccaagggact aattcatggt tgttccaaag
2581 ataatagaaa tgacacagaa ggctttaagt atccattggg acatgaagtt aaccacagtc
2641 gggaaacaag catagaaatg gaagaaagtg aacttgatgc tcagtatttg cagaatacat
2701 tcaaggtttc aaagcgccag tcatttgctc cgttttcaaa tccaggaaat gcagaagagg
2761 aatgtgcaac attctctgcc cactctgggt ccttaaagaa acaaagtcca aaagtcactt
2821 ttgaatgtga acaaaaggaa gaaaatcaag gaaagaatga gtctaatatc aagcctgtac
2881 agacagttaa tatcactgca ggctttcctg tggttggtca gaaagataag ccagttgata
2941 atgccaaatg tagtatcaaa ggaggctcta ggttttgtct atcatctcag ttcagaggca
3001 acgaaactgg actcattact ccaaataaac atggactttt acaaaaccca tatcgtatac
3061 caccactttt tcccatcaag tcatttgtta aaactaaatg taagaaaaat ctgctagagg
3121 aaaactttga ggaacattca atgtcacctg aaagagaaat gggaaatgag aacattccaa
3181 gtacagtgag cacaattagc cgtaataaca ttagagaaaa tgtttttaaa gaagccagct
3241 caagcaatat taatgaagta ggttccagta ctaatgaagt gggctccagt attaatgaaa
3301 taggttccag tgatgaaaac attcaagcag aactaggtag aaacagaggg ccaaaattga
3361 atgctatgct tagattaggg gttttgcaac ctgaggtcta taaacaaagt cttcctggaa
3421 gtaattgtaa gcatcctgaa ataaaaaagc aagaatatga agaagtagtt cagactgtta
3481 atacagattt ctctccatat ctgatttcag ataacttaga acagcctatg ggaagtagtc
3541 atgcatctca ggtttgttct gagacacctg atgacctgtt agatgatggt gaaataaagg
3601 aagatactag ttttgctgaa aatgacatta aggaaagttc tgctgttttt agcaaaagcg
3661 tccagaaagg agagcttagc aggagtccta gccctttcac ccatacacat ttggctcagg
3721 gttaccgaag aggggccaag aaattagagt cctcagaaga gaacttatct agtgaggatg
3781 aagagcttcc ctgcttccaa cacttgttat ttggtaaagt aaacaatata ccttctcagt
3841 ctactaggca tagcaccgtt gctaccgagt gtctgtctaa gaacacagag gagaatttat
3901 tatcattgaa gaatagctta aatgactgca gtaaccaggt aatattggca aaggcatctc
3961 aggaacatca ccttagtgag gaaacaaaat gttctgctag cttgttttct tcacagtgca
```

4021 gtgaattgga agacttgact gcaaatacaa acacccagga tcctttcttg attggttctt
4081 ccaaacaaat gaggcatcag tctgaaagcc agggagttgg tctgagtgac aaggaattgg
4141 tttcagatga tgaagaaaga ggaacgggct tggaagaaaa taatcaagaa gagcaaagca
4201 tggattcaaa cttaggtgaa gcagcatctg ggtgtgagag tgaaacaagc gtctctgaag
4261 actgctcagg gctatcctct cagagtgaca ttttaaccac tcagcagagg gataccatgc
4321 aacataacct gataaagctc cagcaggaaa tggctgaact agaagctgtg ttagaacagc
4381 atgggagcca gccttctaac agctaccctt ccatcataag tgactcttct gcccttgagg
4441 acctgcgaaa tccagaacaa agcacatcag aaaaagcagt attaacttca cagaaaagta
4501 gtgaataccc tataagccag aatccagaag gcctttctgc tgacaagttt gaggtgtctg
4561 cagatagttc taccagtaaa aataaagaac caggagtgga aaggtcatcc ccttctaaat
4621 gcccatcatt agatgatagg tggtacatgc acagttgctc tgggagtctt cagaatagaa
4681 actacccatc tcaagaggag ctcattaagg ttgttgatgt ggaggagcaa cagctggaag
4741 agtctgggcc acacgatttg acggaaacat cttacttgcc aaggcaagat ctagagggaa
4801 ccccttacct ggaatctgga atcagcctct tctctgatga ccctgaatct gatccttctg
4861 aagacagagc cccagagtca gctcgtgttg gcaacatacc atcttcaacc tctgcattga
4921 aagttcccca attgaaagtt gcagaatctg cccagagtcc agctgctgct catactactg
4981 atactgctgg gtataatgca atggaagaaa gtgtgagcag ggagaagcca gaattgacag
5041 cttcaacaga aagggtcaac aaaagaatgt ccatggtggt gtctggcctg accccagaag
5101 aatttatgct cgtgtacaag tttgccagaa aacaccacat cactttaact aatctaatta
5161 ctgaagagac tactcatgtt gttatgaaaa cagatgctga gtttgtgtgt gaacggacac
5221 tgaaatattt tctaggaatt gcgggaggaa aatgggtagt tagctatttc tgggtgaccc
5281 agtctattaa agaaagaaaa atgctgaatg agcatgattt tgaagtcaga ggagatgtgg
5341 tcaatggaag aaaccaccaa ggtccaaagc gagcaagaga atcccaggac agaaagatct
5401 tcaggggct agaaatctgt tgctatgggc ccttcaccaa catgcccaca gatcaactgg
5461 aatggatggt acagctgtgt ggtgcttctg tggtgaagga gctttcatca ttcacccttg
5521 gcacaggtgt ccacccaatt gtggttgtgc agccagatgc ctggacagag gacaatggct
5581 tccatgcaat tgggcagatg tgtgaggcac ctgtggtgac ccgagagtgg gtgttggaca
5641 gtgtagcact ctaccagtgc caggagctgg acacctacct gataccccag atcccccaca
5701 gccactactg a



Appendix B: SNP coding

code	alts	alt 1	alt 2	alt 3	alt 4	rc
A	1	A				T
C	1	C				G
G	1	G				C
T	1	T				A
R	2	A	G			Y
Y	2	C	T			R
S	2	G	C			S
W	2	A	T			W
K	2	G	T			M
M	2	A	C			K
B	3	C	G	T		V
D	3	A	G	T		H
H	3	A	C	T		D
V	3	A	C	G		B
N	4	A	C	G	T	N

Appendix C: The single letter amino-acid code

Code	Amino Acid	Abbreviation
G	Glycine	Gly
P	Proline	Pro
A	Alanine	Ala
V	Valine	Val
L	Leucine	Leu
I	Isoleucine	Ile
M	Methionine	Met
C	Cysteine	Cys
F	Phenylalanine	Phe
Y	Tyrosine	Tyr
W	Tryptophan	Trp
H	Histidine	His
K	Lysine	Lys
R	Arginine	Arg
Q	Glutamine	Gln
N	Asparagine	Asn
E	Glutamic Acid	Glu
D	Apartic Acid	Asp
S	Serine	Ser
T	Threonine	Thr

Appendix D: Primers for allele-specific RT-PCR; cDNA synthesis of cancer predisposition genes

Gene	Primer (reverse)	Sequence (5'-3')
BRCA1	BRCA1 newRT B	tga gat ctt tgg ggt ctt cag cat
	BRCA1 newRT D	aag gcc ttc tgg att ctg gct tat ag
	BRCA1 newRT F	tgt aca atc aag tct tca ctg cc tt
BRCA2	BRCA2 newRT B	ctc atg tat ttt tca ggt ggc aac ag
	BRCA2 newRT D	tgt ggg tat gca ttt gca tct ttt ac
	BRCA2 newRT F	gcc ttt tgg gta tct gca cta ct
	BRCA2 newRT H	ttt gag ttt gga tga cca ttt tgt tg
MSH2	MSH2 newRT B	ccc atg ggc act gac agt taa cac tat g
MLH1	MLH1 newRT B	taa agg aat act atc aga agg caa gta ta
APC	APC newRT B	tga aag gac agt cat gtt gcc agt at
	APC newRT D	tga ctg tgc tcc tcc aac tcc tt
	APC newRT F	gga ctt aaa act gga gtt tgt gcc tg
	APC newRT H	tgt att cac tcc tat ctg agt gcc cc

Appendix E: Characteristics of subjects analysed by CSCE

Code	Sample	G File	MAN Sc1	MAN Sc2	Code	Sample	G File	MAN Sc1	MAN Sc2
A1	9902639	9251	68	57	E1	9704488	11250	15	12
A2	300694	18057	31	22	E2	0302758	18869	15	12
A3	9800105	11517	27	21	E3	901913	13532	15	12
A4	9708037	13169	24	21	E4	0310747	21243	15	12
A5	9903660	13247	24	15	E5	0205468	17601	15	10
A6	9800726	11029	24	15	E6	9703443	10880	14	15
A7	107063	10991	23	23	E7	0304022	19610	14	14
A8	0408756	22283	21	18	E8	9708711	12227	14	13
A9	0000400	13088	21	18	E9	9604191	9473	14	13
A10	304657	9577	21	18	E10	0210043	17247	14	13
A11	0406576	20707	21	15	E11	9706291	11298	14	11
A12	9706003	11048	20	20	E12	940976	8211	14	11
B1	0301989	19368	20	20	F1	0305787	19508	14	11
B2	203267	13970	20	14	F2	0001360	10002	14	11
B3	0307135	20102	20	14	F3	0104086	17050	14	11
B4	105546	16363	19	17	F4	9805481	12091	14	11
B5	9809983	13048	19	16	F5	9706173	11437	14	11
B6	9708294	11581	19	16	F6	9708070	10774	14	10
B7	9701668	9093	19	13	F7	0009226	15867	14	10
B8	95D2662	9726	18	15	F8	0301921	16659	14	10
B9	0308725	18688	18	12	F9	9902978	13054	13	16
B10	0309499	20050	17	20	F10	0204838	19900	13	13
B11	9701165	11660	17	17	F11	9909446	14146	13	13
B12	9606378	10246	17	17	F12	BLANK			
C1	BLANK				G1	94D2334	8257	13	13
C2	0305583	19370	17	17	G2	0211392	9270	13	13
C3	0400676	16656	17	17	G3	0009588	15614	13	13
C4	0406805	21267	17	17	G4	0007101	11676	13	13
C5	0206613	16450	17	14	G5	0403020	15403	13	13
C6	0405479	9376	17	14	G6	0401675	20628	13	13
C7	9902299	13470	17	13	G7	0402292	21074	13	13
C8	9902082	11145	17	11	G8	0309583	19455	13	11
C9	970636	10163	16	16	G9	9906931	13890	13	10
C10	0308469	16487	16	16	G10	9803840	12720	13	10
C11	0104002	16312	16	16	G11	0300959	17969	13	10
C12	0102443	16208	16	16	G12	0104703	15950	13	10
D1	0401352	12294	16	15	H1	0408444	17293	13	10
D2	0005906	15313	16	13	H2	310621	19800	13	10
D3	9908115	12014	16	10	H3	0005707	14370	13	8
D4	0203750	17304	15	16	H4	0208523	17787	12	14
D5	0001268	13316	15	16	H5	9907336	14632	12	13
D6	9806267	12196	15	15	H6	0107284	13847	12	13
D7	9706490	11406	15	15	H7	9809838	13477	12	12
D8	9701744	10244	15	15	H8	9608098	10127	12	12
D9	0107490	16637	15	15	H9	9602198	9380	12	12
D10	0105633	17029	15	15	H10	0401463	18395	12	12
D11	9802376	10352	15	13	H11	0308515	8389	12	12
D12	9906398	14343	15	12	H12	0204053	17824	7	7

Appendix F: Additional BRCA1 primers

Name of primer	Sequence
BRCA1 del12 F	caagaagagcaaagcatgga
BRCA1 del12 R	caagaagagcaaagatgga

Appendix G: Sequence of minigene construct BRCA1 Exons 3-6

gtcataacagctcaaagttgaacttattcactaagaatagctttatttttt
aaataaattattgagcctcatttatttttcttttttctccccccctaccctgctagTCTGGA — Exon 3
GTTGATCAAGGAACCTGTCTCCACAAAGTGTGACCA
CATATTTTGCAAgtaagtttgaatgtgttatgtggctccattattagcttttgtttt
Tgtccttcataacccaggaaacacctaactttatagaagctttacttttcttcaattaagtgagaac
Gaaaaatccaactccatttcattctttctcagagagtatatagttatcaaaagttggttgtaatcat
agttcctggtaaagttttgacatatattatctttttttttttt……

tgcttatgcagcatccaaaaacaattaggaaactattgcttgtaattcacctgccattac
ttttaaatggctcttaagggcagttgtgagattatcttttcatggctatttgccttttg
agtattctttctacaaaaggaagtaaattaaattgttctttctttctttataatttatag — ttc repeat
ATTTTGCATGCTGAAACTTCTCAACCAGAAGAAAGGGC
CTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAA — Exon 5
Ggtatataatttggtaatgatgctaggttggaagcaaccacag
taggaaaaagtagaaattatttaataacatagcgttcctataaaaccattcatcagaaaa
atttataaaagagtttttagcacacagtaaattatttccaaagttattttcctgaaagtt
ttatgggacatctgccttatacaggtattagaaacttactgcctttctctaatgcttcta
gtgtaaaaacttgcagacttatgtaaagtagggctgtatcgccgtgcccccattgtctgt
taatcttgttttatatttttgattgtgtttccttttctttttttttttttttttaagac — Alu JB
agggtcctgctctgtcactgaggctggagtgcagtggcgtgatctcggctcactgtagcc
tctgtctcccagcctcttcctgccttagcctcccaaatagctgggactacaggcacacgc
taccatgcccggccaatttttgtatttttgtagagatgaggttttaccatgttgcccag
gctggtaactcctgagctcaggtgatctgcccacctcggcctcccaaagtgctggggttc
acaggtgtgtgtttatttctatctaattatttacacaaacacaatgtatttatatattgt
gtatctcttctgctacaatgtaaattctatgagagtagtaattttgtctgtctcaacact — MIR 2
gtttttcctaagtttggtacatagtaggcactcagatgcttaaaggaatgaatgaattgt
gctttaattccactttactaaacccaaatctccctttggacattgttatctatgtgtttt
caaagaagtataatcataatttgacagaaatccttgagaggcagaactaagtgagggatt
gggcagggttcagatgttaagaacagtaagctcagcagggtgtgattgctcatgcctata
accctagcactctaggaggctgaggtgggatgattgcttgaggccaggagtttgaaatca
gcctgggcaacatagtgagaccccatctactaccaacaaaataaataaataaataatgtacatg — aat repeat
gtggcatatgcccatagtcctagctacttgggaggctatagtgggaggatagcttgagta
cagaagtctgaggctgcagtgagctatgattgtggcactgcatgctagcctgggcaatag — Alu Jo
agcaagaccctgtctctaaattaaacaaaaaaaaaaagtactctagttttctatgcaatgc
attatatctgctgtggatttagggcagtattatatcagataattttaggcatttggtagg
cttaaatgaatgacaaaaagttactaaatcactgccatcacacggtttatacagatgtca
atgatgtattgattatagaggttttctactgttgctgcatcttatttttatttgtttaca
tgtcttttcttattttagtgtccttaaaaggttgataatcacttgctgagtgtgtttctc
aaacaatttaatttcagGAGCCTACAAGAAAGTACGAGATTTA
GTCAACTTGTTGAAGAGCTATTGAAAATCATTTGTGCTTTT — Exon 6
CAGCTTGACACAGGTTTGGAGTgtaagtgttgaatatcccaagaatgcaact
caagtgctgtccatgaaaactcaggaagtttgcacaattacttt
ctatgacgtggtgataagacctttagtctaggttaattttagttctgtatctgtaatct
atttttaaaaaattactcccactggtctcacacctattttatcaatcgtaaggtgcaca
ttttcacatcttaacatctctgaaattgggaacattttactattgagggtgtgtcattt

Appendix H: Sequence of minigene construct BRCA1 Exons 5-7

ttttaaatg gctcttaagg gcagttgtga gattatcttt tcatggctat ttgccttttg
agtattcttt ctacaaaagg aagtaaatta aattgttctt tctttcttta taatttatag
ATTTTGCATGCTGAAACTTC TCAACCAGAAGAAAG — Exon 5
GGCCT TCACAGTGTC CTTTATGTAAGAATGATATA
ACCAAAAGgt atataatttg gtaatgatgc taggttggaa gcaaccacag
taggaaaaag tagaaattat ttaataacat agcgttccta taaaaccatt catcagaaaa
atttataaaa gagtttttag cacacagtaa attatttcca aagttatttt cctgaaagtt
ttatgggaca tctgccttat acaggtatta gaaacttact gcctttctct aatgcttcta
gtgtaaaaac ttgcagactt atgtaaagta gggctgtatc gccgtgcccc cattgtctgt
taatcttgtt tttatatttt tgattgtgtt tcctttttctt ttttttttt tttttaagacagggtcctgc — AluJB
tctgtcactg aggctggagt gcagtggcgt gatctcggct cactgtagcc
tctgtctccc agcctcttcc tgccttagcc tcccaaatag ctgggactac aggcacacgc
taccatgccc ggccaatttt tgtatttttt gtagagatga ggttttacca tgttgcccag
gctggtaact cctgagctca ggtgatctgc ccacctcggc ctcccaaagt gctggggttc
acaggtgtgt gtttatttct atctaattat ttacacaaac acaatgtatt tatatattgt — MIR2
gtatctcttc tgctacaatg taaattctat gagagtagta attttgtctg tctcaacact
gtttttccta agtttggtac atagtaggca ctcagatgct taaaggaatg aatgaattgt
gctttaattc cactttacta aacccaaatc tccctttgga cattgttatc tatgtgtttt
caaagaagta taatcataat ttgacagaaa tccttgagag gcagaactaa gtgagggatt
gggcagggtt cagatgttaa gaacagtaag ctcagcaggg tgtgattgct catgcctata
accctagcac tctaggaggc tgaggtggga tgattgcttg aggccaggag tttgaaatca — Alu Jo
gcctgggcaa catagtgaga ccccatctct accaacaaaa taaataaata aatgtacatg — aat repeat
gtggcatatg cccatagtcc tagctacttg ggaggctata gtgggaggat agcttgagta
cagaagtctg aggctgcagt gagctatgat tgtggcactg catgctagcc tgggcaatag — Pst I
agcaagaccc tgtctctaaa ttaaacaaaa aaaaaagtac tctagttttc tatgcaatgc
attatatctg ctgtggattt agggcagtat tatatcagat aattttaggc atttggtagg
cttaaatgaa tgacaaaaag ttactaaatc actgccatca cacggtttat acagatgtca
atgatgtatt gattatagag gttttctact gttgctgcat cttatttttа tttgtttaca
tgtcttttct tattttagtg tccttaaaag gttgataatc acttgctgag tgtgtttctc
aaacaattta atttcag GAG CCTACAAGAA AGTACGAGAT TTAG — Exon 6
TCAACT TGTTGAAGAGCTATTGAAAA TCATTTGTGC TTT
TCAGCTT GACACAGGTT TGGAGTgtaa gtgttgaata tcccaagaat
gcaactcaag tgctgtccat gaaaactcag gaagtttgca caattacttt
ctatgacgtg gtgataagac ctttagtct aggttaattt tagttctgta tctgtaatct
atttttaaaa aattactccc actggtctca caccttattt tatcaatcgt aaggtgcaca
ttttcacat cttaacatct ctgaaattgg gaacatttta ctattgaggg tgtgtcattt
gtttaatttg tgtgctttct ttcttagtga tacacgaaat aatagtgcca cttacattgt
tggtgtctta gctttagtga aatacagtat tgataggcaa atttcttagt gttaaggtag
aaaacaagga ctctaaataa ctttgatggt ctgtgtattt gtttttgttt cctaggagta
aaatttccag ttgatttttt aaaatttgat ttttaaaaaa aatcacaggt aaccttaatg
cattgtctta acacaacaaa gagcatacat agggtttctc ttggtttctt tgattataat
tcatacattt ttctctaact gcaaacataa tgttttccct tgtattttac ag ATGCAAAC — Exon 7
AGCTATAATT TTGCAAAAAA GGAAAATAAC TCTCCTGAAC
ATCTAAAAGA TGAAGTTTCTATCATCCAAA GTATGGGCTA
CAGAAACCGT GCCAAAAGAC TTCTACAGAG TGAACCCGAA
AATCCTTCCT TGgtaaaacc atttgttttc ttcttcttct tcttcttctt ttcttttttt tttcttttttt — ttc repeat
ttttttgaga tggagtcttg ctctgtggcc caggctagaa gcagtcctcc tgccttagcc ccctagtag a — Alu Jo
— Alu S

Appendix I: Sequence of minigene construct BRCA1 Exons 9-11

gccacagtag atgctcagta aatatttcta gttgaatatc tgtttttcaa caagtacatt MIR2
tttttaaccc ttttaattaa gaaaactttt attgatttat tttttggggg gaaatttttag
GATCTGAT TCTTCTGAAG ATACCGTTAA TAAGGCAACT Exon 9
TATTGCAGgt gagtcaaaga gaacctttgt ctatgaagct ggtatttttcc tatttagtta
atattaagga ttgatgtttc tctcttttta aaaata
ctccctcaag taggctggtg tttctccaga ctagaatcat ggtattggaa gaaaccttag
agatcatcta gtttagttct ctcattttat agtggaggaa ataccctttt tgtttgttgg
atttagttat tagcactgtc caaaggaatt taggataaca gtagaactct gcacatgctt
gcttctagca gattgttctc taagttcctc atatacagta atattgacac agcagtaatt
gtgactgatg aaaatgttca aggacttcat tttcaactct ttctttcctc tgttccttat
ttccacatat ctctcaagct ttgtctgtat gttatataat aaactacaag caaccccaac
tatgttacct accttcctta ggaattattg cttgacccag gttttttttt tttttttttt
ggagacgggg tcttgccctg ttgccaggat ggagtgtagt ggcgccatct cggctcactg Alu Sc
caatctccaa ctccctggtt caagcgattc tcctgtctca atctcacgag tagctgggac
tacaggtata caccaccacg cccggttaat tgaccattcc atttctttct ttctc ctt repeat
tttttttt ttttgagac agagtcttgc tctgttgccc aggctggagt acagaggtgt
gatctcacct ctccgcaacg tctgcctccc aggttgaagc catactcctg cctcagcctc
tctagtagct gggactacag gcgcgcgcca ccacacccgg ctaatttttg tatttttagt Alu Sx
agagatgggg tttcaccatg ttggccaggc tggtcttgaa ctcatgacct caagtgatcc
accgcctca gcctcccaaa gtgctgggat tacaggcttg agccaccgtg cccagc aacc
atttcatttc aactagaagt ttctaaagga gagagcagct ttcactaact aaataagatt
ggtcagcttt ctgtaatcga aagagctaaa atgtttgatc ttggtcattt gacagttctg
catacatgta actagtgttt cttattagga ctctgtcttt tccctatagT GTGGGAGATC
AAGAATTGTT ACAAATCACC CCTCAAGGAA CCAGGGATGA Exon 10
AATCAGTTTG GATTCTGCAA AAAAGG gtaa tggcaaagtt tgccaactta
acaggcactg aaaagagagt gggtagatac agtactgtaa ttagattatt ctgaagacca tttgggacct
ttacaaccca caaaatctct tggcagagtt agagtatcat tctctgtcaa atgtcgtggt atggtctgat agatttaaat
ggtactagac taatgtacct ataataagac cttctgtaac tgattgttgc cctttcgttt
ttttttttgt ttgtttgttt gttttttttt gagatggggt ctcactctgt tgcccaggct
ggagtgcagt gatgcaatct tggctcactg caacctccac ctccaaggct caagctatcc Alu JB
tcccacttca gcctcctgag tagctgggac tacaggcgca tgccaccaca cccggttaat
ttttgtggt tttatagaga tggggtttca ccatgttacc gaggctggtc tcaaactcct
ggactcaagc agtctgccca cttcagcctc ccaaagtgct gcagttacag gcttgagcca
ctgtgcctgg cctgcccttt actttaatt ggtgtatttg tgtttcatct tttacctact
ggtttttaaa tatagggagt ggtaagtctg tagatagaac agagtattaa gtagacttaa
tggccagtaa tctttagagt acatcagaac cagttttctg atggccaatc tgcttttaat
tcactcttag acgttagaga aataggtgtg gtttctgcat agggaaaatt ctgaaattaa
aaatttaatg gatcctaagt ggaaataatc taggtaaata ggaattaaat gaaagagtat
gagctacatc ttcagtatac ttggtagttt atgaggttag tttctctaat atagccagtt
ggttgatttc cacctccaag gtgtatgaag tatgtatttt tttaatgaca attcagtttt
tgagtacctt gttatttttg tatattttca gCTGCTTGTG AATTTTCTGA GACGGATGTA Exon 11
ACAAATACTG AACATCATCA ACCCAGTAAT AATGATTT
GA ACACCACTGA GAAGCGTGCAGCTGAGAGGC ATCCAGAAAA
GTATCAGGGT AGTTCTGTTT CAAACTTGCA TGTGGAGCCA
TGTGGCACAA ATACTCATGC CAGCTCATTACAGCATGAGA
ACAGCAGTTT ATTACTCACTAAAGACAGAA TGAATGTAGA
AAAGGCTGAA TTCTGTAATA AAAGCAAACA GCCTGGCTTA

Appendix J: Sequence of minigene construct BRCA1 Exons 17-19

ctttaaatag ttccaggaca cgtgtagaac gtgcaggatt gctacatagg taaacatatg LIPA09
ccatggtcga ataactagta ttctgagctg tgtgctagag gtaactcatg ataatggaat
atttgattta atttcag ATG CTCGTGTACA AGTTTGCCAG AAAA
CACCAC ATCACTTTAACTAATCTAAT TACTGAAGAG AC Exon 17
TACTCATG TTGTTATGAA AACAGgtata ccaagaacct ttacagaata
ccttgcatct gctgcataaa accacatgag gcgaggcacg gtggcgcatg cctgtaatcg
cagcactttg ggaggccgag gcgggcagat cacgagatta ggagatcgag ccatcctgg Alu Sy
ccagcatggt gaaaccccgt ctctactaaa aaataaaaaa attagctggg
tgtggtcgcg tgcgcctgta gtcccagcta ctcgtgaggc tgaggcagga gaatcacttg
aaccggggag atggaggttg cagtgagccg agatcatgcc actgcattcc agcctggcga
cagagcaagg ctccgtctca aaaaaaaaaa aaaaaaacgt gaaaaaataa gaatatttgt
tgagcatagc atggatgata gtcttctaat agtcaatcaa ttactttatg aaagacaaat
aatagttttg ctgcttcctt acctcctttt gttttgggtt aagatttgga gtgtgggcca

agtaatatag atgttctccg tccacattaa
tcaaaactat tgacatggat aattcctaat tccttgaaca ctataatgga gatctatagc
tagccttggc gtctagaaga tgggtgttga gaagagggag tggacagata tttcctctgg
tcttaacttc atatcagcct cccctagact tccaaatatc catacctgct ggttataatt
agtggtgttt tcagcctctg attctgtcac caggggtttt agaatcataa atccagattg
atcttgggag tgtaaaaaac tgaggctctt tagcttctta ggacagcact tcctgatttt
gttttcaact tctaatcctt tgagtgtttt tcattctgca g ATGCTGAGT TTGTGTG Exon 18
TGAACGGACACTG AAATATTTTC TAGGAATTGC GGGAGGAAA
A TGGGTAGTTA GCTATTTCTgtaagtataat actatttctc ccctcctccc tttaaca
cct cagaattgca tttttacacctaacgtttaa cacctaaggt ttttgctgat gctgagtctg agttaccaaa
aggtctttaattgtaatact aaactacttt tatctttaat atcactttgt tcagataagc tggtgatgct
gggaaaatgg gtctctttta taactaatag gacctaatct gctcctagca atgttagcat
atgagctagg gatttattta atagtcggca ggaatccatg tgcagcaggc aaacttataa
tgtttaaatt aaacatcaac tctgtctcca gaaggaaact gctgctacaa gccttattaa
agggctgtgg ctttagaggg aaggacctct cctctgtcat tcttcctgtg ctcttttgtg
aatcgctgac ctctctatct ccgtgaaaag agcacgttct tctgctgtat gtaacctgtc
ttttctatga tctctttag G GGTGACCCAG TCTATTAAAG AAAGAAAAAT Exon 19
GCTGAATGAGgtaagtactt gatgttacaa actaaccaga gatattcatt cagtcatata
gtaaaaatg atttgcttc cttccatcaa tgcaccactt tccttaacaa tgcacaaatt ttccatgata

***Appendix K*:**

231 mutant variant primers

Primer	Site of mutagenesis	Sequence
1	c224	GA GCC TAC AAG (C/G/T) AA GTA CTA GAT TT
2	c225	GA GCC TAC AAG A (C/G/T)A GTA CTA GAT TT
3	c225	GA GCC TAC AAG AA (C/G/T) GTA CTA GAT TT
4	c227	GA GCC TAC AAG AAA (T/A/C) TA CTA GAT TTA G
5	c228	GA GCC TAC AAG AAA G (A/C/G)A CTA GAT TTA G
6	c229	TAC AAG AAA GT (C/G/T) CTA GAT TTA GTC A
7	c230	TAC AAG AAA GTA (G/T/A) TA GAT TTA GTC A
8	c231	TAC AAG AAA GTA C (A/C) A GAT TTA GTC A
9	c232	TAC AAG AAA GTA CT (C/G/T) GAT TTA GTC AA
10	c233	TAC AAG AAA GTA CTA (T/A/C) AT TTA GTC AAC
12	c234	TAC AAG AAA GTA CTA G (C/G/T)T TTA GTC AAC T
13	c235	TAC AAG AAA GTA CTA GA (A/C/G) TTA GTC AAC TT
14	c236	AAA GTA CTA GAT (A/C/G) TA GTC AAC TTG
15	c237	AAA GTA CTA GAT T (A/C/G)A GTC AAC TTG T
16	c238	AAA GTA CTA GAT TT (C/G/T) GTC AAC TTG TT

231 wildtype variants

Primer	Site of mutagenesis	Sequence
1	c224	GA GCC TAC AAG (C/G/T) AA GTA CGA GAT TT
2	c225	GA GCC TAC AAG A (C/G/T)A GTA CGA GAT TT
3	c225	GA GCC TAC AAG AA (C/G/T) GTA CGA GAT TT
4	c227	GA GCC TAC AAG AAA (T/A/C) TA CGA GAT TTA G
5	c228	GA GCC TAC AAG AAA G (A/C/G)A CGA GAT TTA G
6	c229	TAC AAG AAA GT (C/G/T) CGA GAT TTA GTC A
7	c230	TAC AAG AAA GTA (G/T/A) GA GAT TTA GTC A
8	c231	TAC AAG AAA GTA C (A/C) A GAT TTA GTC A
9	c232	TAC AAG AAA GTA CG (C/G/T) GAT TTA GTC AA
10	c233	TAC AAG AAA GTA CGA (T/A/C) AT TTA GTC AAC
12	c234	TAC AAG AAA GTA CGA G (C/G/T)T TTA GTC AAC T
13	c235	TAC AAG AAA GTA CGA GA (A/C/G) TTA GTC AAC TT
14	c236	AAA GTA CGA GAT (A/C/G) TA GTC AAC TTG
15	c237	AAA GTA CGA GAT T (A/C/G)A GTC AAC TTG T
16	c238	AAA GTA CGA GAT TT (C/G/T) GTC AAC TTG TT

Appendix L: MDM2 sequence of exon 1, intron 1-2 and exon 2

(Reference sequence = ENST00000258148)

CGCGAAAACCCCGGATGGTGAGGAGCAGgtactggcccggcagcgagcggtcacttttgggtctgggctctgacggtg
tcccctctatcgctggttcccagcctctgcccgttcgcagcctttgtgcggttcgtggctgggggctcggggcgcggggcgcggggcatggg
gcacgtggctttgcggaggttttgttggactggggctaggcagtcgccgccagggaggagggcgggatttcggacggctctcgcggcggt
gggggtgggggtggttcggaggtctccgcgggagttcagggtaaaggtcacgggggccgggggctgcggggccgcttcgKcgcgggc
ggtccggatgatcgcaggtgcctgtcgggtcactagtgtgaacgctgcgcgtagtctgggcgggattgggccggttcagtgggcaggttga
ctcagcttttcctcttgagctggtcaagttcagacacgttccgaaactgcagtaaaaggagttaagtcctgacttgtctccagctggggctattta
aaccatgcattttcccagctgtgttcagtggcgattggagggtagacctgtgggcacggacgcacgccacttttctctgctgatccaggtaag
caccgacttgcttgtagctttagttttaactgttgtttatgttctttatatatgatgtattttccacagatgtttcatgatttccagttttcatcgtgtctttttttc
cttgtagGCAAATGTGCAATACCAACATGTCTGTACCTACTGATGGTGCTGTAACCACCTCACAGA
TTCCAGCTTCGGAACAAGAGACCCTG

gtctccgcgggagttca	MDM2 Fprimer
gtgtgaacgctgcgcgtagtc	MDM2 R primer
KCGGCGCGGGAGGTCCGGATGATCGC	Sequence analysed by pyrosequencer
K	MDM2 SNP309

Appendix M: Characteristics of subjects genotyped for MDM2 SNP309

Number	Family	BRCA1 mutation	Genotype at MDM2 SNP309	Breast ca. yes/no?	age of onset	ovarian ca.? yes/no?	age of onset
9704733	G10477	5382ins1bp ex20	G/G	0	0	0	0
9902509	G12488	T to A @1445, ex 11D	G/G	1	49	0	0
9807451	G12788	185delAG	G/G	0	0	0	0
9809660	G12862	188del11 c24 ex2	G/G	0	0	1	51
0000196	G13481	185delAG ex2	G/G	0	0	1	58
9907064	G14331	1131A>T c338, ex11B	G/G	1	46	0	0
9905701	G14368	3519G>T c1134, ex11	G/G	0	0	1	74
0009467	G14728	ex 20 deletion	G/G	1	44	0	0
0207789	G18636	4693delAA,c1525, ex15	G/G	0	0	1	66
92D1268	G9143	3785del4	G/G	1	38	0	0
9804038	G9452	2594del1bp ex11J	G/G	0	0	0	0
0404582	G9301	4446C>T c1443 ex13	G/G	0	0	0	0
9701443	G10110	3882delAA ex11Q	G/G	1	29	0	0
9604829	G10109	ex 11L mutation	G/T	1	35	0	0
9604831	G10109	ex 11L mutation	G/T	1	42	0	0
9907536	G10129	4184del4 c1355 ex11	G/T	1	29	0	0
93D0059	G10133	3695insT ex11P	G/T	1	34	0	0
0204551	G10241	3825 delAA c1236 ex11	G/T	1	50	0	0
9703225	G10477	5382ins1bp ex20	G/T	0	0	1	68
0400694	G10526	3936C>T c1273, ex11	G/T	0	0	0	0
104738	G12087	4184del4 ex11R	G/T	0	0	0	0
9709525	G12087	4184del4 ex11R	G/T	1	48	0	0
309674	G12488	T to A @1445, ex 11D	G/T	0	0	0	0
9802467	G12488	T to A @1445, ex 11D	G/T	1	56	1	65
9902093	G12788	185delAG	G/T	0	0	0	0
0302623	G13049	4162delG c1348 ex11	G/T	1	30	0	0
9908374	G13521	1294del4, c392, ex11c	G/T	0	0	0	0
9906277	G13521	1294del4, c392, ex11c	G/T	0	0	0	0
9909512	G13868	ex 9-12 deletion	G/T	0	0	0	0
9909511	G13868	ex 9-12 deletion	G/T	1	29	0	0
9908129	G14599	3450delCAAG	G/T	0	0	0	0
0001028	G14871	185delAG ex2	G/T	1	35	0	0
0007221	G15681	1940del4 c607 ex11G	G/T	1	52	0	0
0006667	G15698	3450del4 c1111 ex11Q	G/T	1	40	0	0
0309578	G16335	1260A>T c381 ex11	G/T	1	34	0	0
302184	G16388	3819del5	G/T	0	0	0	0
0201668	G17679	deletion ex21-23	G/T	1	43	0	0
0203915	G17714	2594delC c825 ex11	G/T	1	49	0	0
0308050	G18882	5382insC ex 20	G/T	0	0	0	0
0303340	G19533	2774delCT c886 ex11	G/T	0	0	0	0
0410893	G20172	ex13 duplication	G/T	0	0	0	0
0310494	G20843	3185delA c1022 ex11	G/T	1	33	0	0
0502104	G22746	5272G>A c1718	G/T	1	41	0	0
0500877	G22883	5382insC ex 20	G/T	0	0	0	0
9706377	G3372	4304 G>A ex12	G/T	0	0	0	0
9801581	G3372	4304 G>A ex12	G/T	1	31	0	0
92D0845	G49452	2594del1bp ex11J	G/T	1	30	0	0
9502835	G9094	4184del4 ex 11R	G/T	1	39	0	0
92D0741	G9143	3785del ex11	G/T	1	37	0	0
9504394	G9283	11B mutation	G/T	1	39	1	51
9602053	G9372	4304 G>A ex12	G/T	1	42	0	0
96/5816	G9372	4304 G>A ex12	G/T	0	0	0	0
96/5818	G9372	4304 G>A ex12	G/T	0	0	0	0
93D0319	G9383	5382ins1bp ex20	G/T	1	35	0	0
9608526	G9402	364delT ex6	G/T	1	32	0	0
9708693	G9452	2594del1bp ex11J	G/T	0	0	0	0
9601122	G9452	2594del1bp ex11J	G/T	1	47	0	0
9600902	G9845	185del2	G/T	0	0	0	0
9700787	G9872	4184del4 ex 11R	G/T	0	0	0	0
9700790	G9872	4184del4 ex11R	G/T	0	0	0	0
93D1276	G9872	4184del4 ex11R	G/T	1	31	0	0
9701046	G9931	5382inC, c1755, ex20	G/T	0	0	1	57
9709149	G9931	5382inC, c1755, ex20	G/T	0	0	0	0
9705229	G9931	5382inC, c1755, ex20	G/T	0	0	0	0
9802883	G9931	5382inC, c1755, ex20	G/T	0	0	0	0
9701061	G9931	5382inC, c1755, ex20	G/T	0	0	0	0
95D1158	G9143	3785del ex11	G/T	0	0	0	0
9704896	G9452	2594del1bp ex11J	G/T	0	0	0	0
9808630	G10129	4184del4 c1355 ex11	T/T	0	0	0	0
0310775	G10241	3825 delAA c1236 ex11	T/T	0	0	0	0
0307334	G10526	3936C>T c1273, ex11	T/T	1	33	0	0
0408557	G10526	3936C>T c1273, ex11	T/T	0	0	0	0

References:

1. Antonarakis, S.E. Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Human Mutation*, 1998; **11**: 1-3.
2. Szabo, C., Masiello, A., Ryan, J.E., Consortium, T.B., and Brody, L.C. The breast cancer information core: database design, structure and scope. *Human Mutation*, 2000; **16**: 123-131.
3. *http://www.cancerresearch.org.uk/aboutcancer/statistics.*
4. Easton, D.F., Ford, D. and Bishop, D.T. Breast and ovarian cancer incidence in BRCA1 mutation carriers: The Breast Cancer Linkage Consortium. *Am J Hum Genet*, 1995; **56**: 265-271.
5. Wooster, R. and Weber, B.L. Breast and ovarian cancer. *N Engl J Med,* 2003; **348**: 2339-2347.
6. The CHEK2 breast cancer case-control consortiumCHEK2*1100delC and susceptibility to breast cancer: A collaberative analysis involving 10.860 breat cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet*, 2004; **74**: 1175-1182.
7. Renwick, A., Thompson, D., Seal, S., Kelly, P. and Chagtai, T. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature Genet*, 2006; **38**: 873-875.
8. Seal, S., Thompson, D., Renwick, A., Elliott, A., Kelly, P. *et al.* Truncating mutations in the Fancomi anaemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nature Genet*, 2006; **38**: 1239-1241.
9. Rahman, N., Seal, S., Thompson, D., Kelly, P. and Renwick, A. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genet*, 2007; **39**:165-167.
10. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 2007; 447: 1087-1093.
11. Miki, Y., Swensen, J., Shattuck-Eideus, D., Futreal . P.A., Harshman, K. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science,* 1994; **266**: 66-71.
12. Wooster, R., Bignell, G., Lancaster, J., Swift, S, Seal, S. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature*, 1995; **378**: 789-792.
13. Malkin, D., Li, F.P., Strong, L.C., Fraumeni J.F. Jr, Nelson C.E. *et al.* Germline p53 mutations in a familial syndrome of breast cancer, sarcomas and other neoplasms. *Science*, 1990; **250**: 1233-38.
14. Nelen, M.R., Padberg, G.W., Peeters, E.A., Lin, A.Y., van den Helm, B. *et al.* Localisation of the gene for Cowden disease to chromosome 10q22-23. *Nature Genet*, 1996; **13**: 114-16.
15. Boardman, L., Thibodeau, S.N., Schaid, D.J., Lindor, N.M., McDonnrll, S.K. *et al.* Increased risk for cancer in patients with Peutz-Jeghers syndrome. *Ann Intern Med*, 1998; **128**: 896-99.

16. Pharoah, P.D., Guilford, P. and Caldas, C. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffse gastric cancer families. *Gastroenterology*, 2001; **121**: 1348-53.
17. Dunning, A.M., Ellis, P.D., Mcbride, S., Kirschenlohr, H.L., Healey, C.S. *et al.* A transforming growth factor beta1 signal peptide variant increases secretion in vitro and is associated with increased incidence of invasive breast cancer. *Cancer Research*, 2003; **63**: 2610-15.
18. Welcsh, P.L., Schubert, E.L. and King, M.C. Inherited breast cancer: an emerging picture. *Clin Genet*, 1998; **54**: 447-458.
19. Ford, D., Easton, D.F., Bishop, D.T., Narod, S.A. and Godgar, D.E. Risk of cancer in BRCA1 mutation carriers. *Lancet*, 1994; **343**: 692-695.
20. Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E., Anderson, L.A. *et al.* Linkage of early-onset breast cancer to chromosome 17q21. *Science*, 1990; **250**:1684-1689.
21. Easton, D.F., Ford, D., Bishop, D.T., Ford, D and Crockford, G.P. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *Am J Human Genet*, 1993; **52**: 678-701.
22. Wooster, R., Neuhausen, S., Mangion, J. *et al.* Localisation of a breast cancer susceptibility gene to chromosome 13q12-13. *Science*, 1994; **65**: 2088-1090.
23. Struewing, J.P., Hartge, P., Wacholder, S., Baker, S.M., Berlin, M. *et al.* The risk of cancer associated with specific mutations of BRCA1 and BRCA2 mutations in Ashkenazi Jews. *N Engl J Med*, 1997; **336**: 943-948.
24. King, M.-C., Marks, J.H.and Mandell, J.B. for the New York Breast Cancer Study Group. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*, 2003; **302**: 643-646.
25. Gayther, S.A. and Ponder, B.A. Mutations of the BRCA1 and BRCA2 genes and the possibilities for predictive testing. *Mol Med Today*, 1997; **3**: 168-174.
26. Futreal, P.A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K. *et al.* BRCA1 mutations in primary breast and ovarian carcinomas. *Science*, 1994; **266**:12-122.
27. Teng, D.H., Bogden, R., Mitchell, J., Baumgard, M., Bell, R. *et al.* Low incidence of BRCA2 mutations in breast carcinoma and other cancers. *Nature Genet*, 1996; **13**: 241-244.
28. Rajan, J.V., Marquis, S.T., Gardner, H.P. and Chodosh, L.A. Developmental expression of BRCA2 is colocalised with BRCA1 and is associated with proliferation and differentiation in multiple tissues. *Dev Biol*, 1997; **184**: 385-401.
29. Wessels, L.F., van Welsem, T., Hart, A.A., van 't Veer, L. Reinders, M.J. and Nederlof, P.M. Molecular classification of breast carcinomas by comparative genomic hybridisation: a specific somatic genetic profile for BRCA1 tumours. *Cancer Res*, 2002; **62**: 7110-7117.

30. Jonsson, G., Naylor, T.L., Vallon-Christersson, J., Staaf, J. Huang, J. *et al.* Distinct genomic profiles in hereditary breast tumours identified by array-based comparative genomic hybridisation. *Cancer Res*, 2005; **65**: 7612-7621.
31. Abbott, D.W., Freeman, M.L. and Holt, J.T. Double-strand break repair deficiency and radiation senstivity in BRCA2 mutant cancer cells. *J Natl Cancer Inst*, 1998; **90**: 978-985.
32. Husain, A., He, G., Ventkatraman, E.S. and Spriggs, D.R. BRCA1 up-regulation is associated with repair-mediated resistance to cis-diamminedichloroplatinum (II*). Cancer Res*, 1998; **58**:1120-1123.
33. Howlett, N.G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q. *et al.* Biallelic inactivation of BRCA2 in Fanconi Anaemia. *Science*, 2002; **297**: 606-609.
34. Shen, S.X., Weaver, Z., Xu, X., Li, C., Weinstein, M. *et al.* A targeted disruption of the murine Brca1 gene causes gamma-irradiation hypersensitivity and genetic instability. *Oncogene*, 1998; **17**: 3115-3124.
35. Moynahan, M.E., Chiu, J.W., Koller, B.H. and Jasin, M. Brca1 controls. homology-directed DNA repair. *Mol Cell*, 1999; **4**: 511-518.
36. Moynahan, M.E., Pierce, A.J. and Jasin, M. Brca2 is required for homology-directed repair of chromosomal breaks. *Mol Cell*, 2001, **7**: 263-272.
37. *Tutt, A., Bertwistle, D., Valentine, J., Gabriel, A., Swift, S. et al.* Mutation in Brca2 stimulates error-prone homology directed repair of DNA double-strand breaks occurring between repeated sequences. *Embo J*, 2001; **20**: 4704-4716.
38. Tutt, A. and Ashworth, A. The relationship between the roles of BRCA genes in DNA repair and cancer predisposition. *Trends Mol Med*, 2002; **8**: 571-576.
39. Welcsh, P.L. and King, M. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet*, 2001; **10**: 705-713.
40. Wang, Y., Cortez, D., Yazdi, P., Neff, N. Elledge, S.J. and Qin, J. BASC, a super complex of BRCA1 associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev*, 2000; **14**: 927-939.
41. Zhong, Q., Chen, C.F., Li, S., Chen, Y. Wang, C.C. *et al.* Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. *Science*, 1999; **285**: 747-750.
42. Kerr, P. and Ashworth, A. New complexities for BRCA1 and BRCA2. *Curr Biol,* 2001; **11**: R668-676.
43. Pellegrini, L., Yu, D.S., Lo, T., Anand, S. Lee, M. et al. Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature,* 2002; **420**: 287-93.
44. Yu, D.S., Sonoda, E., Takeda, S., Huang, C.L. Pellegrini L. *et al.* Dynamic control of Rad51 recombinase by self-association and interaction with BRCA2. *Mol Cell*, 2003; **12**: 1029-1041.
45. Lee, J.-S. and Chung, J.H. Diverse functions of BRCA1 in the DNA damage repair response. *Exp Rev Mol Med*, 2001; 2001: 1-11.

46. Scully, R., Anderson, S.F., Chao, D.M., Wei, W., Ye, L. *et al.* BRCA1 is a component of the RNA polymerase II holoenzyme. *Proc Natl Acad Sci.USA*, 1997; **94**: 5605-5610.
47. Kleiman, F.E. and Manley, J.L. Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. *Science*, 1999; **285**: 1576-1579.
48. Monteiro, A.N. BRCA1: exploring the links to transcription. *Trends Biochem Sci*, 2000; **25**: 469-474.
49. Rosen, E., Fan, S., Pestall, R.G. and Goldberg, I.D. BRCA1 gene in breast cancer. *J Cell Physiol*, 2003; **196**: 19-41.
50. Deng, C.X. BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nuc Acid Res*, 2006; **34**: 1416-1426.
51. The Breast Cancer Linkage Consortium. Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. *Lancet*, 1997; **349**: 1505-1510.
52. Johannsson, O.T., Idval, I., Anderson, C., Borg, A., Barkardottir, R.B. *et al.* Tumour biological features of BRCA1-induced breast and ovarian cancer. *Eur J Cancer*, 1997; **33**: 362-371.
53. Lakhani, S.R., Jacquenier, J., Sloane, J.P., Gusterson, B.A., Anderson, T.J. *et al.* Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *J Natl Cancer Inst*, 1998; **90**: 1138-1145.
54. Phillips, K.A., Andrulis, I.L. and Goddwin, P.J. Breast carcinomas arising in carriers of mutations in BRCA1 or BRCA2: are they prognostically different? *J Clin Oncol*, 1999; **17**: 3653-3663.
55. Robson, M., Rajan, P., Rosen, P.P., Gilewski, T., Hirschaut,Y. *et al.* BRCA-associated breast cancer: absence of a characteristic phenotype. *Cancer Res*, 1998; **58**: 1839-1842.
56. Lakhani, S.R., Van de Vijer, M.J., Jacquemier, J., Anderson, T.J., Osin, P.P. *et al.* The pathology of familial breast cancer: Predictive value of immunohistochemical markers oestrogen receptor, PR, HER-2 and p53 in patients with mutations in BRCA1 and 2. *J Clin Oncol*, 2002; **9**: 2310-2318.
57. Marcus, J.N., Watson, P., Page, D.L., Narod, S.A., Tonin, P. *et al.* BRCA2 hereditary breast cancer pathophenotype. *Breast Cancer Res Treat*, 1997; **44**: 275-277.
58. Crook, T., Brooks, L.A., Crossland, S., Osin, P., Barker, K.T. *et al.* p53 mutation with frequent novel condons but not a mutator phenotype in BRCA1 and BRCA2 associated breast tumours. *Oncogene*, 1998; **17**: 1681-1689.
59. Glebov, O.K., McKenzie, K.E., White, C.A., and Sukumar, S. Frequent p53 gene mutations and novel alleles in familial breast cancer. *Cancer Res*, 1994; **54**: 3703-3709.
60. Phillips, K.A., Nichol, K., Ozcelik, H., Knight, J. Done, S.J. *et al.* Frequency of p53 mutations in breast carcinomas from Ashkenazi Jewish carriers of BRCA1 mutations. J *Natl Cancer Inst*, 1999; **91**: 469-473.

61. Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S. *et al.* *Molecular portraits of human breast tumours. Nature*, 2000; **406**: 747-752.
62. van't Veer, L.J., Dai, H.Y., Van de Vijer, M.J., He, Y.D., Hart, A.A. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 2002: 530-6.
63. Gruvberger, S., Ringner, M., Chen, Y.D. and Panavally, S. Oestrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res*, 2001; **61**: 5979-5984.
64. Foulkes, W.D., Stefansson, I.M., Chappuis, P.O., Begin, L.R., Goffin, J.R., Wong, N., Trudel, M., and Akslen, L.A. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *JNCI*, 2003; **95**: 1482-1485.
65. Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S. *et al.* Repeated observation of breast tumour subtypes in independent gene expression data sets. *Proc Natl Acad Sci*, 2003; **100**: 8418-8423.
66. Lakhani, S.R., Reis-Filho, J.S., Fulford, L., Penault-Llorca, F., van der Vijver, M. *et al.* Prediction of BRCA1 status in patients with breast cancer using oestrogen receptor and basal phenotype. *Clin Cancer Res,* 2005; **11**: 5175-5180.
67. Palacios, J., Honrado, E., Osorio, A., Cazorla, A., Sarrio, D. *et al.* Phenotypic characterisation of BRCA1 and BRCA2 tumours based in a tissue microarray study with 37 immunohistochemical markers. *Breast Cancer Res Treat*, 2005; **90**: 5-14.
68. Arnes, J.B., Brunet, J.S., Stefansson, I., Begin, L.R., Wong, N. *et al.* Placental cadherin and the basal epithelial phenotype of BRCA1-related breast cancer. *Clin Cancer Res*, 2005; **11**: 4003-4011.
69. Turner, N., Tutt, A. and Ashworth, A. Hallmarks of "BRCAness" in sporadic cancers. *Nat Rev Cancer*, 2004; **4**: 814-819.
70. Turner, N., Reis-Filho, J.S., Russell, A.M., Springall, R.J., Ryder, K. *et al.* BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene*, 2007; **26**: 2126-2132.
71. Kandel, M.J., Stadler, Z., Masciari, L., Collins, S., Schnitt, S., *et al.* Prevalence of BRCA1 mutations in triple negative breast cancer. *J Clin Oncol, 2006 ASCO Annual Meeting Proceedings Part 1*, 2006; **24**(18S): 508.
72. Abd El-Rehim, D.M., Ball, G., Pinder, S.E., Rakha, E, Paish, C. *et al.* High throughput protein expression analysis using tissue mircoarray technology of a large well-caharacterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analysis. *Int J Cancer*, 2005; **116**: 422-427.
73. Stoppa-Lyonet, D., Ansquer, Y., Dreyfuss, H., Gautier, C., Gauthier-Villars, M. *et al.* Familial invasive breast cancers: worse outcome related to BRCA1 mutations. *J Clin Oncol*, 2000; **18**: 4053-4059.
74 Eerola, H., Vahteristo, P., Sarantaus, L., Kyyronen, P., Pyrhonen, S. *et al.* Survival of breast cancer patients in BRCA1, BRCA2 and non-BRCA1/2 breast cancer families: a relative survival analysis from Finland. *Int J Cancer,* 2001; **93**: 368-372.

75. El-Tamer, M., Russo, D., Troxel, A., Bernadino, L.P., Mazziotta, R. *et al.* Survival and recurrence after breast cancer in BRCA1/2 mutation carriers. *Ann Surg Oncol,* 2004; **11**: 157-164.
76. Brekelman, C.T., Seynauve, C., Menke-Pluymers, M., Bruggenwirth, H.T., Tilanus-Lindthorst, M.M. *et al.* Survival and prognostic factors in BRCA1-associated breast cancer. *Ann Oncol*, 2006; **17**: 391-400.
77. Rennert, G., Bisland-Naggan, S., Barnett-Griness, O., Bar-Joseph, N., Zhang, S. *et al.* Clinical outcomes of breast cancer in carriers of BRCA1 and BRCA2 mutations. *N Engl J Med*, 2007; **357**:115-123.
78 Burke, W., Daly, M., Garber, J., Botkin, J., Kahn, M.J. *et al.* Recommendations for follow-up care of individuals with an inherited predisposition to cancer. II. BRCA1 and BRCA2. Cancer Genetics Studies Consortium. *JAMA*, 1997; **277**: 997-1003.
79. Pichert, G., Bolliger, B., Buser, K., Pagani, O: Swiss Institute for Applied Cancer Research Network for Cancer predisposition Testing and Counselling. Evidence based management options for women at increased breast/ ovarian cancer risk. *Ann Oncol*, 2003; **14**: 9-19.
80. Kriege, M., Brekelmans, C. and Boetes, C. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N Engl J Med*, 2004; **351**: 427-437.
81. Warner, E., Plewes, D., Hill, K.A., Causer, P.A., Zubovits, J.T. *et al.* Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound and mammography, and clinical breast examination. *JAMA*, 2004; **292**:1317-1325.
82. Leach, M.O., Boggis, C.R., Dixon, A.K., Easton, D.F., Eeles, R.A. *et al.* Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). *Lance*t, 2005; **365**:1769-1778.
83. Rebbeck, T.R., Lynch, H.T., Neuhausen, S.L., Narod, S.A., Van't Veer, L.*et al.* Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. *N Engl J Med*, 2002; **346**: 1616-1622.
84. Rebbeck, T.R., Friebel, T., Neuhausen, S.L., van 't Veer, L., Garber, J.E. *et al.* Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers: The PROSE study group. *J Clin Onco*l, 2004; **22**:1055-1062.
85. Narod, S.A., Brunet, J.S., Ghadirian, P., Robson, M., Heimdal, K. *et al.* Tamoxifen and the risk of contralateral breast cancer in BRCA1 and BRCA2 carriers. *Lancet*, 2000; **356**: 1876-1881.
86. Pierce, L.J., Levin, A.M., Rebbeck, T.R., Ben-David, M.A., Friedman, E. *et al.* Ten-year multi-institutional results of breast-conserving surgery and radiotherapy in BRCA1/2-associated stage I/II breast cancer. *J Clin Oncol*, 2006; **24**: 2437-2443.
87. King, M.C., Wieand, S., Hale, K., Lee, M., Walsh, T. *et al.* Tamoxifen and breast cancer incidence among women with inherited mutations in BRCA1 and NRCA2: National Surgical Adjuvant Breast and Bowel Project (NSABP-P1) Breast Cancer Prevention Trial. *JAMA*, 2001; **286**: 2251-2256.

88. Kauf, N.D., Satagopan, J.M., Robson, M.E., Scheur, L., Siegel, B. *et al.* Risk-reducing salpingo-oophorectomy in women with a BRCA1 or BRCA2 mutation. *N Engl J Med*, 2002; **346**:1609-1615.
89. Wapnir, I.L., Anderson, S.J., E.P., M., Geyer, C.E.J., and Jeong, J.H. Prognosis afer ipsilateral breast tumour recurrence and locoregional recurrences in five national surgical adjuvant breast and bowel project node-positive adjuvant breast cancer trials. *J Clin Oncol*, 2006; **24**: 2028-37.
90. Metcalfe, K., Lynch, H.T., Ghadirian, P., Tung, N., Olivotto I. *et al.* Contralateral breast cancer in BRCA1 and BRCA2 mutation carriers. *J Clin Oncol*, 2004; **22**: 2328-2335.
91. Bhattacharyya, A., Ear, U.S., Koller, B.H., Weichselbaum, R.R. and Bishop, D.K. The breast cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. *J Biol Chem*, 2000; **275**: 23899-23903.
92. Moynahan, M.E., Cui, T.Y. and Jasin, M. Homology-directed dna repair, mitomycin-c resistance and chromosome stability is restored with correction of a Brca1 mutation. *Cancer Res*, 2001; **61**: 4842-4850.
93. Quinn, J.E., Kennedy, R.D., Mullan, P.B., Gilmore, P.M., Carty, M. *et al.* BRCA1 functions as a differential modulator of chemotherapy induced apoptosis. *Cancer Res*, 2003; **63**: 6221-6228.
94. De Soto, J.A. and Deng, C.X. Parp-1 inhibitors: are they the long-sought genetically specific drugs for BRCA1/2-associated breast cancers? *Int J Med Sci*, 2006; **3**: 117-23.
95 Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D. *et al.* Specific killing of BRCA2 deficient tumours with inhibitors of poly (ADP-ribose) polymerase. *Nature*, 2005; **434**: 913-7.
96. Farmer, H., McCabe, N. and Lord, C.J. Targeting the DNA repair defect in BRCA1 mutant cells as a therapeutic strategy. *Nature*, 2005; **434**: 917-21.
97. National Human Genome Research Institute Open Access On-Line Breast Cancer Mutation Data Base: http://research.nhgri.nih.gov/projects/bic/index.shtml
98. Mueller, C. and Haworth, A. Draft best practice guidelines for molecular analysis of hereditary breast and ovarian cancer. 2001.
99. Gayther, S.A., Harrington, P., Russell, P., Kharkevitch, G., Garkavtseva, R.F. and Ponder, B.A. Rapid detection of regionally clustered germ-line BRCA1 mutations by multiplex heteroduplex analysis. *Am J Hum Genet*, 1996; **58**: 451-456.
100. Hogervorst, F.B.L., Cornelis, R.S., Bout, M., van Vliet, M., Oosterwijk, J.C. *et al.* Rapid detection of BRCA1 mutations by the protein truncation test. *Nat Genet,* 1995; **10**: 208-212.
101. Opatt, D.M., Morrow, M. and Daly, M. The incidence of BRCA1 and BRCA2 variants of unknown significance varies in different ethinic populations. *J Clin Oncol, 2006 ASCO Annual Meeting Proceedings Part 1*, 2006; **24**(18S): 10002.

102. Peterson, G.M., Parmiagini, G. and Thomas, D. Missense mutations in disease genes: a Bayesian approach to evaluate causality. *Am J Hum Genet*, 1998; **62**:1516-1524
103. Thompson, D., Easton, D.F. and Goldgar, D.E. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet*, 2003; **73**: 652-655.
104. Miller, M.P. and Kumar, S. *Understanding human disease mutations through the use of interspecific* variation. *Hum Mol Genet*, 2001; **10**: 2319-2328.
105. Fleming, M.A., Ostrander, G., Ramirez, C.J., Potter, J., and Ostrander, E.A. Understanding missense mutations in the BRCA1 gene: an evolutionary approach. *Proc Natl Acad Sci*, 2003; **100**:1051-1156.
106. Ramirez, C.J., Fleming, M.A., Potter, J.D., Ostrander, K.O., and Ostrander, E.A. Marsupial BRCA1: conserved regions in mammals and the potential effect of missense changes. *Oncogene*, 2004; **23**:1780-1788.
107. Hurst, L.D. and Pal, C. Evidence for purifying selection acting in silent sites in BRCA1. *Trends in Genetics*, 2001; **17**: 62-65.
108. Huttley, G.A., Easteal, S., Southey, M.C., Tesoriero, A., Giles, G.G. *et al.* Adaptive evolution of the tumour seppressor BRCA1 in humans and chimpanzees. *Nat Genet*, 2000; **25**: 410-413.
109. Abkevitch, V., Zharkikh, A., Deffenbaugh, A.M., Frank, D., Chen, Y. *et al.* Analysis of missense variation in human BRCA1 in the context of interspecific variation. *J Med Genet*, 2004; **41**: 492-507.
110. Williams, R.S., Green, R. and Glover, J.N.M. *Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1*. Nat Struc Biol, 2001; **8**:838-842.
111. Williams, R.S., Chasman, D.I., Hau, D.D., Hui, B., and Lau, A.Y. Detection of protein folding defects caused by BRCA1- BRCAT truncation and missense mutations. *J Biol Chem*, 2003; **278**: 53006-53013.
112. Mirkovic, N., Marti-Renom, M.A., Weber, B.L., Sali, A., and Monteiro, A.N. Structure based assessment of missense mutations in human BRCA1: implications for breast and ovarian cancer disposition. *Cancer Res*, 2004; **64**: 3790-3797.
113. Monteiro, A.N.A., August, A. and Hanafusa, H. Evidence for a transcriptional activation function of BRCA1 C-terminal region. *Proc Natl Acad Sci*, 1996; **93**:13595-13599.
114. Monteiro, A.N.A., August, A. and Hanafusa, H. Common variants of BRCA1 and transcription activation. *Am J Hum Genet*, 1997; **61**: 761-762.
115. Hayes, F., Cayanan, C., Barilla, D. and Monteiro, A.N.A. Functional assay for BRCA1: Mutagenesis of the COOH-terminal region reveals critical residues for transcription activation. *Cancer Res*, 2000; **60**: 2411-2418.
116. Vallon-Christersson, J., Cayanan, C., Haraldsson, K., Loman, N. and Bergthorsson, J.T. Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer. *Hum Mol Genet*, 2001; **10**: 353-360.

117. Phelan, C.M., Dapic, V., Tice, B., Favis, R. and Kwan, E. Classification of BRCA1 missense variants of unknown clinical significance. J Med Genet, 2005; **42**:138-146.
118. Quaresima, B., Faniello, M.C., Baudi, F., Crugliano, T., Cuda, G. *et al.* In vitro analysis of genomic instability triggered by BRCA1 missense mutations. *Hum Mutat*, 2006; **27**: 715-23.
119. Quaresima, B., Faniello, M.C., Baudi, F., Crugliano, T., Di Sanzo M. *et al.* MIssense mutations of BRCA1 gene affect the binding with p53 both in vitro and in vivo. *Oncol Rep*, 2006; **16**: 811-815.
120. Goldgar, D.E., Easton, D.F., Deffenbaugh, A.M., Monteiro, A.N., Tavtigian, S.V. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Gene*t, 2004; **75**: 535-44.
121. Chevenix-Trench, G., Healey, S., Lakhani, S.R., Waring, P., Cummings, M. *et al.* Genetic and histopathological evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown significance. *Cancer Res*, 2006; **66**:2019-27.
122. Osorio, A., Milne, R.L., Honrado, E., Barsoso, A., Diez, O. *et al.* Classification of missense variants of unknown significance in RBCA1 based on clincal and tumour information. *Hum Mutat*, 2007. **28**: 477-485.
123. Liu, H., Cartegni, L., Zhang, M.Q., and Krainer, A.R. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet*, 2001; **27**: 55-58.
124. Burge, C.B., Tuschl, T. and Sharp, P.A. Splicing of precursors to mRNAs by the spliceosome, in *The RNA World*, R.F. Gesteland, T.R. Cech, and J.F. Atkins, Editors. 1999, Cold Spring Harbour Laboratory Press. 525-560.
125. Guth, S. and Valcarcel, J. Kinetic role for mammalian SF1/ BBP in spliceosome assembly and function afer polypyrimidine tract recognition by U2AF. *J Biol Chem*, 2000; **275**: 38059-38066.
126. Liu, Z., Luyten, I. and Bottomley, M.J. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*, 2001; **294**:1098-2102.
127. Brow, D.A. Allosteric cascade of spliceosome activation. *Ann Rev Genet*, 2002; **3**: 333-360.
128. Kielkopf, C.L., Luck, S. and Green, M.R. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev*, 2004. **18**:1513-1526.
129. Cartegni, L., Chew, S.L. and Krainer, A.R. Listening to silence and understanding nonsense. Exonic mutations that affect splicing. *Nature,* 2002; **3**: 285-298.
130. Nielsen, T.W. The case for an RNA enzyme. *Nature*, 2000; **408**: 782-783.
131. Graveley, B.R. Sorting out the complexity of SR protein functions. *RNA*, 2000; **6**: 1197-1211.
132. Birney, E., Kumar, S. and Krainer, A.R. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nuc Acid Res*, 1993; **21**: 5803-5816.
133. Blencowe, B.J. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochemical Sci*, 2000; **25**: 106-110.

134. Zuo, P. and Maniatis, T. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer dependent splicing. *Genes Dev*, 1996; **10**: 1356-1368.
135. Kan, J.L. and Green, M.R. Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. *Genes Dev*, 1999; **13**: 462-471.
136. Liu, H.X., Zhang, M.Q. and Krainer, A.R. Identification of functional exonic splicing enhancer motifs recognised by individual SR proteins. *Genes Dev,* 1998, **12**: 1998-2012.
137. Schall, T.D. and Maniatis, T. Multiple distinct splicing enhancers in the protein-coding sequencers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol Cell Biol*, 1999; **19**: 261-273.
138. Lopez, A.J. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Ann Rev Genet*, 1998. **32**: 279-305.
139. Smith, C.W. and Valcarcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, 2000; **25**: 349-404.
140. Wagner, E.J. and Garcia-Blanco, M.A. Polypyrimidine tract binding protein antagonises exon definition. *Mol Cell Biol*, 2001; **21**: 3281-3288.
141. Krecic, A.M. and Swanson, M.S. hnRNPs complexes: composition, structure and function. *Curr Opin Cell Biol*, 1999; **11**: 363-371.
142. Black, D.L. Protein diversity from alternative splicing. Cell, 2000; **103**: 367-370.
143. Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J. *et al.* Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, 2000; **101**: 671-684.
144. Croft, L., Schandorff, S., Clark, F., Burrage, K., Arctander, P., and Mattick, J.S. ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nat Genet,* 2000; **24**: 340-341.
145. Modrek, B. and Lee, C. A genomic view of alternative splicing. *Nat Genet,* 2002; **30**:13-19.
146. Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., and Loerch, P.M. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, 2003; **19**: 2141-2144.
147. International human genome sequencing consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 2004; **431**: 931-945.
148. Kralovicova, J., Houngninou-Molango, S., Kramer, A., and Vorechovsky, I. Branch site haplotypes that control alternative splicing. *Hum Mol Genet*, 2004; **13**: 3189-3202.
149. Orban, T.I. and Olah, E. Emerging roles of BRCA1 alternative splicing. *Mol Path*, 2003; **56**: 191-197.
150. Lu, M. Transactivation of the p21 promoter by BRCA1 splice variants in mammary eipthelial cells: evidence for both common and distinct activities of wildtype and mutant forms. *Oncogene*, 2000; **19**: 6351-60.
151. Wang, H., Shao, N., Ding, Q.M., Cui, J., Reddy, E.S. and Rao, V.N. BRCA1 proteins are transported to the nucleus in the absence of serum and splice

variants BRCA1a, BRCA1b are tyrosine phosphoproteins that associate with E2F, cyclins and cyclin dependent kinases. *Oncogene,* 1997; **15**:143-57.
152. ElShamy, W.M. and Livingston, D.M. Identification of BRCA1-IRIS, A BRCA1 locus product. *Nat Cell Biol,* 2004; **6**: 954-967.
153. Ashworth, A. Refocusing on BRCA1. *Nat Cell Biol*, 2004; **6**: 916-17.
154. Huber, L.J., Yang, T.W., Sarkison, C.J., Master, S.R.Deng, C.X. and Chodosh, L.A. Impaired DNA damage response in cells expressing an exon 11 deleted murine Brca1 variant that localises to nuclear foci. *Mol Cell Biol,* 2001; **21**: 4005-15.
155. Bachelier, R., Xu, X., Wang, X., Li, W., Naramura, M. *et al.* Normal lymphocyte development and thymic lymphoma development in Brca1 exon 11-deficient mice. *Oncogene*, 2003; **22**: 528-37.
156. Krawczak, M., Reiss, J. and Cooper, D.N. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet*, 1992; **90**: 41-54.
157. Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., and Shiel, J.A. Human gene mutation database (HGMD): 2003 update. *Hum Mutat,* 2003; **21**: 577-581.
158. Teraoka, S.N., Telatar, M., Becker-Catania, S., Liang, T., Onengut, S. *et al.* Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *Am J Hum Gene*t, 1999; **64**:1617-31.
159. Pagani, F., Raponi, M. and Baralle, F.E. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *PNAS,* 2005; **102**: 6368-6372.
160. Wang, Z., Lo, H.S., Gere, S., Hu, Y., Buetow, K.H., and Lee, M.P. Computational analysis and experimental validation of tumour-associated alternative RNA splicing in human cancer. *Cancer Res*, 2003; **63**: 655-657.
161. Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigo, R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters*, 2005; **579**:1900-1903.
162. Cargill, M. Characterisation of single-nucleotide polymorphism in coding regions of human genes. *Nat Genet*, 1999; **22**: 231-238.
163. Cartegni, L., Wang, J., Zhu, Z., Zhang, M.O. and Krainer, A.R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nuc Acid Res*, 2003; **31**: 3568-3569.
164. Fairbrother, W.G., Yeh, R.F., Sharp, P.A. and Burge, C.B. Rescue-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nuc Acid Res*, 2002; **32**: W187-190.
165. Pettigrew, C., Wayte, N., Lovelock, P.K., Tavtigian, S.V., Chenevix-Trench, G. *et al.* Evolutionary conservation analysis increases the colocalisation of predicted exonic splicing enhancers in the BRCA1 gene with missense sequence changes and in-frame deletions, but not polymorphisms. *Breast Cancer Res*, 2005; **7**: R929-R939.
166. Shapiro, M.B. and Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nuc Acid Res*, 1987; **15**: 7155-7174.

167. Brose, M.S., Volpe, P., Paul, K., Stopfer, J.E., Colligon, T.A. *et al.* Characterisation of two novel BRCA1 germ-line mutations involving splice donor sites. *Genet Test*, 2004; **8**: 133-8.
168. Campos, B., Diez, O., Domenech, M., Baena, M., Balman, J. *et al.* RNA analysis of 8 BRCA1 and BRCA2 unclassified variants identified in breast/ ovarian cancer families from Spain. *Hum Mutat*, 2003; **22**: 337.
169. Tesoriero, A., Wong, E.M., Jenkins, M.A., Hopper, J.L., Brown, M.A. *et al.* Molecular characterisation and cancer risk associated with BRCA1 and BRCA2 splice site variants identified in multiple case breast cancer families. *Hum Mutat*, 2005; 26: 495.
170. Keaton, J.C., Nielson, D.R., Hendrickson, B.C., Pyne, M.T., Scheur, L. *et al.* A biochemical analysis demonstrates that the BRCA1 intronic variant IVS10-2 A>C is a mutation. *J Hum Genet*, 2003; **48**: 399-403.
171. Sharp, A., Pichert, G., Lucassen, A., and Eccles, D. RNA analysis reveals splicing mutations and loss of expression defects in MLH1 and BRCA1. *Hum Mutat*, 2005; **24**: 272.
172. Yang, Y., Swamianathan, S., Martin, B.K., and Sharan, S.K. Aberrant splicing induced by missense mutations in BRCA1: clues from a humanised mouse model. *Hum Mol Genet*, 2003; **12**: 2121-2131.
173. McVety, S., Li, L., Gordon, P.H., Chong, G., and Foulkes, W.D. Disruption of an exon splicing enhancer in exon 3 of MLH1 is the cause of HNPCC in a Quebec family. *J Med Genet*, 2006; **43**: 153-156.
174. Auclair, J., Busine, M.P., Navarro, C., Ruano, E., Montmain, G. *et al.* Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. *Hum Mutat*, 2006; **27**: 145-154.
175. Lastella, P., Resta, N., Miccolis, I., Quagliarella, A., Guanti, G. and Stella, A. Site directed mutagenesis of hMLH1 exonic splicing enhancers does not correlate with splicing disruption. *J Med Genet*, 2004; **41**: 72-78.
176. Hoffman, W., Horn, D., Huttner, C., Classen, E., and Scherneck, S. The BRCA2 variant 8204G>A is a splicing mutation and results in an in frame deletion of the gene. *J Med Genet*, 2003; **40**: e23.
177. Hentze, M.W. and Kulozik, A.E. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell*, 1999; **96**: 307-310.
178. Mendell, J.T. and Dietz, H.C. When the message goes awry: disease-producing mutations that influence mRNA content and performance. Cell, 2001; **107**: 411-414.
179. Perrin-Vidoz, L., Sinilnikova, O.M., Stoppa Lyonnet, D., Lenoir, G.M., and Mazoyer, S. The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. *Hum Mol Genet*, 2002; **11**: 2805-2814.
180. Tournier, I., Raux, G., Di Fiore, F., Marechal, I., Leclerc, C. *et al.* Analysis of the allele-specific expression of the mismatch repair gene MLH1 using a simple DHPLC-based method. *Hum Mutat*, 2004; **23**: 379-384.
181. Haber, D.A. Breast cancer in carriers of BRCA1 and BRCA2 mutations: tackling a molecular and clinical conundrum. *J Clin Oncol*, 1999; **17**: 3367-3370.

182. Crook, T., Crossland, S., Crompton, M.R., Osin, P. and Gusterson, B.A. p53 mutations in BRCA1-associated familial breast cancer. *Lancet,* 1997; **350**: 638-639.
183. Cressman, V.L., Backlund, D.C., Hicks, E.M., Gowen, L.C., Godfrey, V., and Koller, B.H. Mammary tumour formation in p53-and BRCA1 deficient mice. *Cell Growth Differ*, 1999; **10**: 1-10.
184. Phillips, K.A. Immunophenotypic and pathologic differences between BRCA1 and BRCA2 hereditary breast cancers. *J Clin Oncol*, 1999; **18**(21s): 107s-112s.
185. Hakem, R., de la Pompa, J.L., Elia, A., Potter, J., and Mak, T.W. Partial rescue of Brca1 early embryonic lethality by p53 or p21 null mutation. *Nat Genet*, 1997; **16**: 298-302.
186. Ludwig, T., Chapman, D.L., Papioannou, V.E., and Efstratiadis, A. Targeted mutations of breast cancer susceptilbility gene homologs in mice: lethal phenotypes of BRCA1, Brca2, Brca1/Brca2, Brca1/p53 nullizygous embryos. *Genes Dev*, 1997; **11**: 1226-41.
187. Mak, T.W., Hakem, A., McPherson, J.P., Shehabeldin, A., Zablocki, E. *et al.* Brcal required for T cell lineage development but not TCR loci rearrangement. *Nature Immunol*, 2000; **1**: 15-17.
188. McPherson, J.P., Lemmers, B., Hirao, A., Hakem, A., and Abraham, J. Collaboration of Brca1 and Chk2 in tumourigenesis. *Genes Dev*, 2004. **18**: 1144-1153.
189. Oliner, J.D., Kinzler, K.W., Meltzer, P.S., George, D.L., and Vogelstein, B. Amplification of a gene encoding a p53-associated protein in human sarcomas. *Nature*, 1992; **358**:15-6.
190. Lundgren, K., Montes de Oca Luna, R., McNeill, Y.B., Emerick, E.P., Spencer, B. *et al.* Targeted expression of MDM2 uncouples S phase from mitosis and inhibits mammary gland developement independent of p53. *Genes Dev*, 1997; **11**: 714-25.
191. Alt, J.R., Bouska, A., Fernandez, M.R., Cerny, R.L., Xiao, H., and Eischen, C.M. MDM2 binds to nbs1 at sites of DNA damage and regulates double strand break repair. *J Biol Chem*, 2005; **280**: 18771-18881.
192. Buesmo-Ramos, C.E., Manshouri, T., Haidar, M.A., Yang, Y., McCown, P. *et al.* Abnormal expression of MDM-2 in breast carcinomas. *Breast Cancer Res Treat*, 1996; **37**: 179-88.
193. Freedman, D.A. and Levine, A.J. Regulation of the p53 protein by the MDM2 oncoprotein - thirty-eighth G.H.A. Clowes Memorial Award Lecture. *Cancer Res*, 1999; **59**:1-7.
194 Bond, G.L., Hu, W., Bond, E.E., Lutzker, S.G., Arva, N.C. *et al.* A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumour suppressor pathway and accelerates tumour formation in humans. *Cell*, 2004; **119**: 591-602.
195. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R. *et al.* Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 1988; **239**: 487-491.

196. Montagna, M., Agata, S., De Nicolo, A., Menin, C., Sordi, G. *et al.* Identification of BRCA1 and BRCA2 carriers by allele-specific gene expression (AGE) analysis. *Int J Cancer*, 2002; **98**: 732-736.
197. Ronaghi, M., Uhlen, M. and Nyren, P. A sequencing method based on real-time pyrophosphate. *Science*, 1998; **281**: 363-365.
198. Lavebratt, C., Sengul, S., Jansson, M. and Schalling, M. Pyrosequencing based SNP allele frequency estimation in DNA pools. *Hum Mutat*, 2004; **23**: 92-97.
199. Langaee, T. and Ronaghi, M. Genetic variation analyses by Pyrosequencing. *Mutat Res*, 2005; **573**: 96-102.
200. Geisler, J.P., Hatterman-Zogg, M.A., Rathe, J.A. and Buller, R.E. Frequency of BRCA1 dysfunction in ovarian cancer. *J Natl Cancer Inst*, 2002; **94**: 61-67.
201. Futreal, P.A., Soderkvist, P., Marks, J.R., Iglehart, J.D., Cochran, C. *et al.* Detection of frequent allelic loss on proximal chromosome 17q in sporadic breast carcinoma using microsatellite length polymorphisms. *Cancer Res*, 1992; **52**: 2624-7.
202. Bonatti, F., Pepe, C., Tancredi, M., Lombardi, G., Aretini, P. *et al.* RNA-based analysis of BRCA1 and BRCA2 gene alterations. *Cancer Genet Cytogenet*, 2006; **170**: 93-101.
203. Gangully, A., Rock, M.J. and Prockop, D.J. Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent induced bends in DNA heteroduplexes. *Proc Natl Acad Sci USA*, 1993; **90**: 10325-10329.
204. Esteban-Cardenosa, E., Duran, M., Infante, M., Velasco, E., and Miner, C. High-throughput mutation detection method to scan BRCA1 and BRCA2 based on heteroduplex analysis by capillary array electrophoresis. *Clin Chem*, 2004; **50**: 313-320.
205. Evans, D.G.R., Eccles, D.M., Rahman, N., Young, K., Bulman, M., and Amir, E. A new scoring system for the chances of identifying a BRCA1/2 mutation outperforms existing models including BRCAPRO. *J Med Genet*, 2004; **41**: 474-480.
206. Xu, C., Chambers, J.A., Nicolai, H., Brown, M.A., Hujeirat, Y. *et al.* Mutations and alternative splicing of the BRCA1 gene in breast/ ovarian cancer families. *Genes Chromosomes Cancer*, 1997; **18**: 102-110.
207. Claes, K., Vandesompele, J., Poppe, B., Dahan, K., Coene, I. *et al.* Pathological splice mutations outside the invariant AG/GT splice sites of BRCA1 exon 5 increase alternative transcript levels in the 5' end of the BRCA1 gene. *Oncogene*, 2002; **21**: 4171-4175.
208. Velasco, E., Sanz, D.J., Perez-Cabornero, L., Duran, M., Infante, M. *et al.* Aberrant splicing of tumour suppressor genes BRCA1 and BRCA2 in familial breast cancer. *Poster, "Pre mRNA splicing and disease" Conference, Cortino d'Ampezzo,* Italy, 2007.

209. Claes, K., Poppe, B., Machackova, E., Coene, I., Foretova, L. *et al.* Differentiating pathogenic mutations from poymorphic alterations in the splice sites of BRCA1 and BRCA2. *Genes Chromosomes Cancer*, 2003; **37**: 314-320.
210. Fortin, J., Moisan, A., Dumont, M., Leblanc, G., and Labrie, Y. A new alternative splice variant of BRCA1 containing an additional in-frame exon. *Biochimica et Biophysica Acta*, 2005; **1731**: 57-65.
211. Cooper, T.A. Use of minigene systems to dissect alternative splicing elements. *Methods*, 2005; **37**: 331-340.
212. Savkur, R.S., Phillips, A.V. and Cooper, T.A. Aberrant regulation of insulin receptor splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet*, 2001; **29**: 40-7.
213. Dredge, B.K. and Darnell, R.B. Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. *Mol Cell Biol*, 2003; **23**: 4687-700.
214. Fraser, C.G., *Biological Variation: from principles to practice*. 2001: AACCPress.
215. Majewski, J. and Ott, J. Distribution and characterisation of regulatory elements in the human genome. *Genome Research*, 2002; **12**: 1827-1836.
216. Kawamoto, S. Neuron-specific alternative splicing of nonmuscle myosin II heavy chain-B pre-mRNA requires a cis-acting intron sequence. *J Biol Chem*, 1996; **271**: 7613-7616.
217. Ruffner, H., Joazeiro, C.A., Hemmati, D., Hunter, T., and Verma, I.M. Cancer-predisposing mutation within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. *PNAS,* 2001; **98**: 5134.
218. Vega, A., Campos, B., Bressac de Paillerets, B., Bond, P.M., Janin, N. *et al.* The R71G BRCA1 is a founder Spanish mutation and leads to aberrant splicing of the transcript. *Hum Mutat*, 2001; **17**: 520-521.
219. Tavtigian, S.V., Deffenbaugh, A.M., Yin, L., Judkins, T., Scholl, T. *et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet,* 2005; **43**: 295-305.
220. Lovelock, P.K., Healey, S., Au, W., Sum, E.Y.M., Tesoriero, A. *et al.* Genetic, functional and histopathological evalutation of 2 C-terminal BRCA1 missense variants. *J Med Genet*, 2006. **43**: 74-83.
221. Yeo, G., Holste, D., kreiman, G., and Burge, C.B. Variation in alternative splicing across human tissues. *Genome Biology*, 2004; **5**: R54-75.
222. Galante, P.A., Sakabe, N.J., Kirschbaum-Slager, N., and de Souza, S.J. Detection and evaluation of intron retention eventsin the human transcriptome. *RNA*, 2004; **10**: 757-765.
223. Le Hir, H., Charlet-Berguerand, N., de Franciscis, V., and Thermes, C. *5'* End-RET splicing: absence of variants in normal tissues and intron retention in phaeochromocytomas. *Oncology*, 2002; **63**: 84-91.
224. Carmel, I., Tal, S., Vig, I., and Ast, G. Comparative analysis detects dependencies among the 5' splice-site positions. *RNA*, 2004; **10**: 828-40.

225. Buratti, E., Baralle, M. and Baralle, E. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nuc Acid Res,* 2006; **34**: 3493-3510.
226. Kreahling, J. and Graveley, B.R. The origins and implications of Aluternative splicing. *Trends Genet*, 2004, **20**:1-4.
227. Balz, V., Prisack, H.B., Bier, H., and Bojar, H. Analysis of BRCA1, TP53 and TSG101 germline mutations in German breast and/ or ovarian cancer families. *Cancer Genet Cytogenet*, 2002; **138**:120-127.
228. Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. dbEST-database for "expressed sequence tags". *Nat Genet*, 1993; **4**: 332-333.
229. Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. Systematic identification and analysis of exonic splicing silencers. *Cell,* 2004; **119**: 831-845.
230. Yeo, G. and Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comp Biol*, 2004; **11**: 377-394.
231. Pagani, F., Stuani, C., Tzetis, M., kanavakis, E., Efthymiadou, A. *et al.* New type of disease casuing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet*, 2003; **12**: 1111-1120.
232. Wiliams, R.S. and Glover, J.N. Structural consequences of a cancer-causing BRCA1 BRCT missense mutation. *J Biol Chem*, 2003; **278**: 2630-5.
233. Hellman, I., Zollner, S., Enard, W., Ebersberger, I., Nickel, B., and Paabo, S. Selection on human genes as revealed by comparison to chimpanzee cDNA. *Genome Res*, 2003; **13**: 831-837.
234. Bustamante, C.D., Nielsen, R. and Hartl, D.L. A maximum likelihood method for analysing pseudogene evolution: implications for silent site evolution in humans and rodents. *Mol Biol Evol*, 2002; **19**: 110-117.
235. Rajeskaran, R., Sudandiradoss, C., Doss, C.G., and Sethumadhavan, R. Identification and in silico analysis of functional SNPs of the BRCA1 gene. *Genomics*, 2007; **90**: 447-52.
236. Messiaen, L.M., Callens, T., Mortier, G., Betsen, D., Vandenbroucke, I. *et al.* Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum Mutat*, 2000; **15**: 541-555.
237. Thomson, S.A., Fishbein, L. and Wallace, M.R. NF1 mutations and molecular testing. *J Child Neurol*, 2002; **17**: 541-555.
238 Blencowe, B.J. Alternative splicing: newinsights from global analyses. *Cell,* 2006; **126**: 37-47.
239. Bougeard, G., Baert-Desurmont, S., Tournier, I., Vasseur, S., Martin, C. *et al.* Impact of the MDM2 SNP309 and p53 Arg72Pro polymorphism on age of tumour onset in Li-Fraumeni syndrome. *J Med Gene*t, 2006; **43**: 531-3.
240. Boersma, B.J., Howe, T.M., Goodman, J.E., Yfantis, H.G., Lee, D.H. *et al.* Association of breast cancer outcome with status of p53 and MDM2 SNP309. *J Natl Cancer Inst*, 2006; **9813**: 911-19.

241. Campbell, I.G., Eccles, D.M. and Choong, D.Y.H. No association of the MDM2 SNP309 polymorphism with risk of breast or ovarian cancer. *Cancer Letts*, 2005; **240**: 195-197.
242. Wasielewski, M., Nagel, J.H., Brekelmans, C., Klijn, J.G., van den Ouweland, A.M.W. *et al.* MDM2 SNP309 accelerates familial breast carcinogenesis independently of oestrogen signalling. *Breast Cancer Res Treat*, 2006; **104:** 153-157.
243. Foulkes, W.D., Stamp, G.W., Afzal, S., Ialani, N., McFarlane, C.P. *et al.* MDM2 overexpression is rare in ovarian carcinoma irrespective of TP53 mutation status. *Br J Cancer*, 1995; **72**: 883-8.
244. Ramus, S.J., Bobrow, L.G., Pharoah, P.D., Finnigan, D.S., Fishman, A. Increased frequency of TP 53 *mutations in BRCA1 and BRCA2 ovarian tumours. Genes Chromosomes Cancer,* 1999; **25**: 91-96.
245. Menin, C., Scaini, M.C., De Salvo, G.L., Biscuola, M., Quaggio, M. *et al.* Association between MDM2-SNP309 and age at colorectal cancer diagnosis according to p53 mutation status. *J Natl Cancer Inst*, 2006; **98**: 285-88.
246. Wilkening, S., Bermejo, J.L., Burwinkel, B., Klaes, R., Bartram, C.R. *et al.* The single nucleotide polymorphism IVS+309 in Mouse Double Minute 2 does not affect risk of familial breast cancer. *Cancer Res,* 2006. **66**: 646-48.
247. Petenkaya, A., Bozkurt, B., Akilli-Ozturk, O., Kaya, H.S., Gur-Dedeoglu, B., and Yulug, I.G. Lack of association between the MDM2-SNP309 polymorphism and breast cancer risk. *Anticancer Res*, 2006; **26**: 4975-77.
248. Alhopuro, P., Ylisaukko-oja, S.K., Koskinen, W.J., Bono, P., Arola, J. *et al.* The MDM2 promoter polymorphism SNP309T>G and the risk of uterine leiomyosarcoma, colorectal cancer and squamous cell carcinoma of the head and neck. *J Med Genet*, 2005; **42**: 694-698.
249. Hong, Y., Miao, X., Zhang, X., Ding, F., Luo, A. *et al.* The role of p53 and MDM2 polymorphisms in the risk of oesphageal cell carcinoma. *Cancer Res,* 2005: **65**: 9582-87.
250. Zhang, X., Miao, X., Guo, Y., Tan, W., Zhou, Y. *et al.* Genetic polymorphisms in cell cycle regulatory genes MDM2 and TP53 are associated with susceptibility to lung cancer. *Hum Mutat*, 2006. **27**:110-7.
251. Ohmiya, N., Taguchi, A., Mabuchi, N., Itoh, A., and Hirooka, Y. MDM2 promoter polymorphism is associated with both an increased susceptibility to gastric carcinoma and poor prognosis. *J Clin Oncol*, 2006; **24**: 4434-40.
252. Bond, G.L., Hu, W. and Levine, A. A single nucleotide polymorphism in the MDM2 gene: from a molecular and cellular explanation to clinical effect. *Cancer Res*, 2005; **65**: 5481-5484.
253. Kinyamu, H.K. and Archer, T.K. Estrogen receptor-dependent proteasomal degradation of the glucocorticoid receptor is coupled to an increase in mdm2 protein expression. *Mol Cell Biol*, 2003; **23**: 5867-81.
254. Copson, E.R., White, H.E., Blaydes, J.P., Robinson, D.O., Johnson, P.W., and Eccles, D.M. Influence of the MDM2 single nucleotide polymorphism SNP309 on tumour development in BRCA1 mutation carriers. BMC Cancer, 2006. **6**: 80-87.

255. Stevens, C., Lin, Y., Sanchez, M., Amin, E., Copson, E. *et al.* A germ line mutation in the death domain of DAPK-1 inactivates ERK-induced apoptosis. *J Biol Chem*, 2007; **282**: 13791-13803.
256. Frank, T.S., Deffenbaugh, A.M., Reid, J.E., Hulick, M. *et al.* Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2 : analysis of 10,000 individuals. *J Clin Oncol,* 2002; **20**: 1480-1490.
257. Mattocks, C., Ward, D., Silibourne, J., Herbert, T., Owen, N. *et al.* Implementation of the SCOBEC high throughput screening facility. *J Med Genet*, 2006. **42** Suppl 1: p. S90.
258. Mattocks, C., Ward, D., Harvey, J.F., and Cross, N.C.P. Design and optimisation of a validated primer set for automated screening of the BRCA1 and BRCA2 genes. *J Med Genet*, 2006. **43** Suppl 1: p. S73.
259. Lixia, M., Ahijian, C., Chao, S., Chaojiang, G., and Congyi, Z. Alternative splicing of breast cancer associated gene BRCA1 from breast cancer cell line. *J Biochem Mol Biol*, 2007; **40**: 15-21.
260. De Brakeleer, S., Bogdani, M., De Greve, J., Decock, J., Sermijn, E. *et al.* Loss of nuclear BRCA1 protein staining in normal tissue cells derived from BRCA1 and BRCA2 mutation carriers. *Mutat Res*, 2007; **619**: 104-112.