DEVELOPMENT OF NEW APPROACHES TO MUTATION DETECTION IN HEREDITARY BREAST CANCER

Mohammed Abdullah A. Aldahmesh

Ph.D.

March 2004

DEVELOPMENT OF NEW APPROACHES TO MUTATION DETECTION IN HEREDITARY BREAST CANCER

Mohammed Abdullah A. Aldahmesh

BSc in Biochemistry
MSc in Medical Genetics

This thesis submitted for

Doctor of Philosophy to

Human genetics division

Faculty of medicine, health and life sciences

March 2004


إهداء


أهدي هذا المجهود إلى


والدتي الحبيبة


حفظها الله


وذكرى والدي برحمه الله

Dedication

I dedicate this entire work to
My Mother, God bless her,
And the happy memory of my
father

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

HUMAN GENETICS RESEARCH DIVISION

Doctor of Philosophy

DEVELOPMENT OF NEW APPROACHES TO MUTATION DETECTION IN HEREDITARY BREAST CANCER

By Mohammed Abdullah Aldahmesh

Mutation analysis of complex genes without individual hotspots for sequence variations, such as *BRCA1*, is time-consuming and expensive. *BRCA1* mutation status has important implications for close relatives and for the individual in terms of future cancer risks but only those with a striking family history can normally access genetic mutation analysis. *BRCA1* spans over 80kb of genomic DNA and comprises 24 exons with pathogenic mutations scattered throughout. Of all currently available methods, direct sequencing has the highest sensitivity, but also the highest cost. Other techniques such as SSCP, DGGE and PTT are more economical but have at best a sensitivity of 90%. Mutation detection sensitivity depends on the method employed and the type of mutation present. Melt-MADGE uses a 96-well plate system for loading DNA samples and a combination of temperature ramp and denaturant to resolve both heteroduplexes and mutant homoduplex bands from PCR amplified alleles. Specifically for examining many subjects in parallel, costs will compare very favourably against SSCP, DGGE, dHPLC and direct sequencing.

At the time of developing the technique I have designed an assay for mutation scanning for individual amplimers. I had already optimised long and nested PCRs and fine-tuned melt-MADGE assay for 39 amplimers required for the entire coding region of the *BRCA1* gene. I have produced high quality results confirming detection of all the common polymorphisms in the largest exon of the gene (exon 11, 3426bp) as well as in other polymorphic exons. In addition, I have discovered a novel polymorphism.
In parallel, I have set out to confirm the detection of mutations by an independent mutation detection method, ARMS assay. I have developed ARMS assay for five SNPs in the *BRCA1* gene to distinguish the polymorphic samples obtained from the mutated samples. In a pilot study, a panel of 100 anonymous samples from the Wessex Regional Genetics Laboratory (WRGL), were examined. The samples were screened over the entire coding region and we were able to detect several mutations, subsequently confirmed by direct sequencing. Matching the results from the diagnostic laboratory, we achieved 93% sensitivity for mutation detection. Further work is now being directed towards determining the reasons for the reduced sensitivity of the assay in specific cases.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Statement of originality

This thesis is the original work of the author. Up to date part of it has been published as an abstract and a full paper (online Technical Tips Online). The work has not been submitted to any other University except to the University of Southampton for the award of the doctorate of philosophy.

Acknowledgements

I am deeply indebted to Professor Ian N.M. Day and Dr. Diana M. Eccles who supervised this work and unsparingly engaged themselves in bringing the present work to this stage and gave me the guidance and help I needed on a daily basis.

My utmost gratitude goes to my previous supervisor Dr. Emmanuel Spanakis, who wholeheartedly gave me much necessary advice and shared with me his experience at the early stage of my project.

I am also grateful to Lesley Hinks, Tricia Briggs, Diane Brown, Carolyn Wallis and Kim Davis for their assistance in the Division during my project without whose precious assistance in the Division this work would have suffered much.

I should also express my thanks to Julie Sillibourne in the Wessex Regional Genetics Laboratory at Salisbury for her co-operation with me during the validation of my method.

I also thank the Government of Saudi Arabia for their financial support in my study. Also, my thanks go to Breast Cancer Campaign for their financial support in the lab work.

I wish to express sincere thanks and deepest gratitude to my mother, brothers and sisters and my wife, for their emotional and practical support and encouraging me with all means throughout my study.

I am also thankful to all my colleagues in the Division for their support, advice and their friendship.

Definitions and Abbreviations

Definitions

Association study: is the study to test whether a genetic marker (polymorphism) occurs more frequently in cases than in controls.

Dominant-negative: Mutation that dominantly affects the phenotype by means of a defective protein or RNA molecule that interferes with the function of the normal gene product in the same cell.

Founder mutation: A high frequency of a mutant gene in a population founded by a small ancestral group when one or more of the founders was a carrier of the mutant gene.

GC-clamp: an artificial GC-sequence introduced to one end of a DNA fragment to stabilise and prevent strand dissociation.

Germline mutations: A heritable genetic change in the germ line cells

Haplotype: A set of genes on a single chromosome.

Hardy-Weinberg equilibrium: A mathematical statement of the concept that gene and genotype frequencies are constant from generation to generation in a large, interbreeding population with random mating in the absence of selection, mutation, or migration.

Heteroduplex: A double-stranded DNA in which the base sequence not completely complementary due to mutation. These heteroduplexes will then be of two types and will be complementary; one could contain T.G mismatch and the other C.A mismatch.

Homoduplex: A hybrid molecule between exactly complementary DNA strands with no mismatches.

Melt-MADGE: is the generic name for a series of high-throughput methods of *de novo* scanning for single-nucleotide variation, both base substitutions and frameshift mutations.

Population study: is the study of the genetic composition and inter-relationships within a population.

SNPs: are single base pair positions in genomic DNA at which different sequence exist in normal individuals in some populations and the frequency of the allele is more than 1%.

Somatic mutations: A non-heritable genetic change occurring within a somatic cell, also known as an acquired mutation.

Splice site: The base sequence at each end of an intron that determines the splice point. The site at the 5' end of the intron is the donor site and the site at the 3' end is the acceptor site.

Sporadic case: A case of a disease with no family history of the disease.

The Breast Cancer Linkage Consortium (BCLC): is a worldwide cooperative network of scientists who share a major interest in inherited breast and ovarian cancer

TIXIS: A program developed by Emmanuel Spanakis, University of Southampton. This software is a modification of the MELT 87 algorithm that was previously devised by Leonard Lerman. TIXIS calculates the theoretical melting profile of a known DNA sequence.

Abbreviations

ARMS	Amplification Refractory Mutation System
BARD1	*BRCA1*-association RING domain protein 1
BRCA1	breast cancer susceptibility gene1
BRCA2	breast cancer susceptibility gene2
BRCT	*BRCA1* carboxy-terminal repeat
CDGE	Constant Denaturant Gel Electrophoresis
DBD	DNA Binding Domain
DGGE	Denaturing Gradient Gel Electrophoresis
dHPLC	Denaturing high-performance Liquid Chromatography
dNTPs	DeoxyNucleotide TriPhosphates
FAMA	Fluorescence-Assisted Mismatch Analysis
H A	Heteroduplex analysis
Long PCR	Long-distance Polymerase chain reaction
MADGE	Microplate-Array-Diagonal-Gel Electrophoresis
Melt-MADGE	Melting Point Analysis by Microplate-Array Diagonal Gel Electrophoresis
PCNA	Proliferation Cell Nuclear Antigen
PCR	Polymerase chain reaction
PTT	Protein Truncated Test
TGGE	Temperature Gradient Gel Electrophoresis
TTGE	Temporal Temperature Gradient Electrophoresis
SNP	Single Nucleotide Ploymorphism
SSCP	Single Strand Conformation Polymorphism

CHAPTER ONE

Research framework and Introduction

Research framework

a) The study area

It is well known that a mutation is any permanent alteration in DNA. In general terms, mutations could be classified into three categories; genomic mutations (e.g. aneuploidy), chromosome mutations (e.g. translocations) and gene mutations (e.g. point mutations). In our study, I have focused on gene mutations, which could occur in either germline cells or somatic cells. The origin of a mutation varies between classes of mutation. However, point mutations could originate by either of two basic mechanisms; errors introduced during the normal process of DNA replication or base changes introduced by mutagens. The effect of a mutation is not always apparent. A single base change may result in the production of nonsense mutation causing termination of the translation of protein and leading to disease. In addition, a single base alteration might result in a silent mutation where the amino acid of the protein remains unchanged (Spanakis and Day 1997). Mutation detection is in high demand in both population studies and clinical diagnostics. Understanding the effect of alterations on the normal biological process can be challenging. Thus, a method with high sensitivity, accuracy, rapid throughput and cost-effectiveness that is simple to set up and operate is the method of choice in the mutation detection field (Cotton 1997).

b) Methodology

Since the cloning of the *BRCA1* gene (Miki et al. 1994) detection of mutations in *BRCA1* has remained a technical challenge. More than 250 mutations have been documented in the Breast Cancer Information Core (BIC). The two most common mutations, the 185delAG and the 5382insC are located at opposite ends of the *BRCA1* gene. A third mutation, 4184del4 in codon 1382 is also seen repeatedly. The large majority of mutations that cause premature truncation of the peptide by frameshift sequence alterations tend to cause a premature stop codon downstream. The nonsense mutations result in stop codons at the site of the single base pair substitution. In addition, mutations outside of the coding region, e.g. regulatory and splice site regions, lead to loss of expression or abnormal translation of the gene. As a result of the discovery of increasing numbers of genes associated with specific diseases, it is increasingly

important to develop simple, low-cost, reliable and high throughput methods to detect sequence variations in DNA sequence. Detection of mutations that are clinically significant requires a simple, sensitive, and cost-effective assay. Technical advances along these lines will allow the development of diagnostic tools based on the variations in DNA sequence. However, a technique such as melt-MADGE will promote mutation detection in diagnostic laboratories and allow more comprehensive population research studies.

c) The benefits of the study

- Developing a highly sensitive, cost effective screening method for *BRCA1*gene mutation.
- Time and cost of screening for point mutations will be reduced in the evaluation of high-risk families.

Introduction

Breast cancer (BC) is the most common malignancy found in women. This disease has received a lot of attention in the literature because of its high prevalence in western countries and the United States (1 in 8 in USA and 1 in 12 lifetime risk in the United Kingdom), and its relatively high frequency worldwide (Offit 1998). However, despite extensive research the genetic component of breast cancer remains to be fully elucidated. Breast cancer development is believed to be a polygenic and multistep process. A mutation in a gene may occur spontaneously at an estimated rate of about 10^{-6} mutations per gene per cell division (Albert et al. 1994*a*).

Mutations may impair normal cellular proliferation and apoptosis due to the dysfunction of one or more of three classes of genes; I- proto-Oncogenes, II-Tumour suppressor genes and recently III- genes that function in DNA damage recognition and repair (Offit, 1998). Tumour suppressor (TS) genes, of which both alleles must be inactivated to observe a phenotypic effect, protect cells from deregulated growth and division. Recently, the concept of 'gatekeeper' and 'caretaker' genes has emerged. Gatekeeper genes act directly to regulate cell proliferation and are rate limiting for tumourigenesis, an example being the retinoblastoma (Rb) gene. Caretaker genes, by contrast, do not directly regulate proliferation but when mutated lead to accelerated conversion of a normal cell to a neoplastic cell, an example is p53. Many caretaker genes are required for the maintenance of genome integrity reviewed by Levitt and Hickson (Levitt and Hickson 2002). However, the three types of genes could act during different phases in the dividing cell. The oncogenes act as accelerators of growth during the G1, or growth phase of the cell cycle. The suppressor genes act as stop signals before the S or synthesis phase of the cell cycle. The third class act as repairmen, identifying and fixing DNA mismatches following DNA replication, before the chromosomes condense in G2 phase for mitosis (M), Figure 1.1 (Albert et al. 1994*b*).

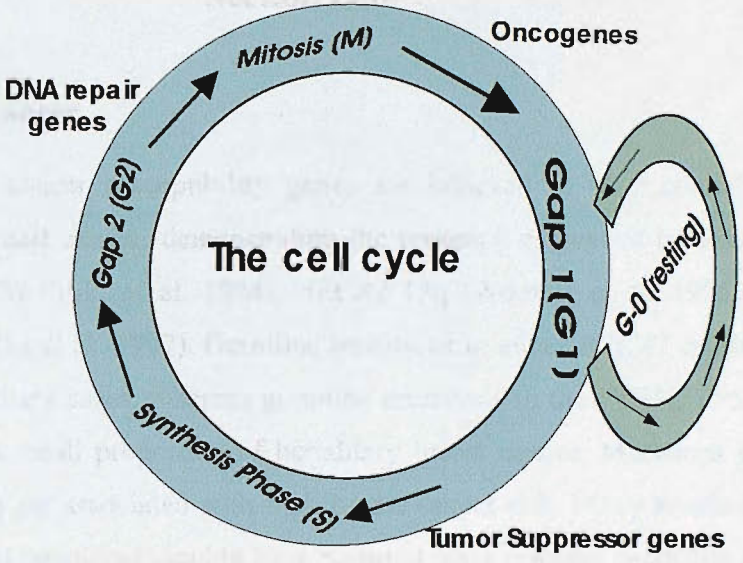


Figure 1.1. The control of the cell cycle. The essential processes, such as DNA replication and cell division, are illustrated. During the interphase, the cell grows continuously; during M phase it divides, DNA replication is confined to the part of interphase known as S phase. G1 phase is the gap between M phase and S phase; G2 is the gap between S phase and M phase. The three classes of regulatory genes are maximally expressed at the phases of the cell cycle where they have maximal effect to regulate the biological process.

Autosomal-dominant hereditary breast cancer syndrome is due to a mutation in either *BRCA1* or *BRCA2* and perhaps other candidate genes. The risk of breast cancer is thought to be variable according to the position of the mutation along the gene and to the genetic background of the ethnic population and to environmental influences.

In the first chapter of this thesis (section one) I review briefly the genes that predispose to breast cancer and have detailed relevant information regarding the *BRCA1* gene.
In Section Two, I introduce the methods that have been used for mutation detection of single nucleotide alterations in DNA sequence and the development of melt-MADGE.

Section One

Breast cancer

1.1 Genetics of breast cancer

To date, several breast cancer susceptibility genes are believed to play crucial roles in the occurrence of familial breast cancer, demonstrating the presence of genetic heterogeneity. These genes include *BRCA1* 17q (Miki et al. 1994), *BRCA2* 13q (Wooster et al. 1995) and Cowden Syndrome (PTEN, 10q) (Li et al. 1997). Germline mutations in either *BRCA1* or *BRCA2* account for the majority of hereditary cases, whereas germline mutations in the PTEN, TP53 and STK11 genes account for only a small proportion of hereditary breast cancer. Mutations in the *BRCA2* gene in chromosome 13q are associated with high breast cancer risk. Many another investigators using linkage analysis and positional cloning have reported other possible candidate gene. BRCA3 has been mapped to chromosome 8 at p12-p22 (Seitz et al. 1997). It is believed this locus may be involved in breast cancer, but no gene has been characterised thus far. Other potential susceptibility or modifier genes have been described. BARD1 on chromosome 2 (*BRCA1*- associated RING-domain 1) was identified through its physical association with BRCA1 but its role in breast cancer predisposition has not yet been fully defined (Sakorafas and Tsiotou 2000). In contrast, the Androgen-Receptor gene (AR-CAG) has been found to be correlated with increased breast cancer risk at an earlier age in BRCA1 gene carriers. Variation in the length of CAG repeats located in exon 1 may alter the androgen binding affinity of the receptor (reduced with the longer repeats) and the increasing number of AR-CAG repeats decreases the age of onset of breast cancer in *BRCA1* gene carriers (Rebeck et al. 1999).

Breast cancer (BC) is caused by multiple somatic mutations. In around 30% of cases low penetrance genes may be implicated whilst in a small proportion (less than 5%) of cases, there is a single high risk dominant inherited genes responsible. *BRCA1* and *BRCA2* are the most frequent causes of dominant inherited high penetrance breast cancer. Hereditary breast cancer is characterised by early onset and bilateral disease. In 1971, Knudson predicted the mechanism of action of tumour suppressor genes (Knudson, Jr. 1971) and his observation indicated that the familial form of retinoblastoma showed an earlier age of onset than the sporadic form and was frequently bilateral. He proposed a two-hit model for the inactivation of a tumour suppressor gene. In familial cancer, the first mutation (germline mutation) is inherited and present in all cells of an individual; the second hit is a somatic mutation resulting frequently from loss of the wild type allele. In sporadic cancer, both hits are somatic. 90-95% of breast cancer is considered sporadic, while the other 5-10% of women suffering from breast cancer have a strong family history of the disease and 80-90% of these patients carry a defective *BRCA1* or *BRCA2* gene in their germlines. *BRCA1* gene mutations have been implicated in 40% of all hereditary breast cancer cases (Hamann 2000).

1.2 Breast Cancer Type 1 (*BRCA1*)

A) *BRCA1* gene Structure

BRCA1 is the first candidate gene to have been identified as a gene causing breast cancer and Miki Y. and co-workers have extensively investigated it since its isolation in 1994. *BRCA1* is a large gene and encodes a mature protein of 1865 amino acide with a total cDNA of 5711 base pair. It is located on chromosome 17p21 and comprises 24 exons spanning more than 80kb with only 22 coding-exons, exon 1a, 1b, and part of exon 24 are non-coding (Figure 2.1), approximately 50% of the coding sequence is contained in a single exon, exon 11. Furthermore, 90.9% of *BRCA1* gene sequence consists of introns and 46.3% of its sequence contains repeat elements. Exon 4 is an *Alu* repetitive motif, consequently exon 4 will not be found in the normal *BRCA1* transcript. The length of the coding regions (exons) ranges from 40bp to 3425bp in length. The intronic sizes range from 403bp to 9193bp and 10 introns are longer than 3kb. The order of the gene on chromosome 17 is: centromere-IFP35-VAT1-RHO7-*BRCA1*-M17S2-telomere (Smith et al. 1996).

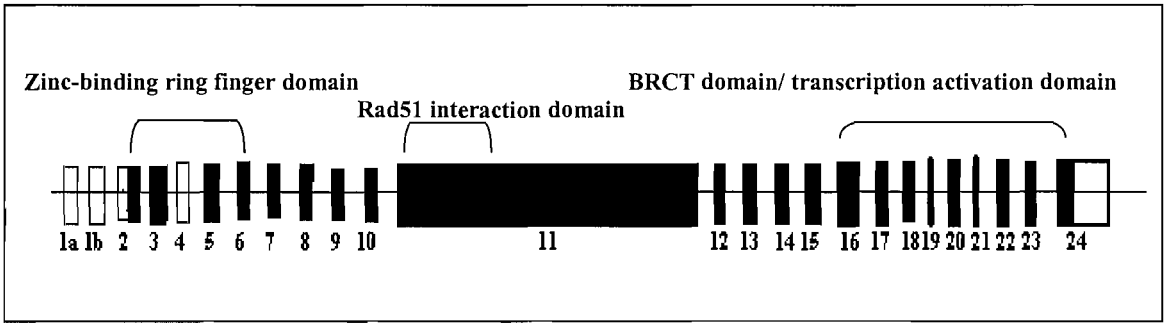


Figure 2.1. The putative structure and functional domains in the *BRCA1* gene. Blank boxes represent the non-coding regions.

The *BRCA1* gene is classified in three domains, corresponding to the amino acids of *BRCA1* protein encoded by the following domains, the first has a zinc or ring finger domain in the N-terminal region comprising exon 2-6. This domain is characterised by a protein-protein interaction. The second domain is located within exon 11 where the Rad51 interaction domain is located. Rad51 is a protein implicated in double-strand DNA repair and recombination. The third domain is the C-terminal region (BRCT) and it is located in the 3`end of the gene, it comprises exon 16-24. BRCT correlates with the transcription activation domain Figure 2.1 (Smith et al. 1996).

B) *BRCA1* expression

Hybridisation of mRNA to labelled fragments of *BRCA1* cDNA showed a wide range of transcripts of mRNA. Particularly, mRNA is abundant in breast, ovary, testis and thymus (Miki et al. 1994). Welcsh and co-worker stated that *BRCA1* mRNA and protein expression is induced during the late G1-early S phase of the cell cycle. In the mitotic cell, *BRCA1*, *BRCA2* and Rad51 interact and are co-localised in the nucleus during the S phase of the cell cycle. *BRCA1* function is regulated by phosphorylation, where it is hyperphosphorylated during the late G1 and S phases by an endogenous kinase activity and functions at the G1-S transition to arrest cell cycle progression by binding hypophosphorylated retinoblastoma protein. In addition, *BRCA1* could have a role for the G2-M checkpoint by controlling the assembly of mitotic spindles and the appropriate segregation of chromosomes to daughter cells (Welcsh, Owens, and King 2000).

C) *BRCA1* protein structure

The *BRCA1* gene encodes a protein of 1863 amino acids with no similarity to other known proteins excluding the two terminal domains, N-terminal domain harbouring ring finger motif and C-terminal domain containing the BRCT domain (Figure 3.1) (Huyton et al. 2000; Williams, Green, and Glover 2001). These locations on the *BRCA1* protein have received extensive attention because of their proposed functional roles in the developing perturbation process during cancer cell formation.

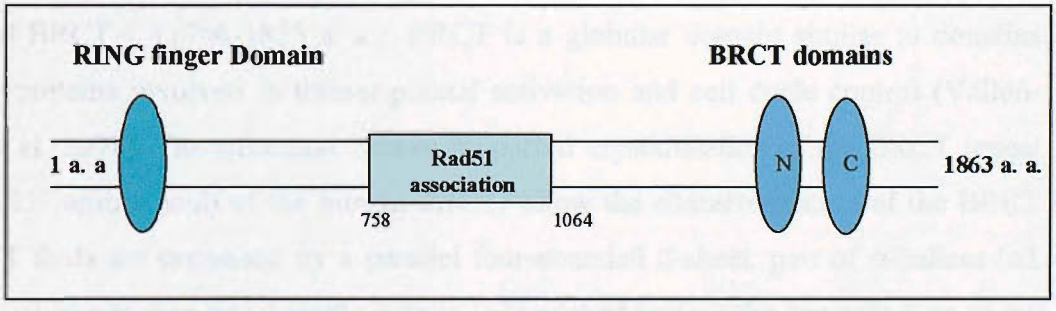


Figure 3.1. Schematic representation of full-length *BRCA1* protein displays the domain structure of *BRCA1*, The RING domain in the N-terminal region and the two BRCT domains in the C-terminal region.

1. RING finger domain

The RING finger domain is located in the N-terminal region of the *BRCA1* protein, which is encoded by exons 2, 3, 5 and 6, believed to be involved in transcriptional activation by interaction with other molecules (Lorick et al. 1999). RING finger domains are distinguished by short antiparallel three standard β-sheet (residues 35-37, 42-44, and 74-76), and three α-

helices (residues 8-22, 46-53, and 81-96) (Brzovic et al. 2001a). Zn^{2+} is a required metal ion to promote structural stability and integrity. It is found at the core of a number of different domains involved in protein-protein interaction including a Zinc finger motif at the RING finger domain (Schwabe and Klug 1994). The RING finger is a zinc-binding domain defined by a conserved pattern of seven cystenes and one histidine residue that act as metal ligand and mediates protein-protein or protein-DNA interactions. These eight residues have distinguishable amino acid sequences that tend to initiate chemical bonds with Zn^{2+} atoms that then form distinct Zn^{2+} binding sites (site I and site II) (Brzovic et al. 2001a). The ligands that form Site I are Cys 24, Cys 27, Cys 44, and Cys 47, and the ligands in site II are Cys 39, His 41, Cys61, and Cys64 that encompass residues 24-64 (Brzovic et al. 2001a). The nucleotide position on the coding DNA of the *BRCA1* gene, which encodes site I, is located between nucleotide number 189 in exon 2 and nucleotide number 260 in exon 5. The coding sequence of site II is overlapped with the region that encodes site I, however, the coding nucleotides for site II are located between nucleotide number 244 in exon 3 and nucleotide number 311 in exon 5.

2. *BRCA1* C-terminus (BRCT)

Breast Cancer C-Terminal domain (BRCT) represents the third region in the BRCA1 protein, which is encoded by exons 16-24 (Welcsh, Owens, and King 2000). The BRCT resides in 336 amino acids of the full length of 1863 amino acid and forms two domains, BRCT-N (1649-1736 a. a.) and BRCT-C (1756-1855 a. a.). BRCT is a globular domain similar to domains found in other proteins involved in transcriptional activation and cell cycle control (Vallon-Christersson et al. 2001). The structural studies by partial crystallisation of the BRCT repeat region (1646-1859 amino acid) of the human *BRCA1* show the characterization of the BRCT subunits. BRCT folds are organised by a parallel four-stranded β-sheet, pair of α-helices (α1 and α3) sited against one face and a single α-helix (α2) packed against the opposite face of the sheet (Figure 4.1).

The order of the subunits of the two BRCT domains on *BRCA1* are BRCT-N β1-α1-β2-β3-α2-β4-α3; amino acid linker of about 20 to 23 a. a.; BRCT-C β1'-α1'-β2'-β3'-α2'-β4'-α3' (Williams, Green, and Glover 2001). However, BRCT repeats from the XRCC1 protein (**X**-Ray Repair **C**ross-**C**omplementing 1) (Thornton et al. 1999) are the homology of the *BRCA1*-BRCT domain. The alignment of both XRCC1-BRCT domains and *BRCA1*-BRCT domains reveals that the arrangement of the α1, α3 and the central β-sheet are conserved in all the repeats. This conservation region is defended by the hydrophobic residues within the α1, α3 and the central β-sheet core (Williams, Green, and Glover 2001).

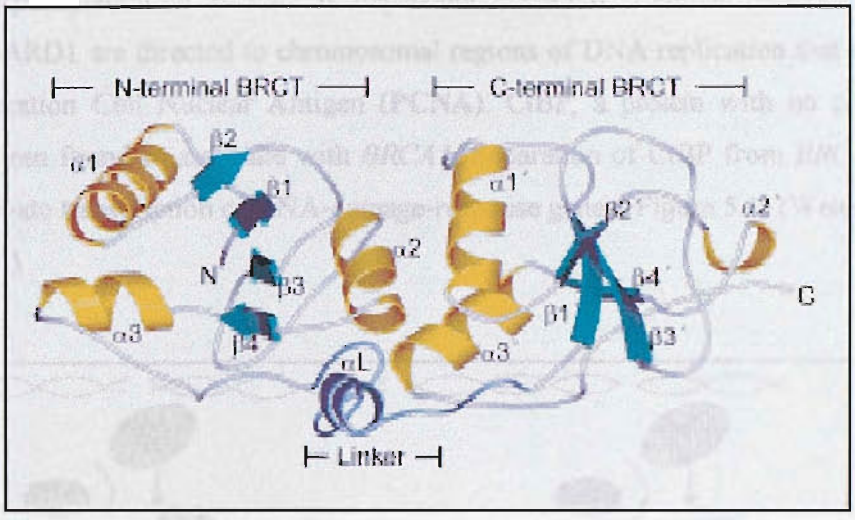


Figure 4.1. Ribbon structure of the BRCT domains on the *BRCA1*. The secondary structure elements in the C-terminal BRCT repeat labelled `prime` to differentiate them from the N-terminal BRCT. This structure was adopted from Williams et al., 2001.

D) *BRCA1* function

The *BRCA1* gene encodes a 220-kDa nuclear phosphoprotein in which structural changes confer susceptibility to familial breast and ovarian cancer (Miki et al. 1994). The *BRCA1* protein (1863 amino acid sequences) has several biological functions, from DNA damage repair, protein degradation, to chromatin remodelling and transcription activation. *BRCA1* is thought to be a multifunctional molecule and does not enter cellular functions alone, since it makes a complex structure with other cellular molecules for optimal performance. Thus, several proteins have been found to interact with the BRCT domain, including TP53, RNA polymerase II holoenzyme and CtBp-interacting protein which support the role of *BRCA1* in transcription regulation (Welcsh, Owens, and King 2000). *BRCA1* is involved in the DNA repair system through interaction with different proteins in response to DNA damage. The interactions of *BRCA1* and *BRCA2* in complexes that activate double strand break repair and initiate homologous recombination links the maintenance of genomic integrity to tumour suppression. Additionally, RAD51, the homologue of bacterial RecA which is required for recombinational repair of double strand DNA breaks, is another of the proteins that is involved in DNA repair (Welcsh, Owens, and King 2000; Zhong et al. 1999).

RAD51 binds with *BRCA1* and *BRCA2*, forms a BRCA1-BRCA2-RAD51 complex, and is co-localised in core form. In addition, Rad50 is introduced during homologous recombination mediated DNA damage repair in association with *BRCA1*. BARD1, *BRCA1* associated RING domain protein-1 (BARD1), is also one of the proteins that is involved in this process; it binds to *BRCA1* via its Zinc finger motif in the RING finger Domain. Within this biological

complex, phosphorylation of *BRCA1* is dependent on ATM. Cellular molecules RAD51, *BRCA1* and BARD1 are directed to chromosomal regions of DNA replication that are marked by the **P**roliferation **C**ell **N**uclear **A**ntigen (**PCNA**). CtBP, a protein with no clear known function has been found to associate with *BRCA1*. Separation of CtBP from *BRCA1* enables *BRCA1* to activate transcription of DNA-damage-response genes (Figure 5.1) (Welcsh, Owens, and King 2000).

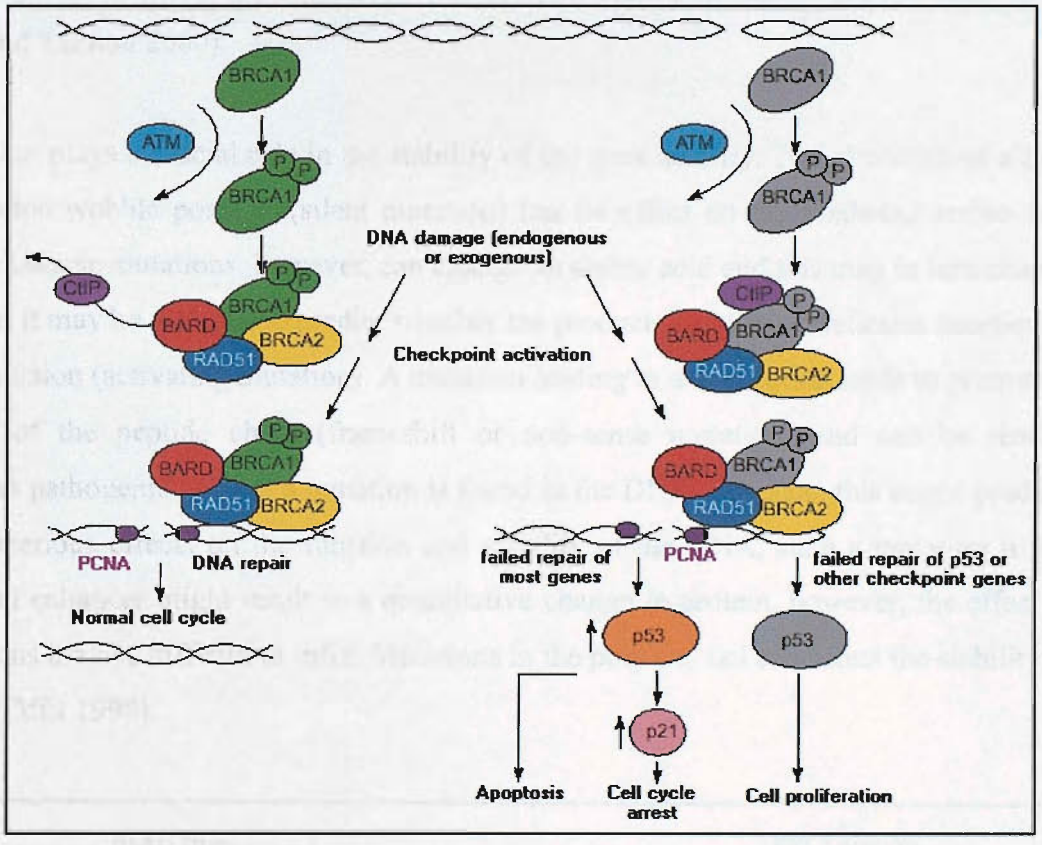


Figure 5.1. *BRCA1* in DNA repair. In this diagram the interaction complex of *BRCA1*, *BRCA2*, BARD1 and RAD51 showed the proposed function for these molecules in DNA repair system. The Figure was adapted from Welcsh, P.L (Welcsh, P.L. et al., 2000).

BRCA1 protein is suggested to be involved in the regulation of the cell cycle. This was supported by observation of the expression of *BRCA1* mRNA and protein during different stages of the cell cycle, existing in the G1 phase and peaking in the S-phase. Moreover, over-expression of *BRCA1* activates $p21^{WAF1/CIP1}$ a P53-independent pathway and inhibits cell cycle progression into S-phase (Somasundaram 2003; Somasundaram et al. 1997).

1.3 Mutations in *BRCA1* gene

BRCA1 mutations are scattered throughout the coding and non-coding sequence of the gene (Figure 6.1). Many mutations have been identified throughout the entire coding sequence of the *BRCA1* gene. However, the majority of the mutations identified in breast and ovarian cancer families have been shown to be due to frameshift and nonsense mutation that lead to the truncation of *BRCA1* protein (Perrin-Vidoz et al. 2002). The incidence of specific mutations have been reported in many breast cancer families in different populations (Backe et al. 1999; Sakorafas and Tsiotou 2000).

Point mutation plays a crucial role in the stability of the gene activity. The alteration of a base pair in a codon wobble position (silent mutation) has no effect on the predicted amino acid sequence. Missense mutations, however, can change an amino acid and this may in turn change function, but it may be difficult to predict whether the product might have deficient function or increased function (activating mutation). A mutation leading to a stop codon leads to premature termination of the peptide chain (frameshift or non-sense mutation) and can be readily interpreted as pathogenic. Where a mutation is found in the DNA sequence, this might produce mild or deleterious effects on the function and stability of the DNA, such a mutation in the promoter and enhancer might result in a quantitative change in protein, however, the effect of such mutations may be difficult to infer. Mutations in the poly (A) tail can affect the stability of the mRNA (Offit 1998).

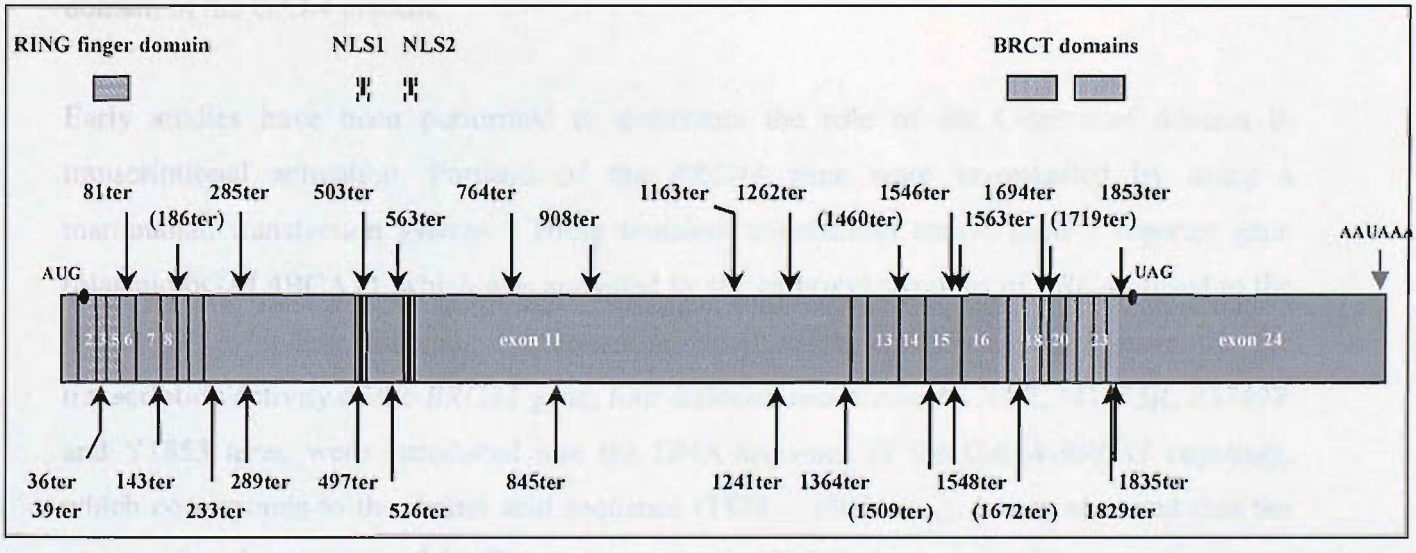


Figure 6.1 Presented some of the mutations scattered throughout *BRCA1* gene. These mutatons are arised in the coding region (Perrin-Vidoz, L, 2002).

Mutations in DNA molecules can occur by one of many mutagenic episodes, endogenous or exogenous. One example of the endogenous mutational mechanism is the high frequency of C to T transitions at the CpG islands in colon cancer, which is caused by deamination of 5-methylcytosines found at CpG sites. On the other hand, exogenous mutational effect is the transversion mutations associated with lung cancer that are caused by benzopyrene mutagens (Xu and Solomon 1996). Another type of mutation is the small deletions that occur at repeated nucleotides as in AA deletion from several A (2800delAA) or T deletions from TTT (1436delT). Other multiple repeat AG repeats (185delAG) and a 40 bp deletion from TGAACT...TGAATC (1294del40). These changes possibly occur during DNA replication as a result of slippage of the template strand and its subsequent misalignment (Trinh and Sinden 1991).

1.3.1 Effect of the mutations in the C-terminal of *BRCA1* gene

Since discovering the *BRCA1* gene (Miki et al. 1994) many investigators have studied the role of this gene product with the intention of elucidating its function in the cell. The *BRCA1* gene has different domains along its sequence and each domain is association with one or more specific functions. One of the functional domains is the C-terminal domain, which is thought to be important in transcription activation. However, many different alterations have been found along the *BRCA1* gene sequence and only mutations that are located within functional domains are likely to give rise to disease. The *BRCA1* C-terminus (amino acids 1528-1863) induces transcription from reporter plasmids in mammalian cells when fused with the DNA-binding domain of the GAL4 protein.

Early studies have been performed to determine the role of the C-terminal domain in transcriptional activation. Portions of the *BRCA1* gene were investigated by using a mammalian transfection system. These transient transfection assays used a reporter gene (plasmid pGAL4BCAT), which was activated by the carboxyl terminus of *BRCA1* fused to the GAL4 DNA-binding domain. To determine what effect mutations might have on the transcription activity of the *BRCA1* gene, four different mutations, A1708E, M1775R, P1749R and Y1853 *term,* were introduced into the DNA sequence of the GAL4-*BRCA1* construct, which corresponds to the amino acid sequence (1528 – 1863 a. a.). It was observed that the transcriptional activation of *BRCA1* was completely abolished when the four mutations were implanted into GAL4-*BRCA1* (1528 – 1863 a. a.). However, upon introducing of M1775R mutation into GAL4-*BRCA1* that corresponded to the amino acids (1760 – 1863 a. a.), it displayed 50% activity of the wild-type *BRCA1*.

Lack of transactivation by introducing Y1853*term* that causes a premature termination codon at nucleotide position 1853, in which the last 11 C-terminal amino acids are deleted, showed that these terminal amino acids determine the transactivation domain in *BRCA1* and are essential for *BRCA1* activity. Although the A1708E and P1749R mutations showed complete loss of function when introduced into *BRCA1* (1528 – 1863 a. a.), the GAL4-*BRCA1* (1760 – 1863 a. a.) which does not contain the mutations sites A1708E and P1749R, which abolish transcription, shows some *BRCA1* activity. It is possible that these alterations might be involved in tertiary folding of the C-terminal region suggesting that these mutations destabilise the tertiary structure. Therefore, *BRCA1* amino acids 1528 – 1863 indicated significant transcriptional activation of the reporter gene when fused to the GAL4 DBD, Figure 7.1. In contrast, the smaller portion (amino acids 1760 – 1863) showed reduced transcription activity (Monteiro, August, and Hanafusa 1996; Chapman and Verma 1996).

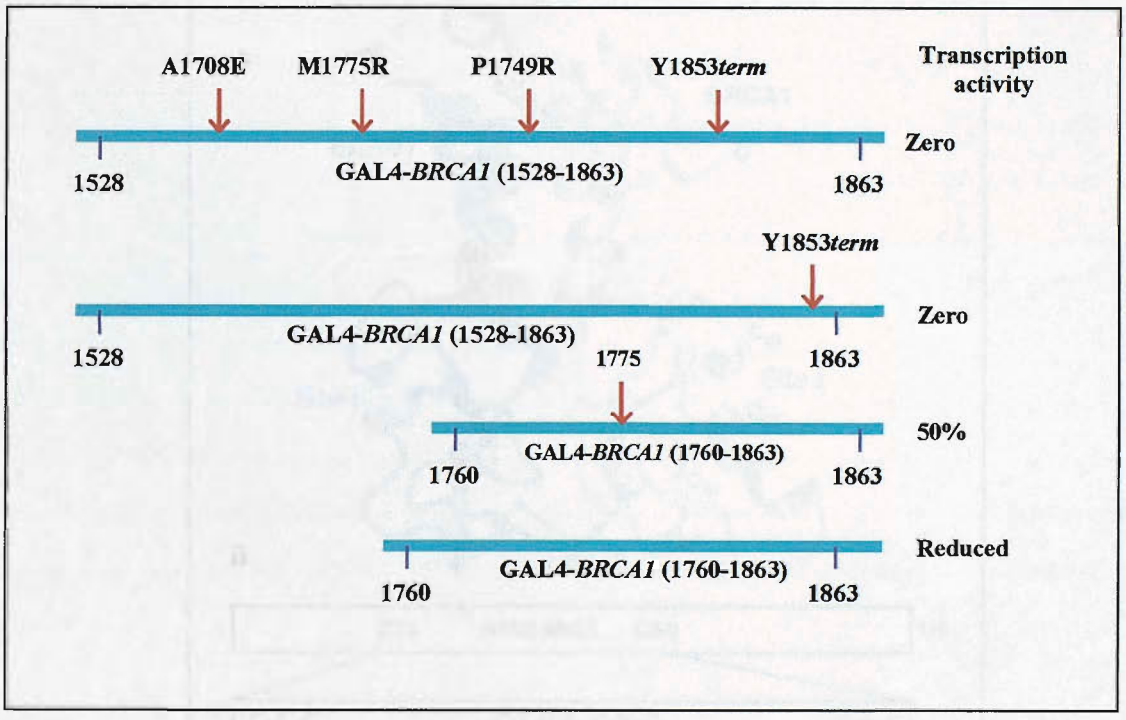


Figure 7.1. Illustration of the mutations affect on the portion sequence of *BRCA1* gene was achieved by construction pGAL4 with C-terminal of *BRCA1* gene to determine the transcription activation of the C-terminal domain.

1.3.2 Effect of the mutations on the RING finger domain of the *BRCA1* gene

The RING finger domain in *BRCA1* has an important role in DNA damage response via its binding site with BARD1 through the ZINC finger motif. Zn^{2+} ions play an essential role in stabilising the *BRCA1* RING domain. Mutations of Cys residues in either site, site I or site II, will disturb the binding affinity to the other binding proteins, especially BARD1 (Figure 8.1). Mutations in site I (Cys 24 and Cys44) cause disturbance of the *BRCA1* RING domain structure and prevent the formation of the heterodimer (Brzovic et al. 2001a). While changes in the structure and dynamics of the second Zn^{2+} binding loop might contribute to the effect of mutations in site II, mutations at Cys 39 / Cys 61 and Cys 64, might lead to the inability of *BRCA1* to work in the ubiquitin transfer reaction by affecting the *BRCA1* RING (E3)-E2 binding interface (Brzovic et al. 2001b).

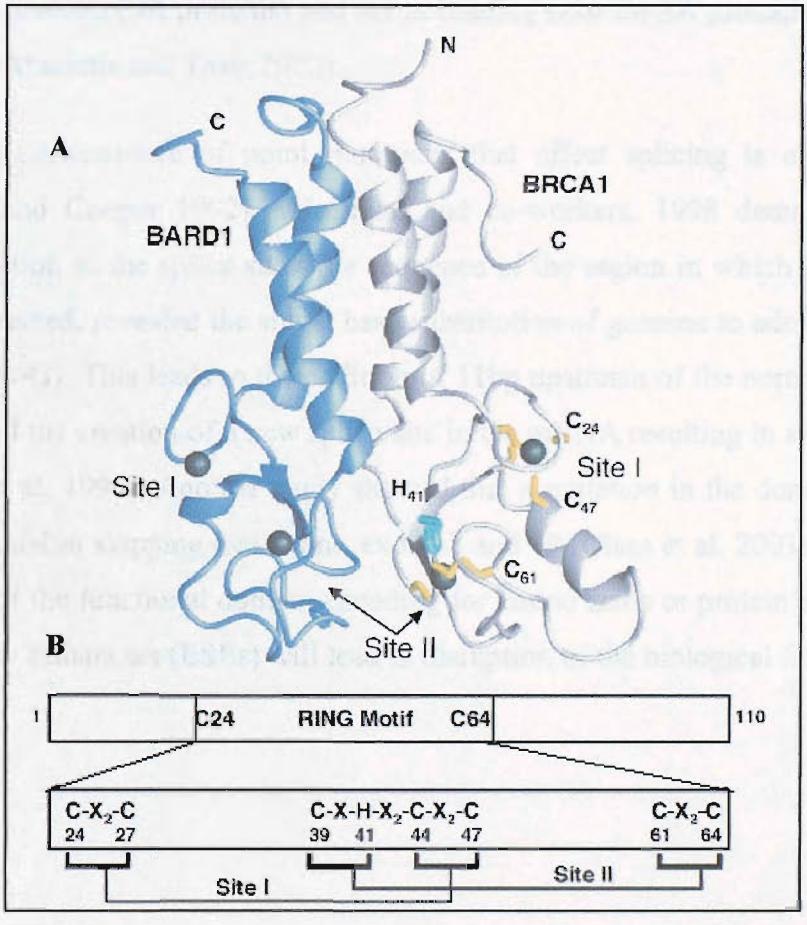


Figure 8.1. Shows the backbone structure of the RING domains of the *BRCA1* and BARD1 proteins. A) The gray colour represents the *BRCA1* subunits and the blue colour represents the BARD1 subunits. The Cys amino acid is in yellow and His is in cyan colour. **B)** shows the *BRCA1* RING motif, residues 24-64, and the arrangment of pairs of Zn^{2+} binding ligands that form sites I and II. Adapted from (Brzovic et al. 2001a).

1.3.3 Mutations affecting exonic splicing

All the genes translated into proteins in eukaryotes are split into coding (exon) and non-coding (introns) regions. The non-coding sequences are removed from the entire sequence before the mRNA is transported to the cytoplasm for translation. However, in the primary mRNA gene sequence, eight nucleotides at the 5` splice site are highly conserved at the junction of exon with intron. While the position at the 3` splice site is recognised by the conserved sequence of 4 nucleotides, preceded by a pyrimidine-rich domain (Shapiro and Senapathy 1987). In general, up to 50% of point mutations responsible for genetic disease in humans cause abnormal splicing. These point mutations can disorder splicing by one of two mechanisms, directly inactivating the splice site or indirectly activating a cryptic splice site or intermediating with regulatory *cis* elements like exonic splicing enhancers (ESE); ESEs are recognised by serine/arginine-rich proteins (SR proteins) and act as binding sites for SR proteins for efficient splicing of an exon (Maniatis and Tasic 2002).

The most common consequence of point mutations that affect splicing is exon skipping (Krawczak, Reiss, and Cooper 1992). Hoffman and co-workers, 1998 demonstrated one example of an alteration in the splice site. The sequence of the region in which the truncated protein had been detected, revealed the single base substitution of guanine to adenine in intron 5 of *BRCA1* (IVS5A>G). This leads to the addition of 11bp upstream of the normal splice site preceding exon 6 and the creation of a new splice site in the mRNA resulting in a short protein product (Hoffman et al. 1998). Another study showed that a mutation in the donor splice site (IVS19+2delT) resulted in skipping two exons, exon 18 and 19 (Claes et al. 2003). Changes in the DNA sequence of the functional domain encoding for amino acids or protein binding sites, like Exonic Sequence Enhancers (ESEs) will lead to disruption of the biological function of the coded protein.

1.4 Penetrance

In the general population, carriers of identical mutations may develop different cancers, the same cancer at diverse ages, or may never develop cancer at all. An important term in genetics is "penetrance". Penetrance is defined as the likelihood that an individual carrying a deleterious mutation will develop the associated disease. For a cancer predisposition gene this is frequently reported as the probability of cancer by the age of 70 years (Begg 2002). A woman with a deleterious mutation in *BRCA1* or *BRCA2* gene may be at higher risk for breast cancer than a woman with similar mutation with no known family history of breast cancer. Cancer risk is possibly modified by numerous unknown genetic risk modifiers (Begg 2002). The penetrance is determined by age; in carriers likelihood of disease increases with increasing age (Satagopan et al. 2001). Variation in clinically expressed phenotypes among people with the same mutation indicates the presence of modifying factors that are either genetic or environmental factors. As an example, an individual with familial hypercholesterolemia that can cause coronary heart disease will increase their risk by smoking, poor diet and lack of exercise (Burke and Austin 2002). Genes with high penetrance may be less influenced by environmental factors and these genes are rare in the population. In contrast, a gene with low penetrance may be much more influenced by environmental factors that modify the risk of having the disease (Begg 2002). In general, however, other genes and the environment are likely to influence the penetrance of a mutation (Vineis, Schulte, and McMichael 2001). The effect of a founder mutation (Neuhausen 1999), which can be studied in multiple individuals, can help determine other influences on penetrance. A lower penetrance of some mutations may be due to one or more of the following factors: 1) functional importance of the protein encoded by the gene, e.g. in key regulatory aspects of the cell cycle, mutation in this type of gene is often associated with high penetrance; 2) functional importance of the mutation; 3) interaction with other genes; 4) onset of somatic mutations; 5) interaction with the environment; 6) existence of alternative pathways that can replace the loss of function (Vineis, Schulte, and McMichael 2001). For frequent mutations in a gene (polymorphisms) if there is any effect at all on disease susceptibility it will be small i.e. the penetrance of polymorphisms will be very low if it has any effect at all.

Different mutation positions in the same gene (*BRCA1*) may be associated with a different pattern of disease e.g., Easton et al., (Easton, Ford, and Bishop 1995) who have studied mutations in the *BRCA1* gene found association with increased susceptibility to breast cancer or ovarian cancer in different proportions. The study by the Breast Cancer linkage Consortium (BCLC) have reviewed family histories from thirty-three families contributed by 11 collaborative groups from the BCLC. These families had to meet certain criteria, including at

least four cases of either ovarian cancer diagnosed at any age or of breast cancer diagnosed before age 60. Linkage studies had produced estimates of penetrance by age 70 in the range of 85% for the *BRCA1* gene.

More detailed analysis have showed that the observed pattern of disease risks could be explained by two different *BRCA1* alleles, allele 1 and allele 2. Allele 1 was conferring a cumulative breast cancer risk of 91% by the age of 79 year with an ovarian cancer risk of 32%, whereas second allele was conferring a breast cancer risk of 70% and an ovarian risk of 84%. Figure 9.1 represents the cumulative risk of breast and ovarian cancer in female carriers, which is estimated to be more than 80% at the age 70 years. The low allele frequency, allele 1 (29%) is estimated to confer a breast cancer risk of 60% by the age 60 years, while ovarian cancer risk is estimated to be 11%, at the age 70 years, the cumulative risk is estimated to be over 80% for breast cancer. The higher frequency allele, allele2 (71% of families), showed relatively lower risk of breast cancer but higher risk of ovarian cancer; allele 2 conferred a breast cancer risk at age 70 of 65% compared to over 80% with allele 1. The estimated penetrance of breast cancer in the selected cases studied by Easton and co-workers have demonstrated evidence of existing modifying factors that conferred a variant risk for different *BRCA1* mutations (Easton, Ford, and Bishop 1995).

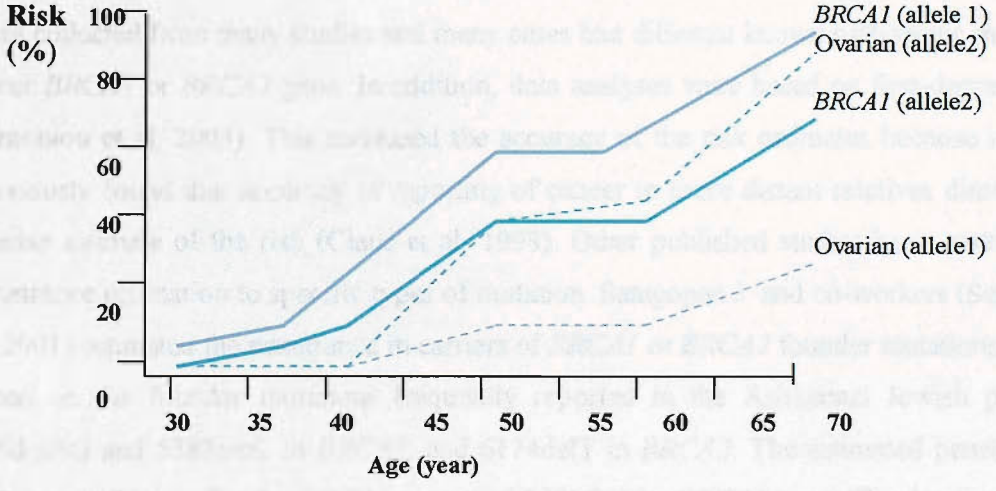


Figure 9.1. Illustrates the correlation between the age of onset and the likelihood for the person to express the phenotype of the disease (Easton et al. 1995).

In a more recent study carried out by Antoniou and colleagues (2003) a meta-analysis of 22 population based studies was presented. This study was designed to estimate the average risk of breast and ovarian cancer in association with *BRCA1* or *BRCA2* mutations. The cases studies were selected to meet certain criteria, subjects had to be sampled independently of family

history, subjects had to have been analysed for *BRCA1* or *BRCA2* mutations and the studies selected were population based for affected cases of female or male breast cancer. The investigators estimated the cumulative risk of breast cancer in this study to be 65% by age 70 years for *BRCA1*-mutation carriers (Figure 10.1).

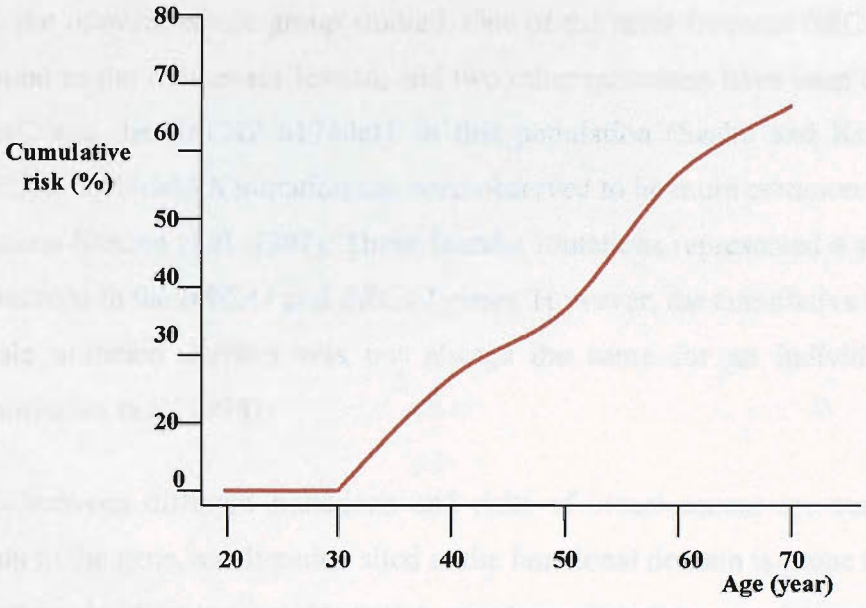


Figure 10.1. Illustrates the estimated penetrance of *BRCA1* by age, (Antoniou A. et al., 2003).

This study forms a new model estimating penetrance for *BRCA1* in large populations; the data were collected from many studies and many cases had different known pathogenic mutations in either *BRCA1* or *BRCA2* gene. In addition, data analyses were based on first-degree relatives (Antoniou et al. 2003). This increased the accuracy of the risk estimates because it has been previously found that accuracy of reporting of cancer in more distant relatives diminishes the precise estimate of the risk (Claus et al. 1998). Other published studies have restricted their penetrance estimation to specific types of mutation. Satagopan J. and co-workers (Satagopan et al. 2001) estimated the penetrance in carriers of *BRCA1* or *BRCA2* founder mutations. This was based on the founder mutations frequently reported in the Ashkenazi Jewish population, 185delAG and 5382insC in *BRCA1*, and 6174delT in *BRCA2*. The estimated penetrance was 46% at age 70 year for the *BRCA1* gene, and 26% for the *BRCA2* gene. This lower estimate in cumulative risk that applies to this population study is consistent with the observation, that higher allele frequency is associated with lower penetrance. However, Antoniou's study may represent a more accurate estimate of penetrance in a population with no clear founder mutation.

1.5 Genotype-Phenotype association

The effect of mutation location on phenotypic expression at the disease level was assessed by analysis of the most common *BRCA1* mutations that suggested these mutations arose from common ancestry (founder mutations). These founder mutations have been detected in every population and, the relevant ethnic group studied. One of the most frequent *BRCA1* mutations is 185delAG found in the Ashkenazi Jewish, and two other mutations have been observed, the *BRCA1* 5382insC and the *BRCA2* 6174delT in this population (Szabo and King 1997). In addition, the *BRCA1* 2804delAA mutation has been observed to be more common in Dutch and Belgian populations (Peelen et al. 1997). These founder mutations represented a small fraction of the many mutations in the *BRCA1* and *BRCA2* genes. However, the cumulative risk of breast cancer in female mutation carriers was not always the same for an individual in these populations (Thorlacius et al. 1998).

The association between different mutations and risks of breast cancer are correlated with mutation location in the gene, an alteration sited in the functional domain is prone to have more effect than mutations in other regions. However, genotype-phenotype correlations suggest that mutations located in the first two thirds of the gene (5`end of the gene) are associated with a significantly greater risk of ovarian cancer in contrast with mutations located in the last third of the gene (3`end of the gene) (Gayther et al. 1995; Eccles et al. 1998). Further evidence for this is seen in patients who have a mutation in *BRCA1* 5382insC, which is located at the 3`end, who are more susceptible to developing ovarian cancer than those who have mutation in *BRCA1* 185delAG, which is located at 5`end (Struewing et al. 1997).

However, it is not necessary that a phenotype would always be explained by finding a point mutation in the *BRCA1* or *BRCA2* genes. A recent study indicated that screening all the coding regions of *BRCA1/2* genes for the mutation in 38 Swiss breast cancer patients by DNA sequencing revealed that 50% had no mutation in either gene. The patients were Caucasian origin. Moreover, patients entered into this study were based on the certain criteria, family history including one first degree on the same side of the family with breast or ovarian cancer, diagnosis of both breast and ovarian cancer at any age, bilateral breast cancer, breast or ovarian cancer before the age of 40 and male breast cancer (Schoumacher et al. 2001). However, more investigations need to be performed to clarify the result of this analysis. An accurate method to clarify Schoumacher finding is to be tested by linkage study. In contrast, may individuals who were tested positively for the founder mutations, 185delAG, 5382insC and 6174delT, showed no sign of developing either breast or ovarian cancer (Moslehi et al. 2000).

1.6 Single Nucleotide Polymorphisms (SNPs)

SNPs are single base pair variations in genomic DNA where the variant allele is frequently present in the general population with no effect on the carriers. A locus is deemed polymorphic where alleles or variants show a frequency of more than 1% in the general population, which contribute to about 90% of human genetic variation and is assumed to result in neutral or benign phenotypic alterations (Collins, Brooks, and Chakravarti 1998). Polymorphism originates as a result of such changes in DNA sequence. The more common types of polymorphism are substitution alterations, which substitute one nucleotide for another. Other types of genetic polymorphism observed in human genes are the insertion and deletion of a section of DNA sequence. For example, variable number tandem repeats (VNTR) may be the site of large alterations at the nucleotide level, possibly deletion or insertion of hundred(s) of base pairs. In addition, simple tandem repeats (STRs) occur with short repeats (1-5 bp) and a smaller size range of allelic length variation (Schork, Fallin, and Lanchbury 2000). Furthermore, *Alu* elements, evident as a visible band in total genomic DNA when it is digested by enzyme *Alu1*, contain a stretch of nucleotides (120-150bp) terminally an adenine (A) base-rich section. Within the genome, Alu elements appear at an estimated one unit/3k on average. These polymorphisms are useful for genetic studies (Smith et al. 1996).

The human genome consists of 3 billion base pairs but only a small fraction, approximately 5%, is coding sequence (Nickerson et al. 1998). Moreover, comparison of two random human chromosomes has revealed an average difference at one in every thousand nucleotides (Kwok et al. 1996). However, in cohort studies using 106 human genes single nucleotide polymorphisms (SNPs) show a similar frequency in exon and intron regions. Thus, in exonic sequences the estimated frequency of SNPs found is one per 346bp, while in intronic sequences the frequency of SNPs is shown to be one per 354bp (Cargill et al. 1999). These single alterations in DNA sequence can have dramatic changes in the coded protein function, because some intron regions contain sequences important for regulatory functions such as the ring binding domain in *BRCA1*, for example BARD1 and changes in splice site binding affinity. Cargill and co-workers have indicated in his study of human genes that each gene has an average of about four cSNPs each with low frequency of approximately 2% in the population. Thus, 50% of cSNPs lead to missense mutation and the other 50% lead to silent mutation (Cargill et al. 1999). Schork and colleagues (Schork, Fallin, and Lanchbury 2000) observed SNPs scattered though the gene and found SNPs in introns creating cryptic splice sites. This resulted in duplication of a whole exon as in exon 13 of the *BRCA1* gene (Mazoyer et al. 1998) or deletion of whole exon as in exon 17 of the *BRCA1* gene (Montagna et al. 1999;

Puget et al. 1997). The latter group have also identified SNPs in exons destroying splice donor/ accepter site and causing protein to be truncated. SNPs in a promoter could affect gene expression and in enhancer elements could reduce the transcription activation (Schork, Fallin, and Lanchbury 2000).

In recent studies, it has been shown that 90% of missense mutations identified in clinical samples lead to impaired protein stability by changing the chemical properties of the residues in the DNA sequence that is involved in molecular stability. On the other hand, 70% of SNPs, that code for missense mutations found in the general population (not selected for inherited disease) are neutral alterations in which there is no influence on the protein function or structure (Wang et al. 2001). The application of SNP identification has been widely expanded. They are more stable and easier to work with than the microsatellite repeat variants that are involved in gene mapping (Collins, Brooks, and Chakravarti 1998) which had profound implications for disease-marker correlations and in association studies (Chakravarti 1999). In addition, discovering and studying SNPs will provide validation, quality control, and comparison of different methods of detecting SNPs (Collins, Brooks, and Chakravarti 1998). However, it has been commonly accepted that polymorphisms in the general population are practically neutral.

1.7 Association Studies in Human Genetics

The examination of DNA sequence variation to identify genes that may influence multifactorial diseases and the advantages of studying a single nucleotide polymorphism (SNP) in a DNA sequence will provide valuable information in relation to the phenotype of various diseases. The completion of the Human Genome wide sequence, which encodes 30,000- 40,000 different genes, will also facilitate the epidemiological studies in large populations in which genetic changes correlated with phenotype may reveal associations (Day et al. 2001). In epidemiological studies, the population under investigation should be stable. Genetic heterogeneity is one of the principles for human population studies (Schork, Fallin, and Lanchbury 2000). Different methods are used to identify the candidate loci associated with any disease; segregation analysis is used to determine what model of inheritance is most likely in relation to the phenotype; linkage analysis tests whether, in a family, a candidate locus is transmitted with the phenotype using genetic markers of known chromosomal location followed by positional cloning. The more powerful study methodology for large populations and small genotypic effect is an association study in ethnically homogenous populations. However, association studies are population-based case-control studies that are aimed to detect

any relationship between variations in nucleotide sequence in a gene and phenotypic consequences (Gambaro, Anglani, and D'Angelo 2000). A common form of population genetics analysis is to study single markers, to determine a link to susceptible genes that cause disease. This approach has been used in many studies using genetics markers. It not only provides a description of the structure of genetic variation found within a general population, but also allows estimation of the rate of gene transmission in the population (Gambaro, Anglani, and D'Angelo 2000). Centralisation of molecular genetics and epidemiological sciences for studying the nucleotide variations, at population level, and its influence on the gene function and structure has further advanced research for identifying a single marker associated with disease. For example, the angiotensinogen gene marker, which is commercially available, can be used to provide evidence of a relationship of this gene to hypertension (Day et al. 2001).

Population associated studies can be used for the investigation of complex disease. A complex disease is a multistage process and its susceptibility depends on gene and gene-environment interactions. A result of heterogeneity of genes in the general population, individuals may respond differently to environmental factors; the more convincing associations are these where a biologically plausible cause and effect between a gene polymorphism and a disease (Gambaro, Anglani, and D'Angelo 2000). However, different mutations in a single gene might contribute to different disease phenotypes or have slight effects on the pattern of inheritance, and as a consequences, common disease may be influenced by one or more mutations at the same locus (Kwiatkowski 2000). Polymorphisms and mutations occurring in introns may affect biological function. Introns contain sequences that are important for gene regulatory functions, for example as enhancers and silencers (Cargill et al. 1999). However, the study of Brookes (Brookes 1999) suggested that, in an association study SNP analysis should be performed in the region of the gene that would be predicted more likely to have functional consequences such as non-synonymous cSNPs and promoter variants.

Section Two

Mutation Detection Methods

Mutation detection has recently found important applications in functional genomic and clinical diagnosis. Many mutation detection techniques have been developed. These include, for example, *S*ingle *S*tranded *C*onformational *P*olymorphism (*SSCP*)/*H*eteroduplex *A*nalysis (*HA*), *D*enaturing *G*radient *G*el *E*lectrophoresis (*DGGE*), protein truncated test (PTT), Denaturing High-performance Liquid Chromatography (dHPLC) and DNA sequencing (Cotton 1997). However, these approaches do not lend themselves well to studying many samples (>1000 sample) in parallel. Analysis of long sequences of DNA for searching alterations by nucleotide sequencing techniques is time consuming and expensive. Therefore, The *M*elt Microplate-*A*rray-*D*iagonal-*G*el *E*lectrophoresis (Melt-*MADGE*) technique (Day et al. 1999a; Day et al. 1998) has significant advantages for mutation scanning. It generates high throughput at less expense and less time with high sensitivity in mutation detection. We have developed a Melt-MADGE assay for the *BRCA1* gene in order to undertake mutation research studies of less highly selected cohorts. There are two classes of mutation detection method. I-specific methods that identify specific characterised sequence variations such as allele-specific amplification and allele-specific oligonucleotide hybridisation. II- Scanning methods that detect uncharacterised sequence variations such as single strand conformation polymorphism (SSCP), heteroduplex analysis (HA) and denaturing gradient gel electrophoresis (DGGE).

2.1 Nondenaturant methods

2.1.1 Single Strand Conformation Polymorphism (SSCP) method

SSCP (Orita et al. 1989) is based on the relationship between electrophoretic mobility of single stranded DNA and its folded conformation (secondary structure), which depends on the nucleotide sequence. Different base substitutions might lead to different secondary structures because of different intramolecular interactions. In this method the fragments of interest are amplified by PCR and the double stranded DNA is first denatured and, then, electrophoresed in a nondenaturing polyacrylamide gel so that the single strands fold and migrate in a sequence dependent manner. So, variation in a single nucleotide within the sequence of single strand, will present a different three-dimensional conformation and show unique electrophoretic mobility in which the SSCP test takes advantage of the fact that these different conformations run at different rates through an acrylamide gel (Nataraj et al. 1999). The disadvantage of this method is that the temperature must be maintained throughout the gel run. Moreover, the method sensitivity is observed to be decreased by increasing the length of the DNA fragment (600bp). The optimum fragment size run by SSCP is between 150-200bp for good resolution and high sensitivity (Cotton 1997).

2.1.2 Heteroduplex analysis (HA)

Heteroduplex analysis (H. A.) method is similar to the SSCP method and it depends on the conformation of duplex DNA in non-denaturant gel, which consist of complementary strands that have a single-base pair mismatch. Separation of homoduplexes (perfectly annealed complementary strands) from heteroduplexes (formed when the partially complementary strands from two allelic loci anneal imperfectly to form a double strand with a mismatch), the mismatch fragment tends to slow the mobility during DNA migration in gel electrophoresis, that is possibly due to a conformational change or the dragging effect of the bulge through the gel electrophoresis (Nagamine, Chan, and Lau 1989).

In PCR preparation for H. A. method, PCR products were denatured by heating and then allowed to slowly reanneal. In a heterogeneous DNA sample such as a tumour biopsy, both the wild-type and mutant DNA are amplified during PCR, and the heteroduplex formed by a post-PCR annealing step. On the other hand, the wild type and mutant sample can be amplified separately and then mixed together before denaturation and annealing reaction to form heteroduplex. There are two types of heteroduplexes; the first type is formed when the sequence is altered in the two DNA fragments in one or more single-base mutations. When the two strands are annealed during the PCR reaction the shape of the heteroduplex appears as a bubble and is recognised by its mobility in the gel. The second type is formed when the sequence is altered in the two fragments as a small insertion or deletion, the conformation of dsDNA appears as bulge heteroduplex. When formed DNA molecules are separated in non-denaturing polyacrylamide gel, heteroduplex and homoduplex exhibit distinct electrophoresis motilities (Nagamine, Chan, and Lau 1989).

2.1.3 Protein Truncated Test (PTT)

The protein truncation test (PTT) is a method that specifically detects mutations that lead to the termination of mRNA translation and subsequent protein truncation. The specific mutations that cause protein truncation are caused by frameshift mutations (small insertion/deletion mutations), non-sense mutations and splice site mutations. PTT is based on a combination of RT-PCR transcription/translation and selectively detects translation-terminating mutations. The first application of the PTT method was to detect point mutations in DMD-patients and carrier females (Roest et al. 1993). Certain mutations such as single base substitutions, which do not cause truncated protein, will not be detectable by the PTT method. The Protein Truncated Test method could analyse fragments of up to 2kb, which other methods such as SSCP could not (Hogervorst et al. 1995).

2.1.4 The Amplification Refractory Mutation System (ARMS)

The ARMS technique is a simple, rapid and reliable method for the detection of a known point mutation in a test DNA sample. The ARMS test is often used in situations where high quality and accuracy results are essential, such as the clinical diagnosis of inherited disease. This method is based on the observation that oligonucleotide primers with mismatched 3`-termini will not function as polymerase chain reaction (PCR) amplimers under appropriate conditions. This observation is used to develop ARMS primers which allow amplification of test DNA if the target sequence is present in the test sample, and do not allow amplification if the target sequence is not present. Target sequences which differ by a single point mutation can be easily discriminated (Newton et al. 1989). The ARMS assay for a DNA polymorphism consists of two complementary PCR reactions. The first contain an ARMS primer specific for target DNA containing the normal variant and will not amplify target DNA containing the mutant variant. The second reaction contains the mutant specific primers and will not amplify normal DNA, however, this method, ARMS, known to be a simple and economical SNP genotyping method involving a single PCR reaction followed by gel electrophoresis. High throughput genotyping could be achieved by combining the ARMS method with the MADGE technique (ARMS-PCR-MADGE) (Ye et al. 2001).

2.2 Denaturant methods

Separation techniques that are based on electrophoresis of DNA samples in denaturing gel for identification of single base changes in a segment of DNA are widely applied in many laboratories throughout the world for screening different genes. All the current methods such as, DGGE, TTGE, CDGE and dHPLC that are based on capillary electrophoresis and require no denaturant gel, are merged into one principle which involves partially splitting the dsDNA during electrophoresis. Our approach, melt-MADGE, operates using a similar principle.

2.2.1 Denaturing Gradient Gel Electrophoresis (DGGE)

The principle of conventional DGGE (Myers, Maniatis, and Lerman 1987) is based on the fact that the two strands of DNA molecule separate when subjected to heat and chemical denaturant, a gradient of urea and formamide. Separation of DNA strands by the denaturation is influenced by the disruption of hydrogen bonds between A-T and G-C nucleotides; the hydrogen bonds in a GC pair are more stable than in an A-T pair. Thus, the GC pair will melt at a higher temperature than the AT pair and at a slightly higher concentration of urea and formamide.

DGGE has been shown to detect differences in the melting domain of small fragments that differ by a single base change. PCR products are run on a polyacrylamide gel, which contains an increased gradient of the denaturant agent urea and formamide. As a result of complete dissociation of the DNA into a single strand during the run, an artefact GC-sequence is introduced to one end of the sequence, usually in the highest melting domain end to prevent complete dissociation of DNA molecules and create a non-melting sequence (Cotton 1997).

DNA molecules that differ by a single nucleotide show a different melting temperature compared to a wild-type DNA. A heteroduplex will always lie higher in the gel at the end of the electrophoresis than the wild-type homoduplex that migrates further into the gel. A single base substitution will result in a non-Watson-Crick pair at the site of substitution, while an insertion or deletion will result in a molecule containing one base looped out of the helix (Lerman and Silverstein 1987). Because the GC-clamped primers are expensive, DGGE could be performed with longer PCR products up to 1000bp. However, the advantage of increased melting domain numbers would be balanced by the disadvantages of decreased mobility of the amplicon. In addition, the mutation detection rate (the sensitivity) would decrease (Nollau and Wagener 1997).

2.2.2 Denaturing high-performance Liquid Chromatography (DHPLC)

DHPLC is a recently developed method of comparative sequencing based mainly upon heteroduplex detection. It is a sensitive method for detecting single base alterations as well as small insertions and deletions. Separation of DNA is achieved by its migration through a column, non-polar stationary phase, and the hydroorganic mobile phase, resulting in the formation of a positive surface potential. Retention of dsDNA molecules is achieved by electrostatic interaction between the positive charges on the surface generated by triethylammonium ions adsorbed at the stationary phase and the negative surface potential generated by the dissociated phosphodiester groups of the sugar-phosphate backbone of DNA (Wagner et al. 1999). The method has efficiency to screen DNA fragment at size range of 150 to 700bp with estimated sensitivity approaching 98%. Although dHPLC shares similar principles to the conventional DGGE method, it does not require an introduced GC-clamp (Wagner et al. 1999). However, the method does not reveal the nature of the mutation or the location which has to be established by sequencing of the fragment (Wagner et al. 1999).

2.2.3 Developing mutation screening method; Melting Point Analysis By Microplate-Array-Diagonal-Gel Electrophoresis (Melt-MADGE)

DNA **melt**ing analysis by **M**icroplate **A**rray **D**iagonal **G**el **E**lectrophoresis is the generic name for a series of high-throughput methods of *de novo* scanning for single-nucleotide variation, both base substitutions and frameshift mutations. The standard techniques to detect minor changes in DNA e.g., SSCP (Single-Strand Conformation Polymorphism) analysis, heteroduplex analysis and DGGE (Denaturing Gradient Gel Electrophoresis) analysis reveal all sequence changes, including those pathogenic and polymorphisms. However, they do not have the capacity for high throughput and low cost. Melt-MADGE technique (Figure 11.1) was designed by Ian Day in 1994 (Day et al. 1995a) and the first application of this technique was for analysis of the LDLR gene. I have developed a rapid and efficient test, Microplate-Array-Diagonal-Gel Electrophoresis (Melt-MADGE) for use with the breast cancer gene type I (*BRCA1*).

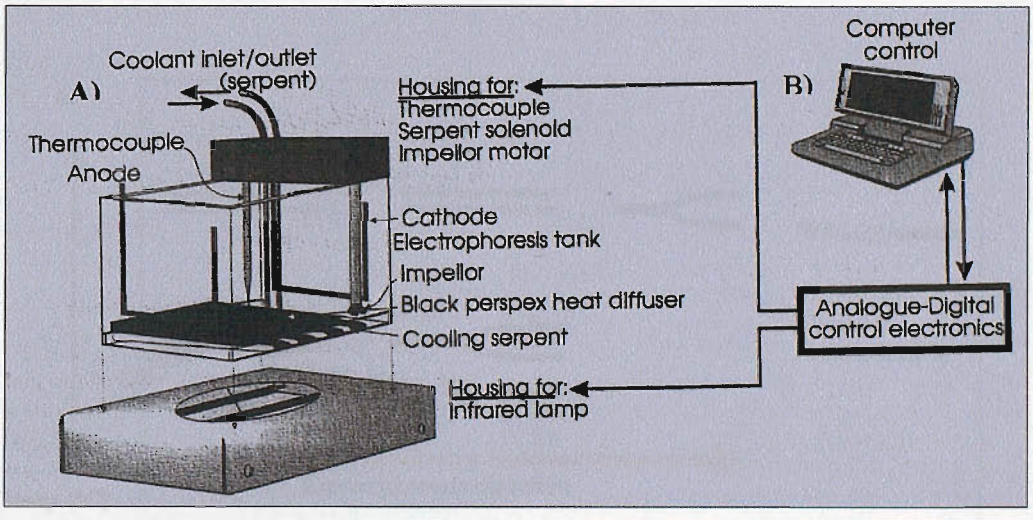


Figure 11.1. Illustrates the schematic of early constructed melt-MADGE apparatus. A) A cuboid tank to accommodate 2 litres of buffer and 10-12 gels. The tank is made of silica glass (heat resistant). It contains an impellor to ensure continuous stirring and even temperature in the tank; an infrared lamp below the tank and the black Perspex platform in the base of the tank to absorb infrared irradiation; a serpent cooling pipe passing through the base of the tank to enable cooling by a cooling recirculator; thermocouples in the tank for feedback of temperature, B) an analogue digital control boxinterfacing computer control by software with thermocouple feedbacks.

a) Melt-MADGE Theory

The theory behind melt-MADGE is based on the same theory of DNA thermodynamics as DGGE, CDGE and dHPLC. However, this technique allows the separation of DNA molecules only differing by single base changes that cause them to migrate differently in a denaturing polyacrylamide gel, which is subjected to an optimised ramping temperature; the polyacrylamide gel contains constant urea in homogenous buffer. The temperature at which the DNA duplex melts is influenced by the hydrogen bonds formed between complimentary base pairs and GC-rich regions that melt at higher temperatures than AT-rich regions and prevent complete strand separation (Figure 12.1). The two strands of a DNA molecule are expected to melt partially, because at one end of the double stranded DNA sequence a GC rich area is introduced known as, a GC clamp (Myers et al. 1985b) that is stable at higher temperatures. The other end of the double stranded DNA is more AT-rich so will partially split at a certain point along the gel at a lower temperature, which will be determined by the nucleotide sequence.

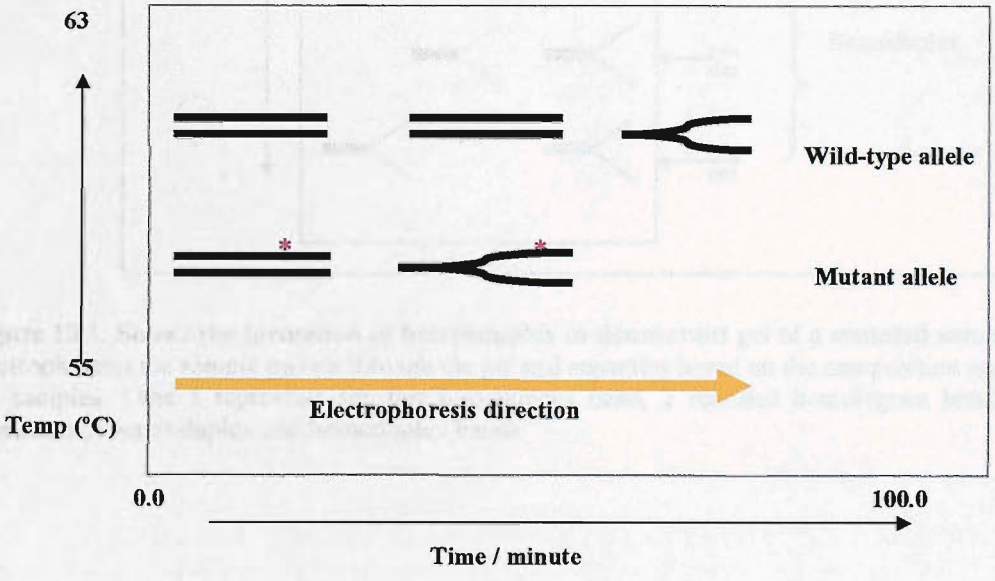


Figure 12.1. Schematic of mutation effect on the DNA molecules. The PCR product of wild-type and mutant allele differing by single nucleotide demonstrated different migration in the gel while subjected to denaturant conditions, temperature and urea.

When a double-stranded DNA segment is separated by electrophoresis, as at its melting point, it partially denatures and consequently slows down its rate of migration dramatically by intercalating with the gel matrix. Mutations will alter the amplicon's melting behaviour and cause mutant molecules to have an altered pattern of migration Figure 13.1.

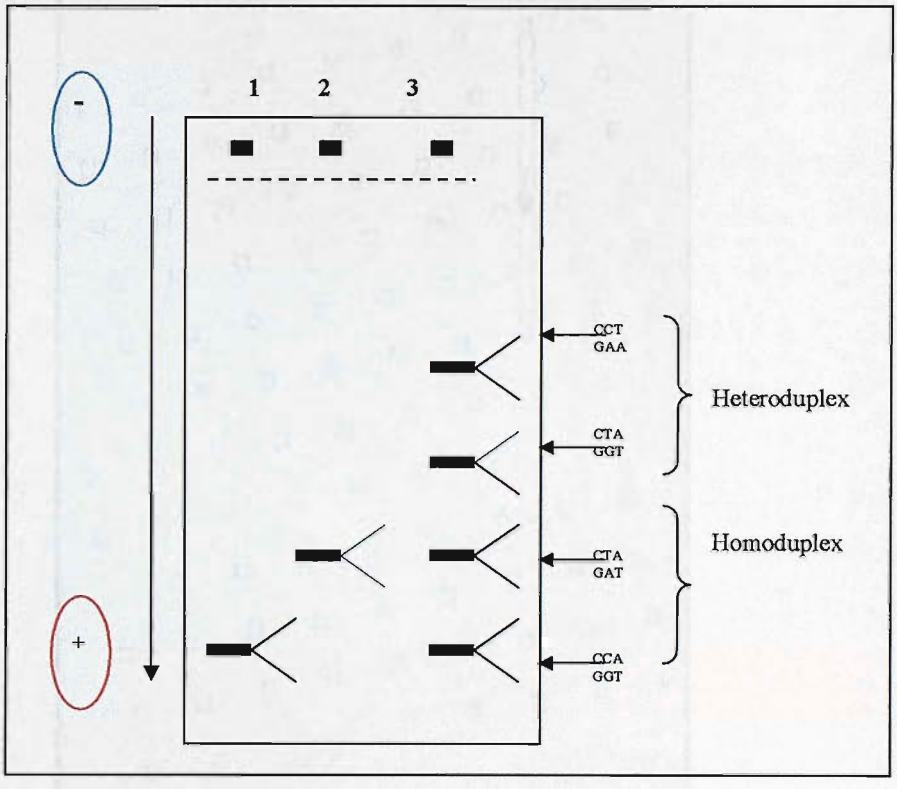


Figure 13.1. Shows the formation of heteroduplex in denaturant gel of a mutated sample. During electrophoresis the sample travels through the gel and separates based on the composition nucleotides in the samples. Lane 1 represents the fast homologous band, 2 retarded homologous band and 3 the formation of heteroduplex and homoduplex bands.

b) MADGE former

MADGE format is fully compatible with the industry standard 96-well microplates used for PCR. The polyacrylamide gel is anchored on glass and provides 96 26.5mm tracks by having the 8 rows of 12 wells (Figure 14.1).

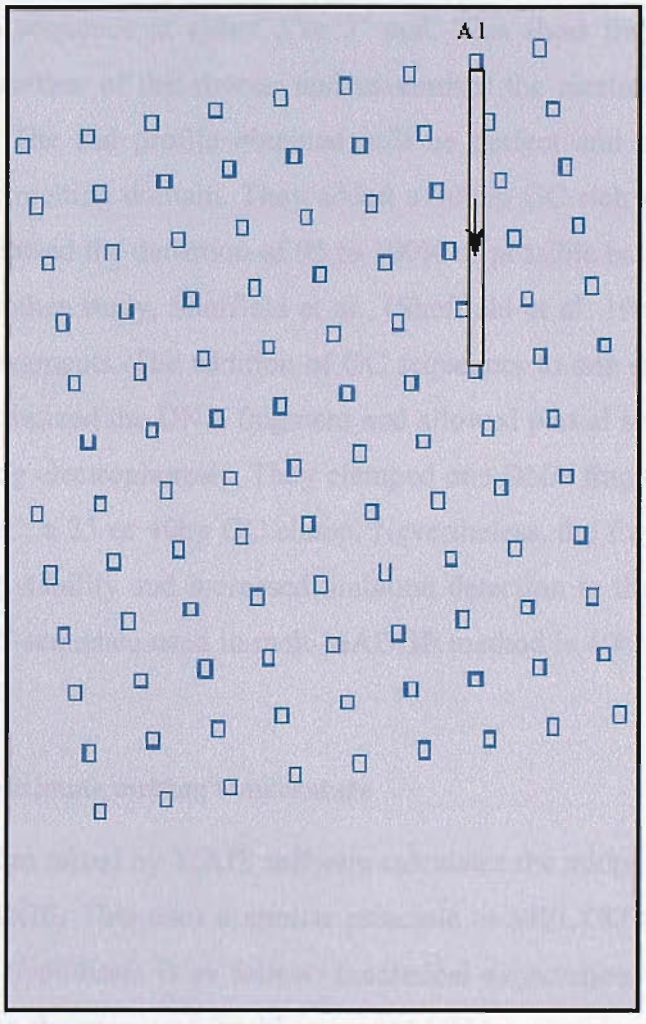


Figure 14.1. MADGE format an array of 96 wells compatible with PCR microplates. Using this template 10 gels offer 1000 tracks and can run in parallel in a 2L tank.

The combination of temperature control with MADGE format offers a low technology platform for the study of several categories of genetic markers. It also enhances our ability to explore single nucleotide variation at large- population scale where parallel analysis of many samples is required. The MADGE format permits rapid parallel inquiry of the same amplimer from many subjects at once in the prototype tank, which is made of 0.5cm thick polypropylene. Its dimensions are 23cm long, 11cm wide and 15cm high (Day et al. 1995b; Day et al. 1999b)

c) Introducing GC clamp to prevent dissociation of DNA strand

The old method, DGGE, can be used to distinguish two DNA molecules that differ by as little as a single base substitution. To increase the number of single base changes that can be distinguished by DGGE, Myers et al., 1985 (Myers et al. 1985b) introduced the idea of stabilizing one end of the DNA. In this the variable sequence needs to be clamped by implanting a GC-rich sequence at either 5`or 3` end. This short fragment of GC-sequence inhibits complete separation of the strands and maximises the electrophoretic retardation by partial DNA melting The flat profile obtained will be perfect and this enhances mutation detection in the lower melting domain. They added a 300bp GC rich sequence to a fragment sized 135bp, which allowed the detection of 95 to 100% of possible base pair alterations in the target fragment. In another study, Sheffield et al., (Sheffield et al. 1989) used PCR to attach GC-clamps to DNA fragments. The addition of GC sequences to one end of the 5` end of one of the PCR primers stabilized the DNA fragment and allowed partial separation of the dsDNA in the denaturant during electrophoresis. They clamped one DNA fragment with two different sequence lengths of GC, a 25 or 40bp GC clamp. Nevertheless, the fragment with a 40bp GC clamp displayed more stability and increased mutation detection in the first melting domain. Thus, the length of GC-sequence used in melt-MADGE method is 40bp attached to one of the primers.

d) Calculating the approximate melting temperature

The theoretical algorithm raised by TIXIS software calculates the midpoint temperature, *Tm,* of dsDNA for Melt-MADGE. This uses a similar principle to MELT87 that was developed by Lerman in 1987. The hypothesis is as follow: theoretical expectation of *Tm* is based on the knowledge that wherein the unwound double stranded DNA molecule each base pair is in one of two states; a stacked and fully intact duplex formed by hydrogen bonding, or a non-bonded single strand form. The intrinsic stability of each base pair is presumed to depend on its base pair type and the stacking interactions with neighbouring base pairs. A melting map exhibits *Tm* for each individual base pair in a DNA segment indicating that the melting of nearby base pairs is closely coupled over substantial lengths of the molecule despite their individual differences in stability. The expectation of the midpoint temperature, *Tm*, at which each base pair is in 0.5 equilibrium between the helical and melted form of the dsDNA fragment is determined by repetition of the probability calculation at a closely spaced series of temperature steps (Lerman and Silverstein 1987).

2.5 Cost-effectiveness for screening mutations in *BRCA1* gene

Most genetics analyses are costly and time-consuming. In such large genes as *BRCA1* with many mutations scattered throughout the entire coding sequence, screening for mutations and/or polymorphisms in large populations is extremely difficult. A recent study by Sevilla C. et al., 2002 (Sevilla et al. 2002) compared techniques for cost-effectiveness. They used the *BRCA1* gene as a model for this and applied it in a single country, France, within three biomedical laboratories. In table 1.1, seven different techniques were used to analyse the *BRCA1* fragments, and the cost per fragment of *BRCA1* gene was estimated. The number of fragments used in this context ranges from 25 to 35 fragments. 14 fragments of exon 11 with 21 small exons of the *BRCA1* gene analysed using DS, dHPLC, SSCP and DGGE techniques, and 6 fragments of exon 11 with 21 exons using the HA method, while exon 11 fragmented into 4 segments and was analysed by FAMA and PTT together with the 21 other exons of the *BRCA1* gene (Sevilla et al. 2002).

Cost factor	DS	DHPLC	SSCP	DGGE	HA	FAMA	PTT
Consumable supplies	16.7 €	1.4 €	3.6 €	1.3 €	1.9 €	7.90 €	7.2 €
Equipment	5.20 €	1.6 €	0.3 €	0.2 €	0.1 €	6.70 €	0.2 €
Personal	7.20 €	0.9 €	2.3 €	3.0 €	1.0 €	12.6 €	1.7 €
Total cost	29.1 €	3.9 €	6.2 €	4.5 €	3.0 €	27.2 €	9.1 €

Table 1.1. Shows cost comparison of different techniques for individual fragments of *BRCA1* gene. Adapted from Sevilla C. 2002 (Sevilla et al. 2002).

The estimated cost per fragment for the complete genetic testing process, from receiving the blood samples to releasing the genotype results is showing in table 1.1. The estimated cost in the above table for dHPLC, SSCP, DGGE, AH, FAMA, and PTT did not cover DNA sequencing cost for the genotype variant that might detected. However, the maximum numbers of the fragments tested per year are limited, in which 7378 fragments equivalent to only 86 patients are screened annually by direct DNA sequence (DS). Full gene sequencing works out as the most expensive screening method. In addition, HA represents the highest number of fragments to be screened per year, in which 29295 fragments equivalent to 473 patients can be analysed (Sevilla et al. 2002). Consequently, screening of large genes, e.g. *BRCA1*, in large numbers of pateints is limited by the expense and the time consumed in performing labour intensive genetic analysis.

The hypothesis and development

The working hypothesis of this study is that many different rare mutations may be acting through similar or common biochemical pathways to create the same phenotype. A complex disease like breast cancer, although relatively common in the global population, may be caused by a large number of different mutations each of which is quite penetrant but rare. H-ras gene, one of the ras family Oncogenes, showed a significant but low risk mutation for breast cancer associated with rare variant at a minisatellite locus (Krontiris et al. 1993). Thus, mutation in *h-ras* gene was found to be less frequent in human disease compared with the *k-ras* gene. This could be due to functional diversity between both genes and *k-ras* being more susceptible to mutation than *h-ras*. As yet, this has not been fully explained (Ellis and Clark 2000). The "common disease rare variant" hypothesis is an alternative to the currently popular "common variant common disease" hypothesis, built on the paradigm of Alzheimer's disease. The latter is strongly associated with a common polymorphism, however, no common polymorphism has been found to be associated with breast cancer. A large amount of evidence suggests that breast cancer has multifactorial aetiology with both genetic and environmental components. However, genetic factors seem to have a strong influence, at least in a subset of cases with a family history of disease. These patients frequently have mutations in one (or both) of two large genes, *BRCA1* and *BRCA2*. More than 500 mutations have been reported in *BRCA1*, but the vast majority of them have been found only once. Therefore, no statistically sound association has been established between each mutation and susceptibility to cancer.

To measure the contribution of any rare mutation to a phenotype, one must study very large population samples. Population studies for rare gene mutation are limited by the availability of a suitable population and the cost and time involved in mutation detection. Development of method of scanning the *BRCA1* gene for mutations, with the potential to process thousands of samples per day (melt-MADGE) would 1) much expand the potential of diagnostic laboratories and 2) enable definition of the full spectrum of population variation in *BRCA1* and more complete investigation of genotype-phenotype correlations. *BRCA1* and *BRCA2* gene mutations in hereditary breast cancer represent examples of the "rare disease-rare variants" hypothesis, in which rare mutations (often protein truncating and often unique to a family) account for a rare form of a disease. This contrasts with the "common disease-common variants" hypothesis in which common polymorphisms give a small increase in disease risk, as exemplified by heterozygous carriers in population breast cancer susceptibility. It remains less certain what genotype-phenotype relationships might exist for other mutations (and polymorphisms) in genes such as *BRCA1* and *BRCA2* in influencing features such as breast

cancer heritability, mean age of onset, human characteristics and population attributable risk. Practical problems to explore these questions, and even to characterise cases currently submitted to diagnostic laboratories, concern availability of suitable phenotype collections and cost and efficiency of mutation scanning or direct re-sequencing. This thesis is directed to establish approaches, which would facilitate a broader evaluation of the mutational spectrum in relation to relevant phenotypes. Additionally, such development (both methods and generated) would be of practical diagnostic importance.

Specifically, the work has involved development and application of a new technique, melt-MADGE for mutation scanning of the whole of the *BRCA1* gene. A range of performance comparisons has been made with other established approaches to mutation scanning, forming the basis to undertake genotype-phenotype studies and diagnostic mutation scanning on a much larger scale than was previously possible.

CHAPTER TWO
MATERIALS AND METHODS

2.1 Materials

2.1.1 DNA samples

a) Standard DNA samples

Twenty in-house DNA samples were used for the initial work to standardise and test the assay. These DNA samples (from anonymised male and female volunteers) were taken from the DNA bank in the Human Genetics division, University of Southampton, and had been equalised to 10ng/μl by Tricia Briggs.

b) Working DNA samples

Samples were collected from Breast or Ovarian cancer patients. These samples were collected from the Wessex Regional Genetics Laboratory at Salisbury. Samples were equalised to 10ng/μl by using picogreen method:

- 6 test DNA samples from patients with previously identified mutations in BRCA1 (anonymous)
- 200 DNA samples (anonymous) from breast cancer patients with a family history of breast and / or ovarian cancer.

2.1.2 Polymerase chain reaction (PCR) materials

a) Oligonucleotides

1. Primers for Long-distance DNA amplification:

   Nine pairs of primers were designed to give a set of long PCR amplicons spanning the *BRCA1*gene.The primers were ordered from MWG-Biotech (http/www.mwgdna.com), Appendix C.

2. Primers for short DNA amplification:

   Sixty-one pairs of primers were designed to cover all the coding regions of *BRCA1* gene, the non-coding region and some flanking intronic sequence. The primers were ordered from MWG-Biotech (http/www.mwgdna.com) Appendix A.

3. Primers for Amplification Refractory Mutation System (ARMS).

   Five sites of common polymorphisms in *BRCA1* gene were requested for ARMS amplification, and the primer sets were designed according to the mismatches in the polymorphism site. The primers were ordered from MWG- Biotech (http/www.mwgdna.com), Appendix B.

b) Enzymes

1. *Taq* DNA polymerase (5u/μl), supplied by GIBCO BRC, Life Technologies, cat. # 18038-026. Taq DNA was used to amplify short amplicons.
2. *Pwo* DNA polymerase supplied by Roche Diagnostics, Lewes, UK. The Pwo Taq enzyme was used to amplify long-distances of DNA with Taq DNA polymerase in a PCR reaction.

c) Chemical materials for PCR reaction

The following chemicals were used for preparing short PCR reaction.

- **D**eoxy**N**ucleotide **T**ri**P**hosphates (dNTPs), 40μmol each were obtained from Promega, Cat # U1240.
- $MgCl_2$ (5 – 3 mM)
- 1M KCl supplied by BDH, Poole, UK.
- 1M Tris pH8.3 was supplied by USB, cat. No. US75825, Ohio, USA.
- 0.01% gelatine was supplied by Amersham Life Science, Buckinghamshire, UK.

The following chemicals and reagents were used for L-D PCR amplification:

- Long PCR Buffer
- 10mM dNTP, Promega, Cat # U1240.
- 100Mpmol/ul primers, MWG-Biotech, (http/www.mwgdna.com)
- 50mM $MgCl_2$, provided with Taq DNA Polymerase from Promega.
- 5M Betaine, Sigma-Aldrich Company Ltd., Poole, UK.

2.1.3 Accessories of melt-MADGE Kits

- MADGE format, 96-well plate (MadgeBio, Grantham, UK).
- Skirted Thermo-Fast 96, 0.2ml Skirted 96 tube plate, Abgene House, (www.abgene.com).
- Omni Seal mat. Hybaid Omniseal TD mat, Hybaid Limited, cat. No. HB-TD-MT-SRS-5.
- Glass Plates were cut from standard float glass to 110mm x 170mm and were obtained from a local glass merchant (Shirley Glass, Shirley, Southampton).
- Stationery rubber bands,
- Flexible Silicon rubber tubes, the diameter is 1mm internal and 2mm external.
- 2 litre prototype Tank for 12-gel plates capacity (23x11x15cm tank carrying the electrodes) was provided together with a removable gel-track by the Medical Electronics and Engineering Department at Southampton General Hospital.

- Water-circulating glass coil (serpent) was provided by the Chemistry Department at the University of Southampton.
- Programmable heating/cooling water bath circulator (model RT111, Neslab Inc., Runcorn, UK).
- Digital thermometer (Tracker 120, Data Track instruments, New Milton, UK).
- 200Volts/1Amps output power supply, BIO RAD laboratories, inc., CA, USA

2.1.4 Chemicals and Regents for preparing denaturant gels

- Urea, BDH Laboratory supplies, Product No. 44387 7J, Poole, UK.
- Acrylamide-bis-acrylamide solution (19:1) stock solution, Seven Biotech Ltd., Worcestershire, UK.
- Ethylenediaminetetra-acetic Acid disodium salt (Na-EDTA), BDH Laboratory supplies, Product No. 100938B, Poole, UK.
- Tris base, BDH, Poole, Dorset, UK
- Glacial Acetic Acid, 100%. BDH Laboratory supplies, Product No. 10001A5, Poole, UK.
- Ammonium PeroxodiSulphate (APS), BDH Laboratory supplies, ANALAR 100323W, Poole, UK.
- NNN`N-TetraMethylEthyleneDiamine (TEMED), BDH Laboratory supplies product No. 443083G, Poole, UK.

2.1.5 Chemicals for staining the gel

- SYBR Gold Dye, Molecular Probes, Eugene, OR, USA.
- Ethidium bromide stock solution obtained from Gibco BRL. Life Technologies, Paisley, Scotland, UK.
- Vistra Green Dye, Molecular Probes, Eugene, OR, USA

2.1.6 Common Chemicals and materials

- γ-Methacryloxypropyltrimethoxysilane, SIGMA, EEC No., 219-785-8.
- Ethanol (CH3CH2OH), BDH, Poole, Dorset, UK.
- Sodium Hydroxide pellets (NaOH), BDH, Poole, Dorset, UK
- Microcentrifuge tubes (1.5ml), Alpha Laboratories Ltd., Hampshire, UK.
- Disposable pipette, Alpha Laboratories Ltd., Hampshire, UK.

- Falcon microarray plates (96-well), from Becton, Dickinson Labware, Becton, Dickinson and Co., California, USA.
- Plate sealers supplied by B.I.S. Ltd., Kirkham, Lancastershire, UK.
- Multiple channel Finnpipette (8 channels), supplied by Labsystem, Helsinki, Finland.
- Gilson pipetman pipettes (1000μl, 200μl, 100μl and 20μl)
- Disposable tips were supplied by Life Science International Ltd., Basingstoke, Hampshire, UK.
- 96-pin passive replicator (MADGEBio).

2.1.7 Instruments

- Flourimager (Model595, Molecular Dynamics Amersham Pharmacia Biotech, Little Chalfont, UK).
- Thermal Cycler PCR, DNA Engine Tetrad. MJResearch, INC., USA.
- Centrifuge, International Equipment Company (IEC), centra® MP4, Model No. 23690, USA.
- Microcentrifuge, 1.5ml tube, International Equipment Company (IEC), centra® MP4, Model 230, USA
- ABI PRISM 377 DNA Sequencer supplied by PE Applied Biosystems.

2.1.8 Buffers and Solutions.

a) 10X L-PCR Buffer

Long PCR buffer was prepared using 140mM Ammonium acetate $(NH_4)_2 SO_4$ (Sigma-Aldrich Company Ltd., Poole, UK), 500mM Tris-HCl (BDH) pH 8.9.

b) 10X T.A.E. stock buffer (2.5L. dH2O)

TAE buffer was prepared as 10x concentrated stock that was diluted as required.

- 121g Tris was dissolved in sufficient dH_2O.
- 9.3g Na-EDTA.
- 28.5ml glacial Acetic Acid was added to the solution and the volume was brought up to 2.5L of dH_2O.

c) 10X Polymerase mix (20ml)

Reagents	volume required	final concentration
1M KCl	10ml	0.5M
1M Tris pH 8.3	02ml	0.1M
0.1% gelatine	02ml	0.01%
2mM dNTPs (set)	1.6ml	0.2mM
Distilled water	4.4ml	____

d) 1M KCl solutions

A 7.45g was dissolved in 100ml dH_2O.

Polymerase mix was used at a 1x concentration in all standard PCRs.

e) 1M Tris pH 8.3

121g was dissolved in 100ml dH_2O, and adjusted pH to 8.3 with concentrated HCl,

f) 0.1 % gelatin solution

0.5g of gelatin was added to 50ml of distilled water, and then 5ml was added into 45ml of distilled water to give the total volume 50ml of 0.1% solution.

g) 5M Betaine

A stock solution of Betaine ($C_5H_{11}NO_2$), Sigma-Aldrich Company Ltd., was prepared at 5M, and stored at 4C for up to six months. 58.55g was dissolved in 100ml distilled water.

h) Sticky Silane

2.5ml (0.5%v/v) γ-methacryloxypropyltrimethoxy-silane,

2.5ml (0.5%v/v) glacial Acetic Acid were added to 500ml Ethanol

i) SYBR Gold / Vistra Green staining preparation

12μl was added into 100ml of 1xTAE buffer

2.2 Methods

2.2.1 Construction of melting profile and primer design

The Melt-MADGE method is relatively simple once optimised. However, significant steps of preparative work must be undertaken before using the technique to screen for alterations in a candidate gene. In designing primers for Melt-MADGE, it is necessary to consider the length and the GC/AT ratio. In addition a GC clamp is usually placed at either 5'or 3'end of the amplicon. The melting profile for each of the fragments used in this study was constructed using the TIXIS program developed by Emmanuel Spanakis, University of Southampton. This software is a modification of the *MELT 87* algorithm that was previously devised by Leonard Lerman. TIXIS calculates the theoretical melting profile of a known DNA sequence

2.2.2 PCR amplification

a) Long-distance amplification PCR

Long distance PCR pre-amplification from genomic DNA provided the template for the amplification of the 39 target fragments of the *BRCA1* gene (Figure 15.2). Primers for Long PCR were designed by using primer3 (http://www.genmewi.mit.edu/cgi-bin/primer/primer3_www.cgi) with requirements set by input of parameters such as length (between 25-34bp) and theoretical melting temperature (between 68°C-72°C). Some of the Long PCR primers within specific regions in the gene were hard to design because of the *Alu* elements in those regions. Those Long PCR primers were requested and provided kindly by Nathalie J van Orsouw (van Orsouw et al. 1999), at the Institute for Drug Development, Cancer Therapy and Research Centre, USA. Primer sequences are shown in Appendix (C).

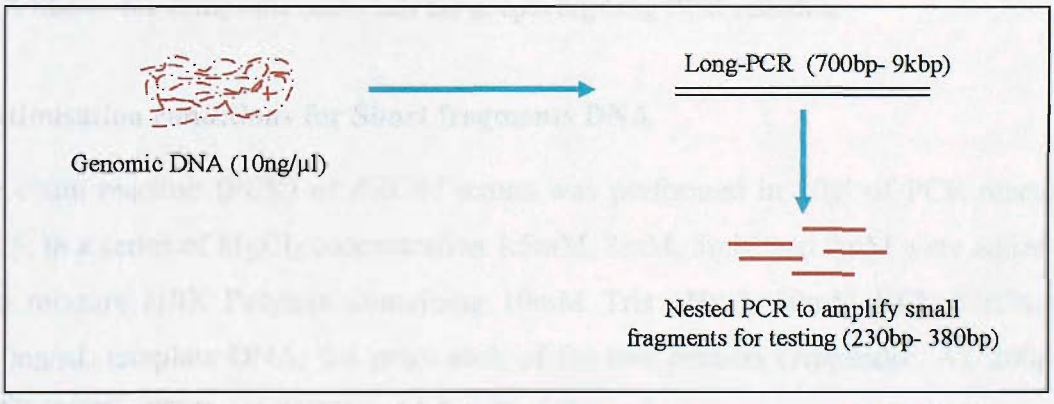


Figure 15.2. Long PCR allows us to make much more DNA from a small stock for testing

PCR reaction mixture per sample reaction as follow

Long PCR amplification was performed in 20 μl PCR reaction by the protocol that Ms Lesley Hinks developed (University of Southampton, Human Genetics Division) (table 2.2). 2 μl of genomic DNA template (10ng/ μl) was aliquoted into a 96-well PCR plate and placed onto a thermal Cycler (DNA engine Tetrad) for 10 minutes at 80°C for dryness. 20 μl of PCR reaction mix was added to each of the DNA samples. The plate was then sealed with an adhesive sealing sheet, and placed onto a thermal cycler (DNA engine Tetrad). PCR cycling program started with an initial denaturant temperature step at 94°C for 2 minutes, then followed by 35 cycles, denaturing step for 20seconds at 94°C, gradient annealing temperature for 20seconds at 68°C and extension at 72°C (1 min/ 1kb product length). Then followed by a final extension at 68°C for 20 minutes. The long PCR products were then electrophoresed in 1% agarose gel in TBE buffer to check the long amplicons in parallel with a 1Kb size ladder.

Ingredient	Concentration	Volume required
H_2O	-----	10.74 μl
Long PCR Buffer	10x	2.0 μl
10mM dNTP	250uM (10ul of each in 100ul)	0.5 μl
100pmol/ul primers	400nM	0.08+ 0.08 μl
50mM MgCl	2mM	0.8 μl
5M Betaine	1.3M	5.2 μl
Gibco Taq	1.0unit	0.2 μl
1/250 Pwo	0.008units	0.4 μl

Table 2.2. Shows the composite materials for preparing long PCR reaction.

b) PCR Optimisation conditions for Short fragments DNA

Polymerase chain reaction (PCR) of *BRCA1* exons was performed in 20μl of PCR reaction tube. Initially, in a series of $MgCl_2$ concentration 1.5mM, 3mM, 5mM and 7mM were added to the reaction mixture (10X Polymix containing 10mM Tris pH8.3, 50mM KCl, 0.01%w/v gelatine), 10ng/μL template DNA, 0.4 pmol each of the two primers (Appendix, A), 200μM each of dATP, dCTP, dTTP and dGTP and 1.0 unit of *Taq* polymerase.

Sample preparation and PCR amplification conditions

PCR Master Mix per one reaction:

Ingredients	Volume (μl)
Deionised H_2O	12μl
3-5 mM titration of $MgCl_2$	2μl
(10x polymix)	2μl
PCR primers 0.4pmol (Forward and Reverse)	2μl
Taq DNA Polymerase	0.1μl

Table 3.2. Shows the reagents for amplification of the short PCR amplicons.

18μl of PCR Master Mix was aliquot into the 96-well plate, and then 2μl of each DNA template (10ng/μl) was added to the wells to make the final volume 20μl. The plate was sealed with the adhesive sealing sheet using Omni Seal mat and then placed on a thermal cycle (DNA engine Tetrad, MJ Research). PCR cycling program started with an initial denaturant temperature step at 94°C for 3 minutes, then followed by 30 cycles denaturing steps for 30seconds at 94°C, gradient annealing temperature for 30seconds at 55°C – 75°C and extension for 30seconds at 72°C. Then a final extension step at 72°C for 3 minutes was added.

2.2.3 Setting up the conditions for the electrophoresis

The denaturising conditions that are used for Melt-MADGE analysis of *BRCA1* gene fragments are temperature and urea. The urea concentration was between 4M and 6M, which was dependent upon the melting behaviour of the fragment. A urea concentration of 4M is optimal for fragments with melting temperatures in the melting profile found to be between 55°C and 63°C. A 1M urea concentration will reduce the melting point by approximately 2.0°C. The gel contains polyacrylamide / bisacrylamide (39:1) solution in 1x TAE buffer. Its concentration is dependent on the amplicons fragment size. We found a 5% polyacrylamide gel is optimal, which satisfied the clear resolution of the fragments between 350bp to 450bp, while 6% polyacrylamide gel is optimal for the fragments 200bp to 350bp in length. The addition of NNN`N-TetraMethylEthyleneDiamine (TEMED) and Ammonium PeroxodiSulphate (APS) to the mixture will catalyse the acrylamide to polymerise. (APS) has an additional role as it accommodates the ionic strength in the gel as it appears in the buffer during the electrophoresis.

a) Assembling glass plate-gel former

A glass plate is wiped with sticky silane. This is important for the polyacrylamide gel to adhere firmly to the glass. The plastic gel former is then overlaid with the glass plate. The gel mix is carefully poured onto the plastic-gel-former-glass-plate assembly to avoid creation of air bubbles within the gel. A 250g weight is then placed on the plate to maintain pressure between plate and teeth. The gel is allowed to set for a minimum of 30 minutes and maximum of two hours. After the gel had set, the glass plate was removed, bearing its open-face 96-well gel. The sample volume loaded is 2µl. Using a 96-channel pipette, the 96 samples can be loaded at once. The well must be filled with dH_2O. The distilled water is added on the side of the gel and the second clean (non-silanised) glass plate is carefully slid along the gel, because air trapped in wells will probably lead to reduce band resolution. The glass-gel-glass assembly is secured with two stationary rubber bands. The two long sides of the gel are sealed with long pieces of flexible silicon-rubber tubing, stretched and inserted through the gap between the glass plates and released to seal the gel. This will insure the bands in the gel run straight along the gel (Spanakis E., Al-dahmesh M.A., and Day INM 2001) (Figure 16.2).

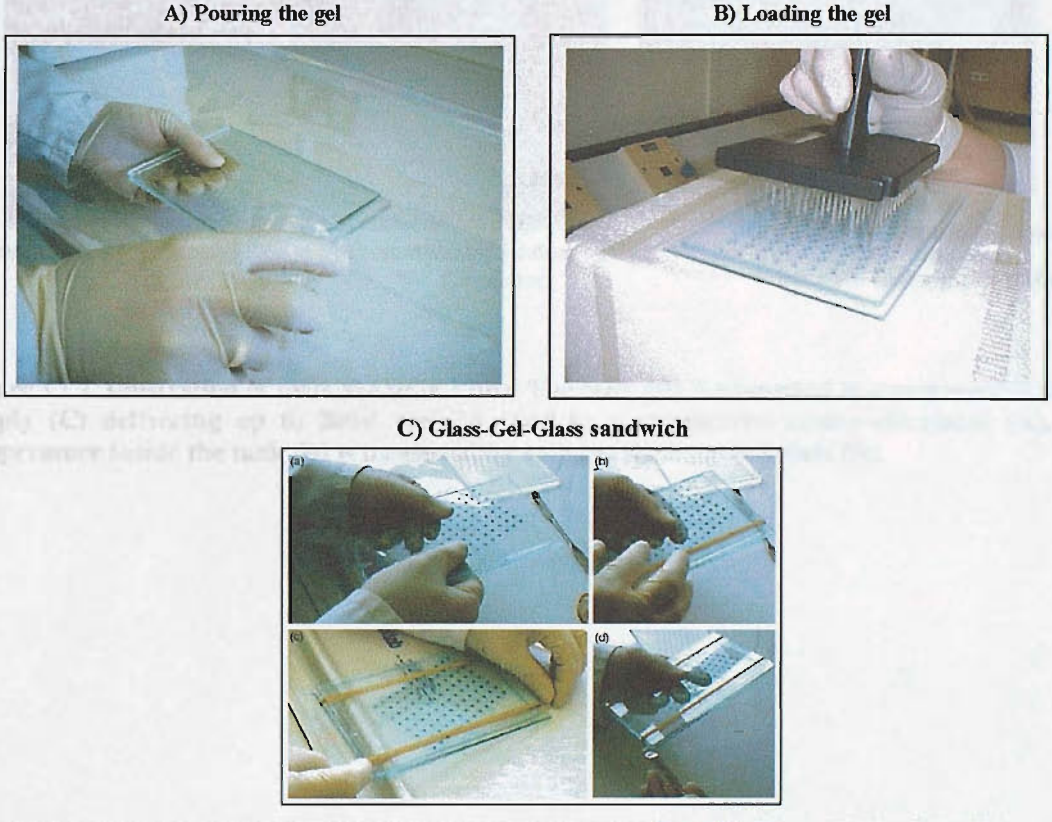


Figure 16.2. Illustrates melt-MADGE method procedure. A) the gel-former and the glass plate are held together and the gel is poured in the space between them. B) After loading, C) the gel is covered with a second glass plate (Ca). The glass-gel-glass 'sandwich' is secured with two stationery rubber bands (Cb). The long edges of the gel are sealed with ~15cm-long pieces of silicon rubber tubing (Cc,Cd).

b) Electrophoresis of PCR products

A small electrophoresis tank under good temperature control is required. Our prototype tank (Figure 17.2a) is 23cm long (anode to cathode), 11cm wide and 15cm high, with a 12 gel maximum capacity. It has two platinum electrodes, a propeller, a water-circulating glass coil and a removable gel rack. The tank is made of 0.5cm thick polypropylene. The electrodes are connected through the cover of the tank to a commercial 200V, 1A power supply (Figure 17.2b). Perfect spatial homogeneity is achieved by vigorous stirring. The glass coil is connected to a programmable heating-cooling circulator (Figure 17.2c) that can produce a linear temperature ramp up to 70°C. measuring the temperature in the tank was performed by using a high precision digital thermometer (Figure 17.2d).

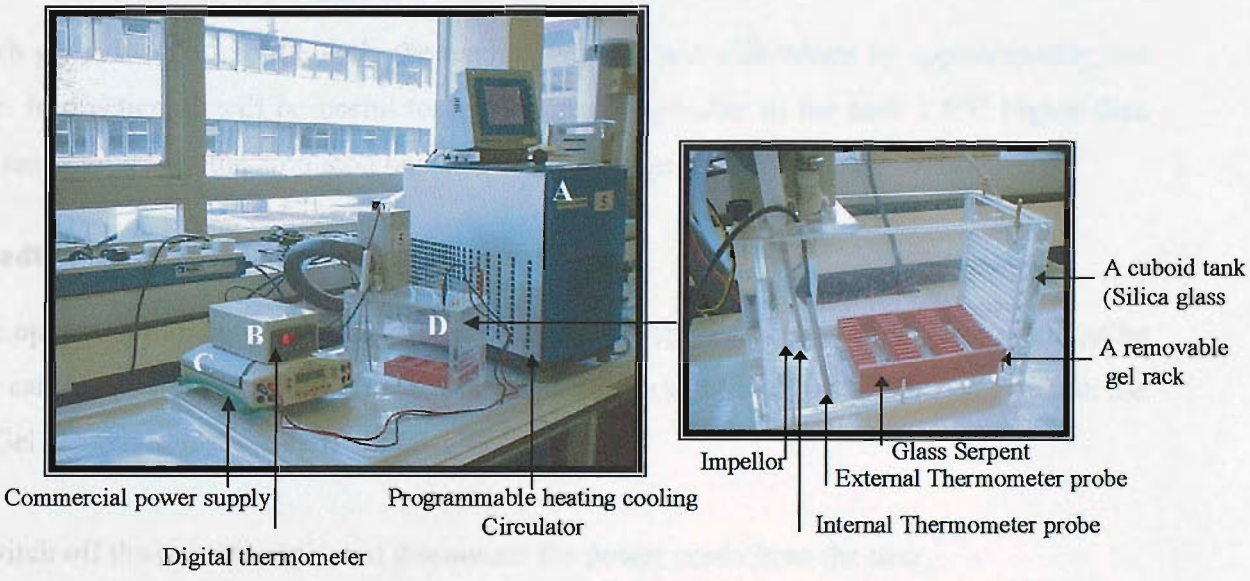


Figure 17.2. Illustration of melt-MADGE units. The tank (D) is connected to a commercial power supply (C) delivering up to 200V and 2A, and to a commercial cooler–circulator (A). The temperature inside the tank (D) is measured by a digital thermometer unit (B).

C) Pre-heating the running buffer

It is important to accommodate the running buffer to electrophoresis PCR products by preheating the buffer.

- Filling the electrophoresis tank, the prototype tank, with 2 litres of 1x TAE buffer.
- Replace the lid on top of the tank, attach the power cords to the anode and cathode on top of the tank, and turn the power, a propeller, and digital thermometer on.
- Setting the desire temperature by programmable heating-cooling circulator will maintain the temperature to an optimum reading. It could take 40 minutes to for the system to heat the buffer.

As each gel is loaded into the tank, the reading temperature will reduce by approximately one degree. In practice, it will be useful to heat the running buffer in the tank 1.5°C higher than initial temperature for the amplicon before lowering the gel in the tank.

D) Loading the glass-gel-glass in the prototype Tank

As the optimum temperature is approached in the prototype Tank, the following steps must be taken, carefully with attention to safety, because of the electric current and hot water in the tank. Gel electrophoreses is carried out at 50 Volts.

1. Switch off the power supply and disconnect the power cords from the tank.
2. Remove the tank lid.
3. 4ml 20% APS must be added to the 2L 1XTAE buffer, in order to equalise the ionic strength in the buffer and in the gel.
4. Immerse the glass-gel-glass sandwich in the tank.
5. Replace the lid on the tank and connect the power cords.
6. Switch on the electricity to the tank, adjusted to 50volts.
7. Run the stored program, by pressing start ramp key. When the run is finished, after 100 minutes, switch off the electric source.
8. Take the lid off and carefully pick up the glass-gel-glass sandwich out of the tank.

E) Temperature control

The temperature controller maintains the desired buffer temperature and controls the temperature ramp in the Melt-MADGE tank. The actual and set buffer temperatures are displayed in degrees Celsius. The temperature ramping is adjusted by inputting two steps:

1. Initial temperature to final temperature for one hour
2. Final temperature to final temperature for thirty to forty minutes (Figure18.2)

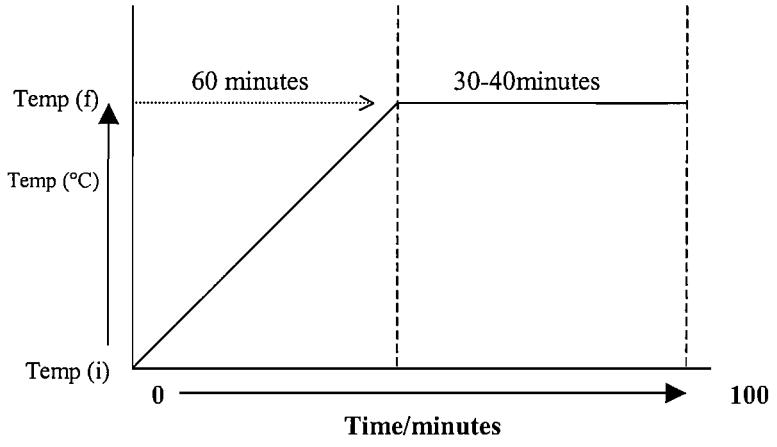


Figure 18.2. The diagram showed the two steps taken for controlling the ramping temperature during electrophoresis; i = initial temerature, f = final temperature.

F) Staining and imaging

After electrophoresis, the gel is washed with warm water for approximately 20 seconds to remove dissolved- Urea that might be adhesive in the gel, which could have an effect on the gel staining. Gels were stained in Vistra Green, which stains double strands in denaturant gels and give less background. 10-15 minutes staining is adequate to stain the gel in a dark room. The gel is then scanned on a Fluorimager 595 (Molecular Dynamic, Sunnyvale, CA.USA).

2.2.4 Developing of ARMS assay

a) PCR Optimisation conditions for ARMS assay

ARMS Polymerase chain reaction (PCR) of *BRCA1* exon 11 was performed in 10μl of PCR reaction tube. Initially, a series of $MgCl_2$ concentration 2.0mM, 2.5mM, and 3.0mM were added to the reaction mixture (1X PCR buffer, 1.3mM of betain, 0.2mM of dNTPs). The upper control primer was added at different concentrations at 0.05mM, 0.1mM, 0.2mM and 0.3mM. The allele specific primer and control primer were between 0.7 to 1.0mM (for primers sequence see Appendix B). Then 2μL of 10ng/μL template DNA was added to each well, and sufficient dH_2O was added to make up the total volume 10 μL and 1.0 unit of *Taq* polymerase was added. The plate was sealed with the adhesive sealing sheet using Omni Seal mat and then placed on a thermal cycle (DNA engine Tetrad, MJ Research).

b) Thermocycle of Polymerase Chain Reaction

The PCR cycling program started with an initial denaturant temperature step at 92°C for 3 minutes, followed by 30 cycles denaturing steps for 30seconds at 92°C, gradient annealing temperature for 30seconds at 55°C – 65°C and extension for 30seconds at 70°C. Then followed by a final extension step at 72°C for 3 minutes.

2.2.5 DNA sequencing

DNA sequencing was performed on an ABI PRISM^{TM} 377 DNA sequencer (Perkin Elmer) using DNA sequencing Kit, BigDye^{TM} Terminator Cycle Sequencing Ready Reaction Kit, Warrington, UK. DNA samples were processed out for sequencing in five steps:

1- PCR amplify DNA fragments.
2- Purify PCR products.
3- Cycle sequencing.
4- Purify extension products.
5- Electrophoresis on the ABI.

The primers for sequencing were chosen from those used for Melt-MADGE and the clamp was omitted. PCR reaction and cycling program for amplifying DNA fragments were set up as optimised for exons/fragments of *BRCA1* gene Section (2-21). Wizard® PCR Preps DNA purification System supplied by Promega was used to purify PCR products. The PCR reaction was performed in 20μl reactions. Each PCR reaction mixture contained 8.0μl Terminator Ready Reaction (A 2.5x dilution buffer for the Big Dye Terminator kit was prepared using 200mM Tris-HCl pH 9.0 and 5mM MgCl2. This was used to replace half of the recommended

quantity of Big Dye ready reaction mix), 5.0μl DNA Template (PCR purified products), 3.2μl working primer (3.2pmol/μl) and 3.8μl ddH_2O. The PCR cycling program started with an initial denaturant step at 96°C for 10 seconds, then annealing temperature at 50°C for 5seconds and extension at 60°C for 4mins for 25 cycles. A 1°C/second-ramp speed between steps of 4°C was set. Excess dye terminators were removed from PCR products and amplicons were saturated by using Ethanol / Isopropanol precipitation method as described in ABI PRISM™ BigDye™ Terminator Cycle Sequencing Ready Reaction Kit Protocol.

2.2.6 Statistic for haplotyping analysis

Estimation of haplotype frequencies from genotypic data in humans in the absence of family data is not direct because haplotype phase is not always known. Several methodologies permit the derivation of haplotypes by inferring phase, each one with its own advantages. For the sake of comparison, haplotype frequencies were estimated by the Expectation-Maximization (EM) algorithm using the Arlequin program (Schneider, Roessli, and Excoffier 2000), by Gibbs sampling using the Phase program (Stephens, Smith, and Donnelly 2001).

a) Arliquin method

The EM algorithm has been shown to estimate common haplotype frequencies with accuracy when Hardy-Weinberg equilibrium (HWE) is satisfied and when sample sizes are reasonably large (>100 chromosomes), although the EM algorithm can still give good results when HWE is not strictly satisfied (Fallin and Schork 2000; Tishkoff et al. 2000). To ensure that the global maximum likelihood estimate was found, the EM algorithm was run 100 times using different random initial conditions. The number of iterations was limited to 16000, a number large enough for guaranteeing the achievement of estimates differing by less than 1% with the estimates of maximum likelihood in 99% of cases (Guo and Thompson 1992).

b) Phase method

Haplotype frequencies were inferred (absolutely or to some predefined level of confidence) by Gibbs sampling, a type of Markov-chain Monte Carlo algorithm (Gilks, Richardson, and Spiegelhalter 1996), using the Phase program (Stephens, Smith, and Donnelly 2001). It has been suggested that this approach increases the accuracy in the estimation of haplotype frequencies, often reducing the error rates of the other method frequently used for haplotype reconstruction (expectation maximization algorithm) by >50% (Stephens, Smith, and Donnelly 2001). The predefined level of confidence was set at ≥ 90% and the numbers of iterations and burns-in performed were 10000, each iteration consisting of performing 100 steps through the Markov chain.

CHAPTER THREE

RESULTS

3.1 Computer generation of melting profile and primer design

Primers were designed in the intronic regions to amplify groups of exons in the *BRCA1* gene. Exon 11 and 24, due to their length, were split to 16 and 7 overlapping fragments respectively, to ensure the full coverage of the coding region. The length of the exons and fragments fall between 250bp-350bp (Figure 19.3).

The addition of a 40-60 base pair GC-clamp to one of the PCR primers was necessary to ensure that the region screened is in the lowest melting domain, and to protect dsDNA from complete denaturation. The temperature range for the Melt-MADGE can be calculated from the melting profile by determining the lowest and highest non-GC clamp melting temperature of the DNA sequence. From the calculated low and high temperature, adding urea to the gel can lower the theoretical melting temperature. A denaturant urea gel will lower the theoretical melting temperature of DNA by 2.5°C for every mole of urea per litre. Table 4.3 shows the melting temperature predicted by TIXIS program and melting temperature obtained by experimental results at constant temperature.

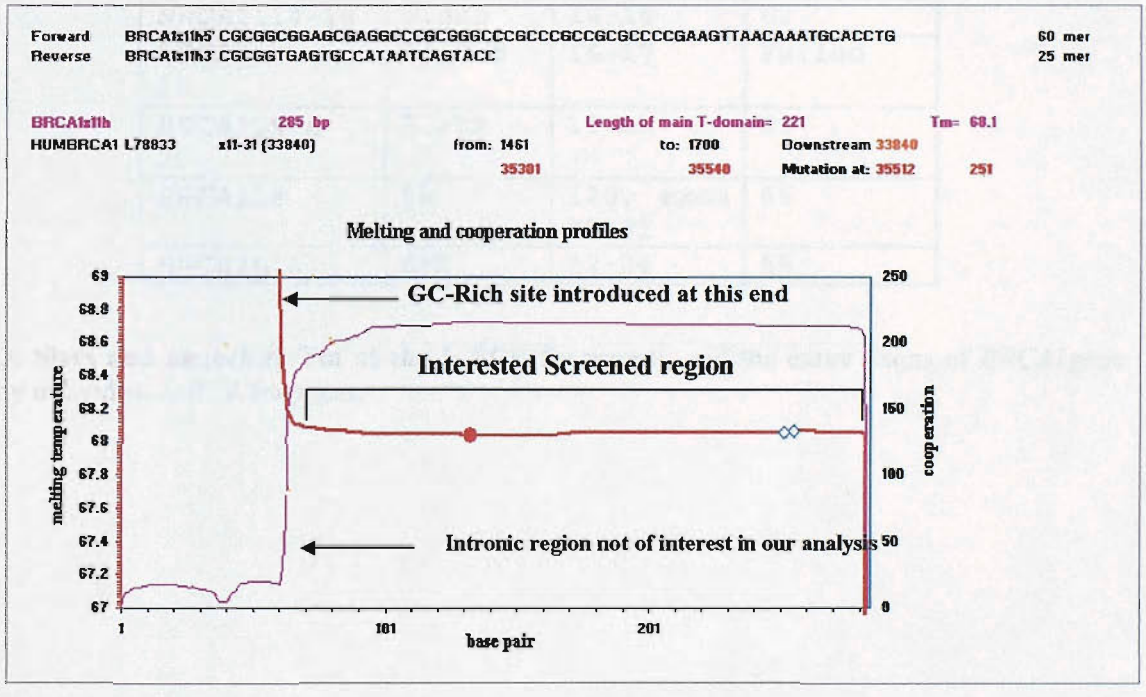


Figure 19.3. Attachment of G/C nucleotides generated by TIXIS program: Theoretical Melt map of 285bp fragment without GC-clamp at 3`-end (pink line) melt temperature varies across fragments compared with melt map of 285bp with GC-clamp (red line) showing steady melting profile.

3.2 PCR optimisation results

a) Long Distance PCR (L-PCR)

BRCA1 gene has 24 exons with 5711 nucleotides coding sequences that can be amplified by eleven long-distance PCR segments. Eleven long-distance PCR were tested to provide the templates from genomic DNA for the amplification of the 54 nested fragments.

The length of the L-PCR product varies between 700bp and 9kb (table 4.3). Because of the length of individual long sequences, two amplimers failed to amplify the targeted long sequence. Specific proof reading enzyme might be needed to improve this.

L-PCR ID	Length (bp)	Exons covered	Annealing Tm (°C)
*BRCA1*L1	3kb	1a, 1b, 2	68
*BRCA1*L2	700bp	3	63
BRCAL1N5-9	11.3kb	5-9	Failed
*BRCA1*LN10-11	4.8kb	10-11	65
*BRCA1*LN12-13	9kb	12-13	65
*BRCA1*L5	4.7kb	I12, exon12-13	65
*BRCA1*L14-16	5.8kb	14-16	63
*BRCA1*LN14-17	10.7kb	14-17	Failed
*BRCA1*LN18-20	7.2kb	18-20	65
*BRCA1*L8	8k	I20, exon 20-21	65
*BRCA1*L10	6kb	22-24	65

Table 4.3. Sizes and annealing Tm of the L-PCR fragments, and the entire exons of *BRCA1*gene covered by individual L-PCR fragment.

Electrophoresis L-PCR products:

L-PCR products were loaded in 1% agarose gel 1x T.B.E. buffer and electrophoresed for 60 minutes at 150volts. Figure 20.3 shows an example of L-PCR products covered by different regions of *BRCA1* gene.

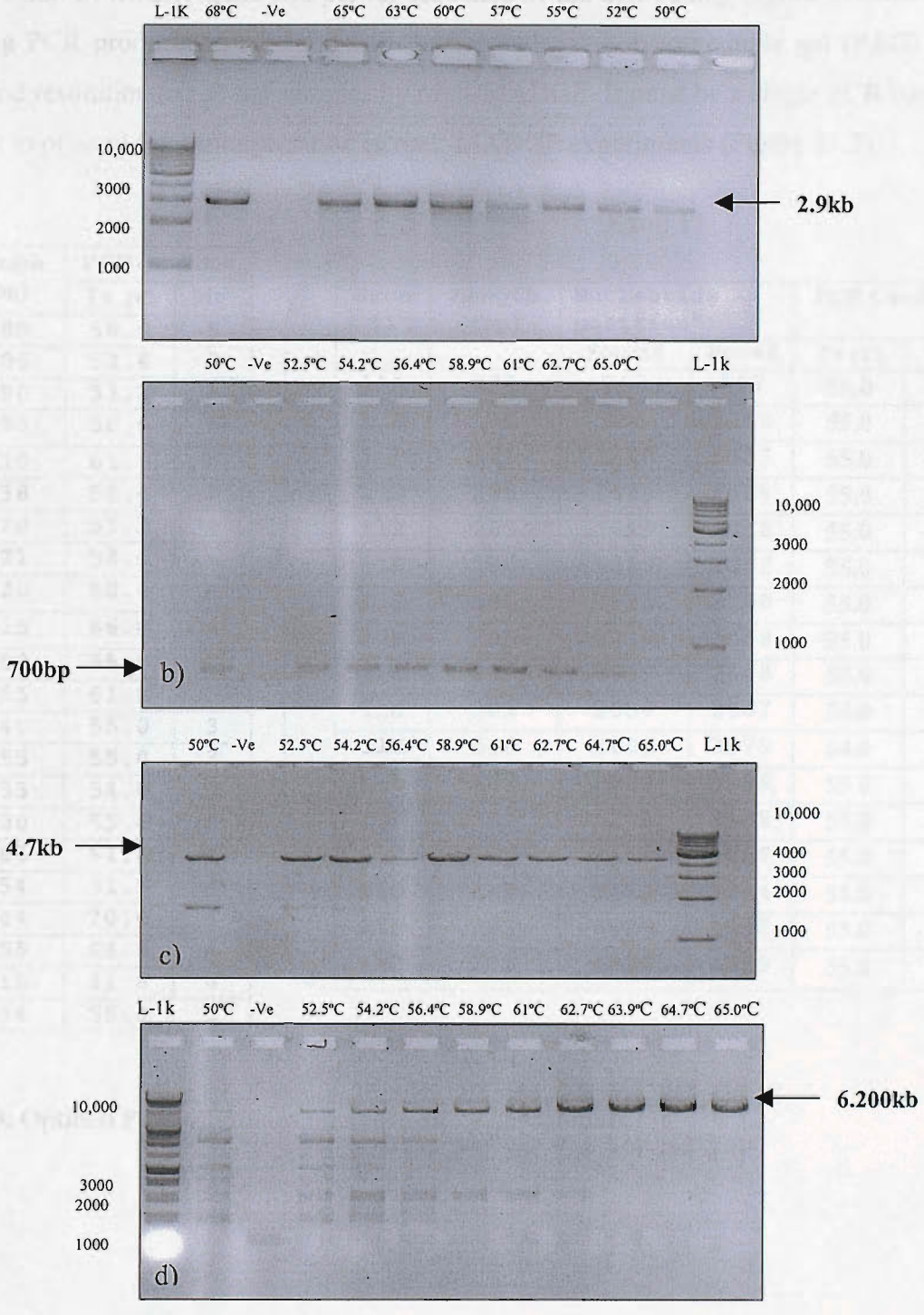


Figure 20.3. Examples of L-PCR products, which amplified template for nested PCR. a) L-PCR product covered exon 1a to exon 2 (2.9Kb). b) L-PCR product covered exon 3 of *BRCA1* gene (700bp). C) L-PCR product covered intron 12 contained GT repeat and exon 13, (4.7Kb). d) L-PCR product covered exon 22 to 24 of *BRCA1* gene (6195bp). *L-1k; ladder of 1kb step*

b) Short PCR products

Fifty-four amplicons were adjusted for the final magnesium concentration and annealing temperature. Table 5.3 shows PCR conditions chosen for individual amplicons of the entire coding region of *BRCA1* gene and partial sequence of the non-coding region in exon 24, 24g. Checking PCR products using 1X T.B.E. buffer and 5% Polyacrylamide gel (PAG) showed good band resolution to run the samples by Melt-MADGE. It must be a single PCR band in the gel track to prevent false interpretation in melt-MADGE experiments (Figure 21.3).

Exon #	Length (bp)	PCR condition	
		Ta (°C)	Mg^+
2	180	56.0	5
3	305	53.4	3
5	290	53.2	4
6	290	58.4	5
7	310	61.8	5
8	338	58.4	5
9	270	53.2	4
10	321	58.4	4
12	320	58.4	4
13	325	66.8	4
14	300	55.5	4
15	355	61.8	3
16	445	55.0	3
17	255	55.0	3
18	353	54.0	4
19	200	55.0	3
20	285	61.8	4
21	254	51.5	4
22	244	70.0	4
23	255	64.0	4
24A	318	61.8	4
24G	234	55.0	3

Exon 11

Exon 11	Length (bp)	Nucleotide position[a]		PCR Condition	
		Started	Ended	Ta (°C)	Mg^+
11A	277	768	997	55.0	3
11B	345	960	1258	55.0	3
11C	285	1220	1457	55.0	3
11D	285	1429	1658	55.0	3
11E	285	1619	1858	55.0	3
11F	285	1819	2058	55.0	3
11G	285	2021	2258	55.0	3
11H	285	2219	2458	55.0	3
11I	305	2399	2648	55.0	3
11J	285	2634	2867	55.0	3
11K	330	2802	3079	64.0	4
11L	287	3022	3299	55.0	3
11M	307	3203	3458	55.0	3
11N	285	3419	3657	55.0	3
11O	289	3619	3844	55.0	3
11P	285	3820	4057	55.0	3
11Q	265	4020	4239	55.0	3

Table 5.3. Optimal PCR conditions for short *BRCA1* amplimers.

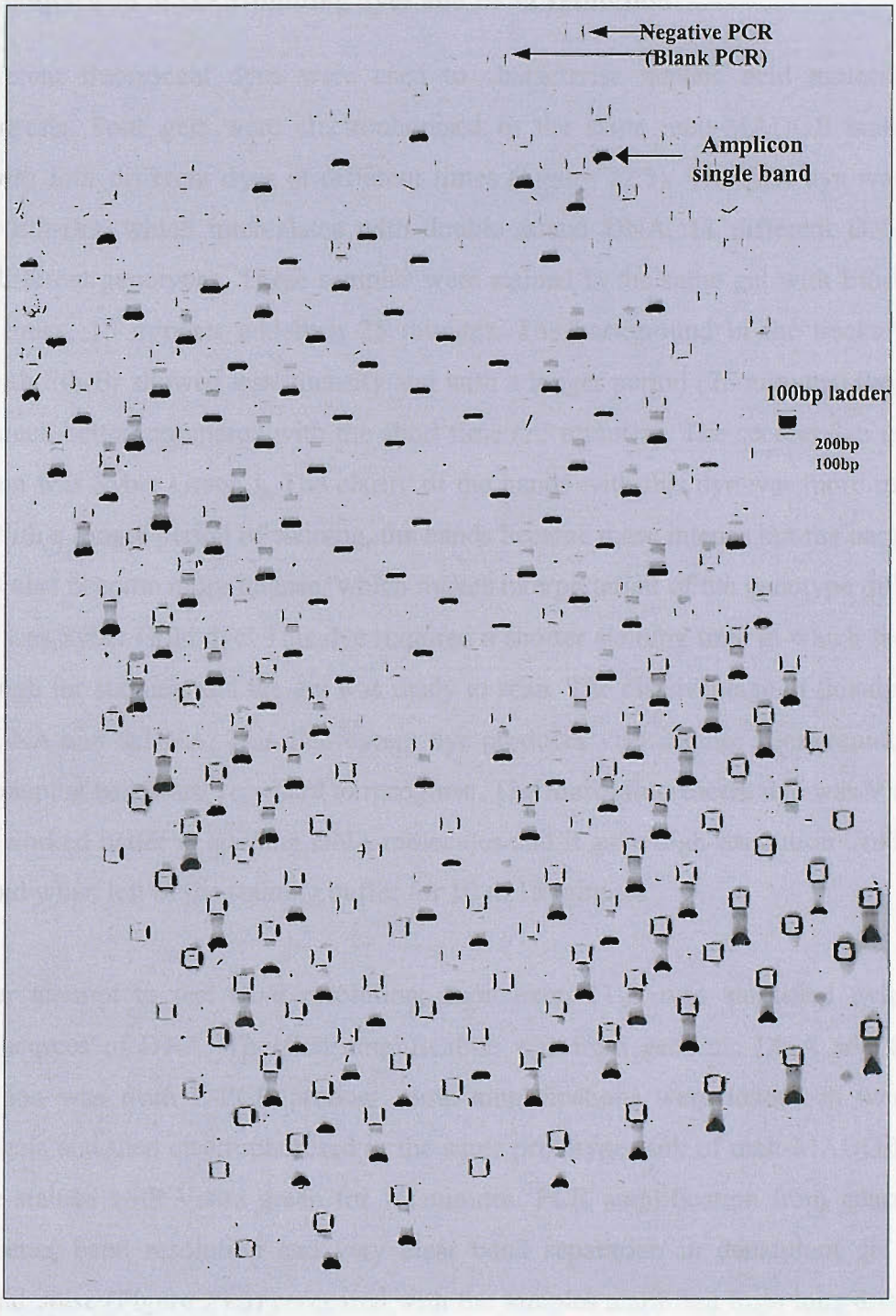
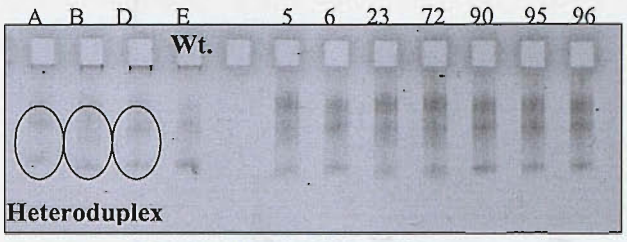


Figure 21.3. Checking PCR gels, represent some of the amplicons that will be run by Melt-MADGE. The amplicon in this experiment showed a single band, which reduces the background in melt-MADGE track. The blank wells are negative PCR.

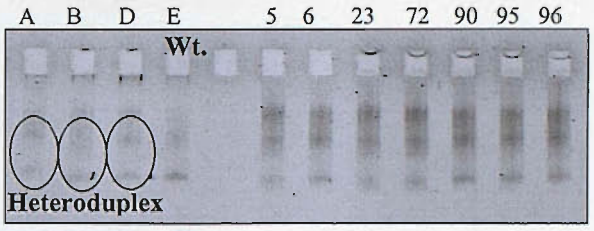
3.3 The comparison of DNA-binding dyes and band resolution

Four different fluorescent dyes were used to characterise nucleic acid molecules in gel electrophoresis. Four gels were electrophoresed in the same melt-MADGE tank and then stained with four different dyes at different times (Figure 22.3). The first dye was ethidium bromide (Eth-Br), which intercalates with double strand DNA. 11 different DNA samples showed different genotypes. These samples were stained in the same gel with Eth-Br for two different times, 15 minutes and then 25 minutes. The background in the tracks of the gel stained with Eth-Br showed less intensity and with a longer period (25 minutes) the resolution became much better, compared with the short time (15 minutes). The second dye used in this comparison was Syber Green I. The clarity of the bands with this dye was more intense than Eth-Br. With a longer period of staining, the bands became more intense but the background in the tracks also become more intense, which makes interpretation of the genotype difficult. The third dye was Syber Gold dye. This dye required a shorter staining time in which five minutes were enough for staining and the gel was ready to scan. The disadvantage of this dye is that it stains ssDNA and dsDNA. This fluorescent dye produces very intense background, in which the heteroduplex bands are very hard to recognise. The fourth fluorescent dye was Vistra green. This dye worked better in staining DNA molecules and it gave high-resolution bands and less background when left in the staining buffer for 10 to 15 minutes

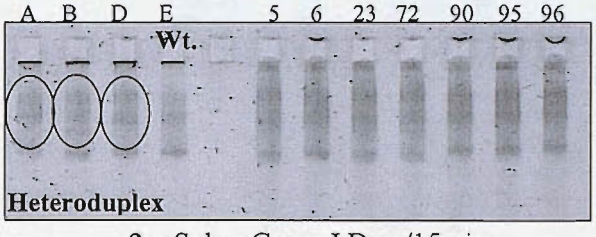
In another attempt to test band resolution, exon exon 11-B was amplified twice by two different sources of DNA. The first amplification was from genomic DNA and the second amplification was from L-PCR product. Both amplifications were loaded in two identical prepared gels and then electrophoresed in the same prototype tank of melt-MADGE. The two gels were stained with Vistra green for 15 minutes. PCR amplification from genomic DNA showed better band resolution and very clear band separation in denaturant gel with less background noise (Figure 23.3) compared with the samples amplified from long distance-PCR (L-PCR) (Figure 24.3).



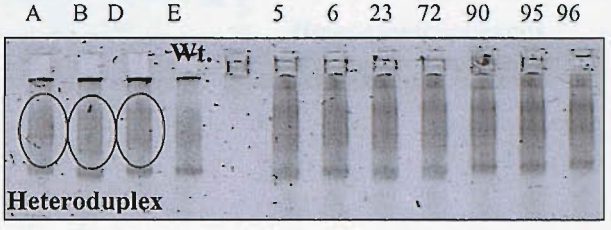
1a- Eth.Br. Dye /15mins



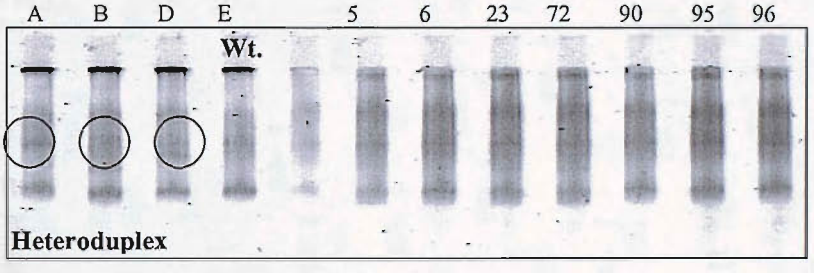
1b- Eth.Br. Dye / 25mins



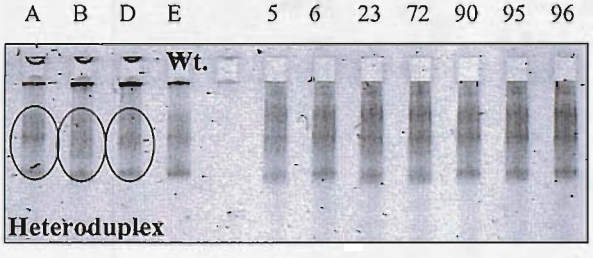
2a- Syber Green I Dye /15mins



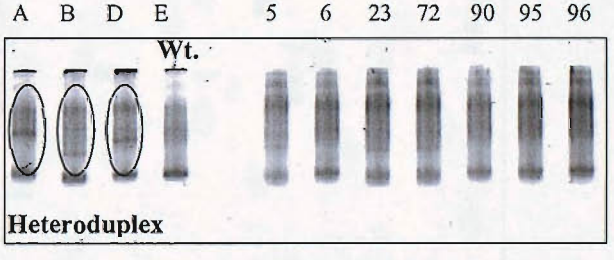
2b-SyberGreen I Dye /25mins



3- Syber GoldDye/5mins



4a- Vistra Green Dye / 20mins



4b- Vistra Green Dye / 25mins

Figure 22.3. Represents the comparison of different fluorescent dyes. The dyes used to stain DNA molecules melted in denaturant gel by Melt-MADGE method. The samples used in this experiment are four positive controls with previously known genotype, A, B, D and E these samples were amplified from genomic DNA. The circled bands in samples A, B and C showed the heteroduplex bands in those samples, and sample E is wild-type control DNA. The samples 5, 6, 23, 72, 90, 95 and 96 are random DNA samples from patients previously diagnosed with breast cancer with unknown genotype. These samples were amplified from L-PCR product. Panel 1a and 1b show the samples stained with Eth-Br for 15 and 20 minutes respectively. Panel 2a and 2b stained with Syber Green I for 15 and 25 minutes respectively. Panel 3, Syber Gold stained the DNA molecules in short time (5 minutes), which gives strong signals with intense background. Panel 4a and 4b, the gel stained with Vistra Green for 15 and 25 minutes respectively.
Wt; wild-type

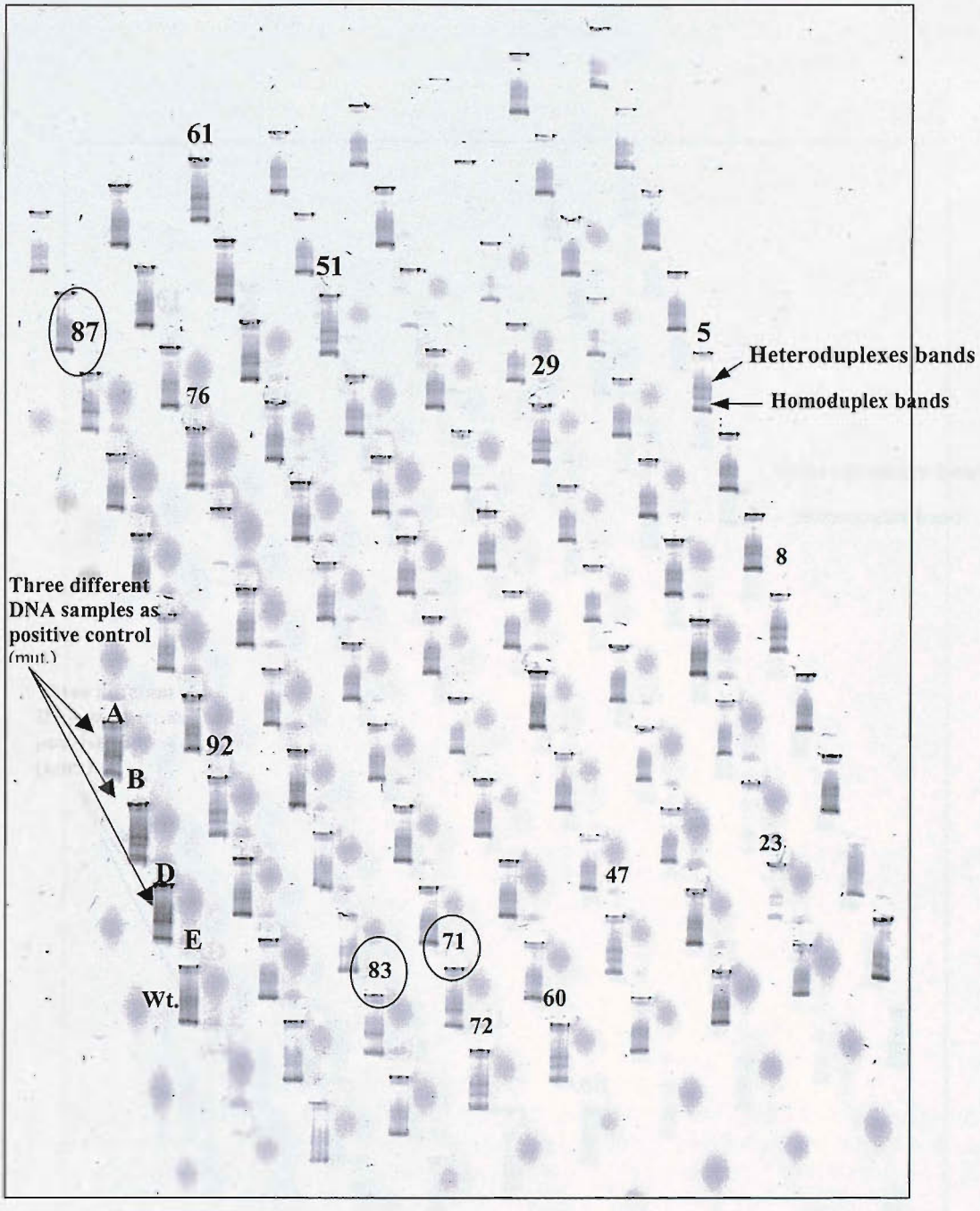


Figure 23.3. Represents exon 11-B amplicon amplified from direct genomic DNA. The gel stained with Vistra green and it revealed three different mutations, circled numbers 71, 83, and 87 (1138delG, 1218insA, 1131A>T, respectively), and 10 samples revealed the polymorphism 1187A>G, 5, 8, 23, 29, 47, 51, 60, 61, 72, 76 and 92. The A, B, D (1138delG, 1218insA, 1131A>T, respectively) and E is wild-type DNA. The bands are more resolved in this gel compared with the gel in Figure 24.3.

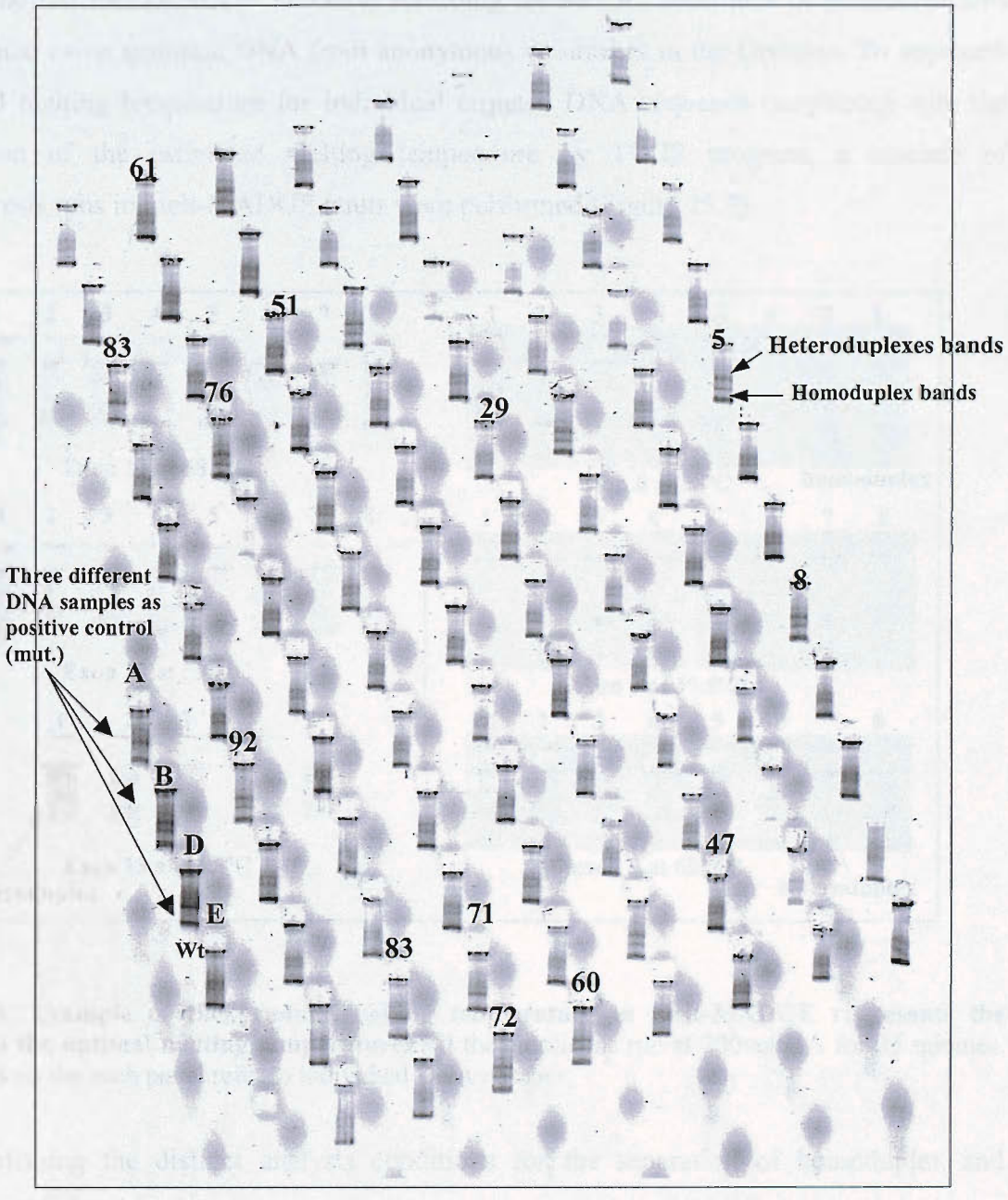


Figure 24.3. Represents exon 11-B amplicon amplified L-PCR. The gel stained with Vistra green and it revealed three different mutations, 71, 83, and 87 (1138delG, 1218insA, 1131A>T, respectively). 10 samples revealed the polymorphism, 1187A>G, 5, 8, 23, 29, 47, 51, 60, 61, 72 and 92. The A, B, D are 1138delG, 1218insA, 1131A>T, respectively, and E is a wild-type DNA. The bands in this gel are less well resolved compared with the gel in Figure 23.3.

3.4 Optimisation of the Melt-MADGE technique for *BRCA1* gene by employing constant temperature

The efficiency of melt-MADGE mutation screening for *BRCA1* mutations or polymorphisms was examined using genomic DNA from anonymous volunteers in the Division. To approach the optimal melting temperature for individual targeted DNA sequence (amplicon) with the consideration of the estimated melting temperature by TIXIS program, a cascade of electrophoresis runs in melt-MADGE tanks were performed (Figure 25.3).

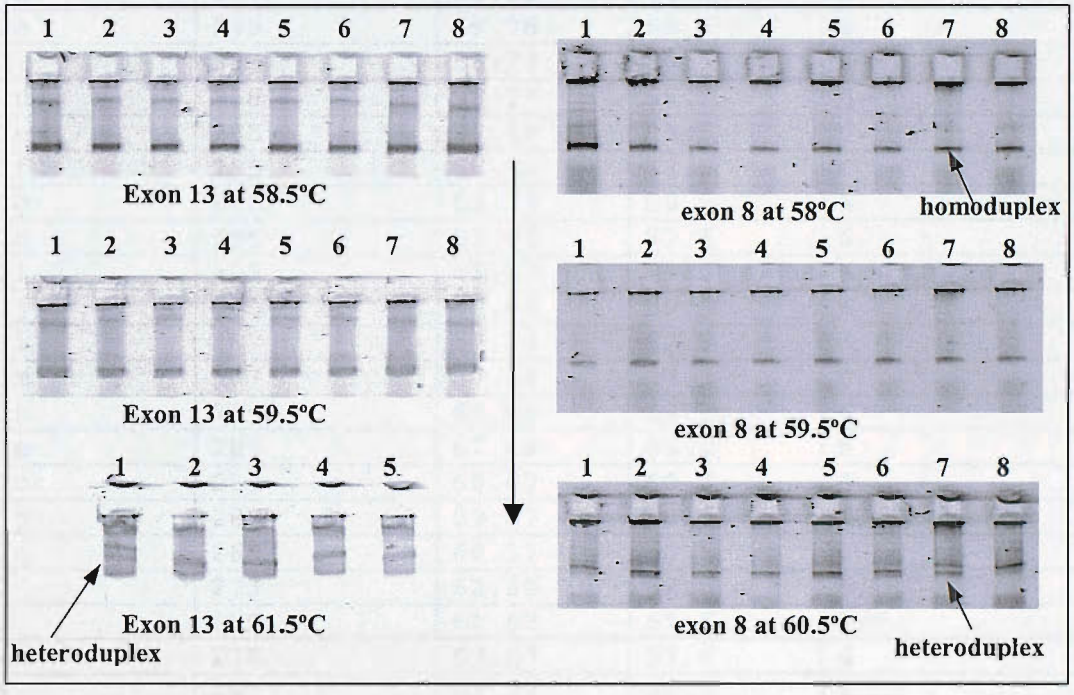


Figure 25.3. Example of the gradual melting temperature in melt-MADGE represents the approach to the optimal melting temperature. All the amplicons run at 200volt/1A for 35 minutes. The numbers on the each panel refer to individual DNA number.

After establishing the distinct analysis conditions for the separation of homoduplex and heteroduplex molecules, the optimum melting temperatures of all the amplicons at constant temperature with 4M urea is achieved. The identification of samples with polymorphisms or mutations is a direct procedure. Table 6.3 lists the 54 fragment analyses performed using Melt-MADGE at a constant temperature. I was able to detect all the polymorphisms and different mutations in patients with hereditary breast cancer.

Exon /fragment ID	Exon/fragment size (bp)	*Predicted Tm (°C)	‡Observed *Tm* (°C)	Acrylamide Gel concentration (%)
2	303	57.20	57.5	6
3	305	57.94	57.5	6
5	290	55.33	55.6	6
6	270	58.80	58.0	6
7	290	57.17	58.5	6
7	310	57.24	58.5	6
8	338	60.27	60.5	5/6
9	227	55.86	55.6	6
10	321	59.11	58.5	6
11a	277	60.52	60.5	6
11b	345	59.78	59.5	6
11c	285	58.73	58.5	6
11d	258	58.18	58.5	6
11e	285	57.56	57.5	6
11f	285	58.33	58.5	6
11g	285	59.79	59.5	6
11h	285	58.06	57.5	6
11i	305	59.25	59.5	6
11j	285	60.04	59.5	6
11l	287	57.34	57.5	6
11m	307	57.64	57.5	6
11k	330	59.60	59.80	6
11n	285	57.46	56.0	6
11o	289	60.67	60.7	6
11p	285	59.07	58.5	6
11q	265	60.37	60.5	6
12	273	62.30	62.5	6
13	325	61.02	61.5	5/6
14	278	57.67	57.6	6
15	281	61.27	61.5	6
16/16a	445/ 337	62.30	63.3	5
17	255	57.44	57.6	6
18	353	58.44	58.5	6
19	200	56.29	56.4	6
20	285	61.86	62.5	6
21	254	62.79	62.5	6
22	288	63.36	63.1	6
23	191	63.90	62.8	6
24a	300	66.32	66.5	6
24g	234	60.97	59.5	6

Table 6.3 predicted and observed *Tm* for *BRCA1* gene amplimers. The 54 amplimers were designed by TIXIS program, and their predicted Tm was relatively compatible with the Tm, which was observed by melt-MADGE. Appendix D shows examples of melt-MADGE results that represent the comparative Tm values in the table.

**Predicted by TIXIS program and corrected by (-10°C) for an effect of 4M urea in gel.* Judged as approximately 50% retardation of homoduplex.

‡; The temperature was recorded in the tank based on the amplicon melting during the electrophoresis run

3.4.1 GC-content and melting temperature

The melting temperature of the amplicon is increased by a higher percentage of G-C base pair content in the DNA sequence. Figure 26.3 represents the linearity between the theoretical Tm estimated by TIXIS program and empirical Tm observed during the analysis, which is monitored by Melt-MADGE method. The use of a GC-Clamp (Figure 3.1) at one end of the amplicon is not included in the estimated melting temperature for the amplicon.

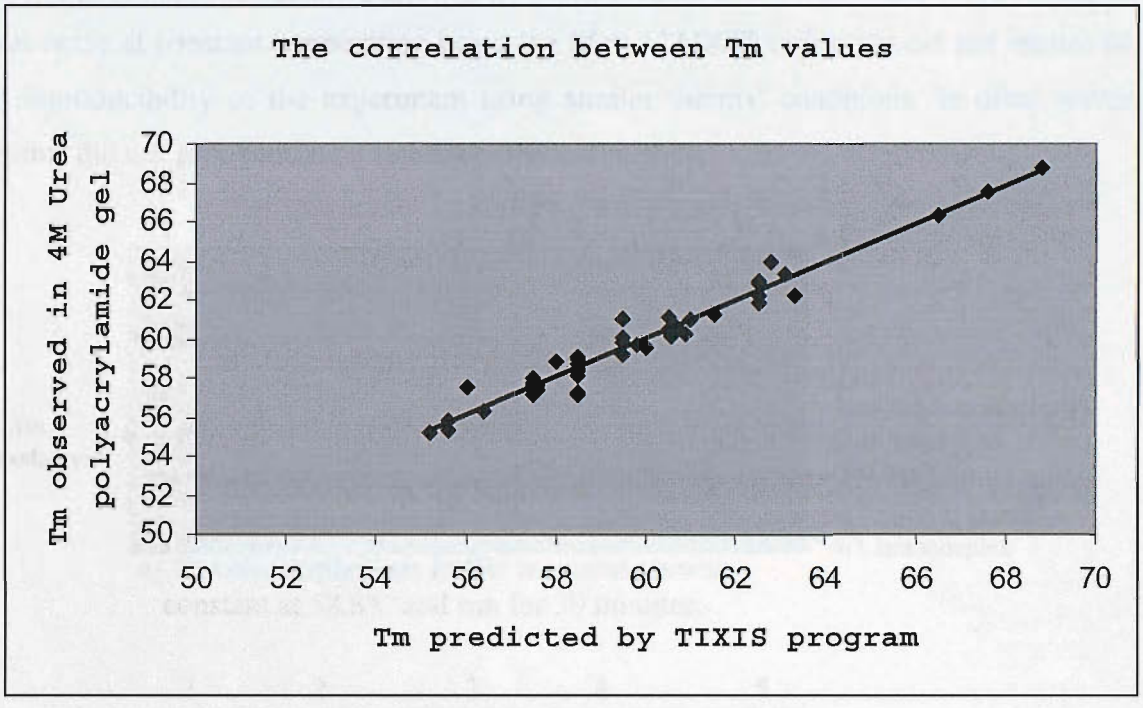
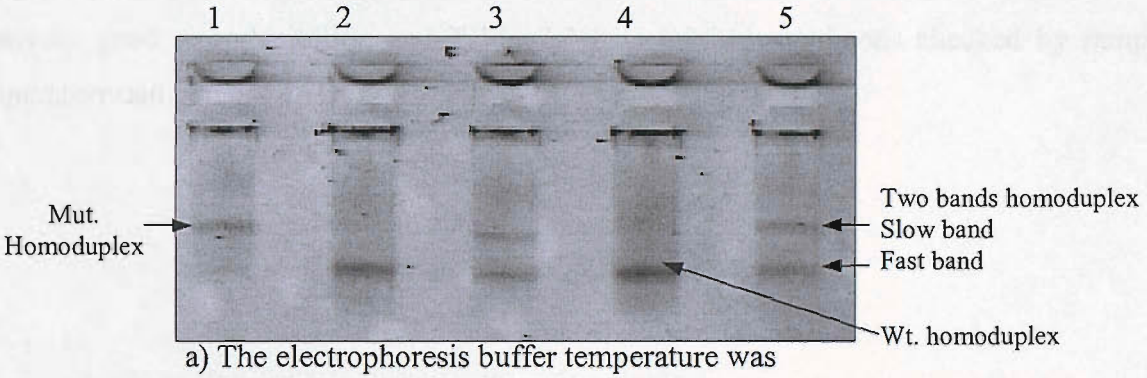


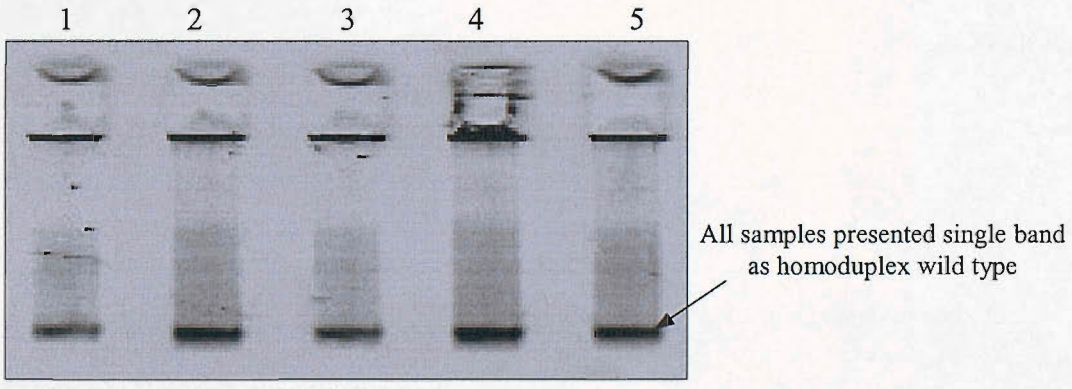
Figure 26.3. The relation between Tm predicted by TIXIS program and Tm observed in 4M Urea polyacrylamide gel during *melt*-MADGE analysis produced from table 3.1. The difference in these two values might result from statistical error in the program or result from false estimated Tm from *melt*-MADGE run.

3.4.2 Constancy of amplicon`s melting temperature

The problems encountered in our testing of melt-MADGE are associated with constancy of the melting resolution of targeted DNA sequence. It was noted that slightly different running times and temperatures occurred during electrophoresis when repeating the identical experiment environments for the amplicon, the band could either over melt and appear as condensed background near to the original well or the run could end with no melting for the amplicon. The latter required re-running at a higher temperature (Figure 27.3). Unfortunately, running the amplicon twice at constant temperature using the Melt-MADGE technique did not ensure the optimal reproducibility of the experiment using similar thermal conditions. In other words, experiments did not give consistent reproducible results.



a) The electrophoresis buffer temperature was constant at 58.8°C and run for 30 minutes.



b) The electrophoresis buffer temperature was constant at 58.8°C and run for 30 minutes.

Figure 27.3. Represents fragment 11-G with five different DNA control samples as an example of the constent melting temperature. The variation in the same fragment at the similar conditions was not identical. The variation in panel (a) represents the polymorphism at 2201C>T, while in panel (b) was not resolved

3.5 Examining *BRCA1* by employing thermal ramp electrophoresis

Since the melt-MADGE assay at constant temperature for the *BRCA1* amplicons did not resolve mutation detection, the ramping temperature was employed to replace the constant temperature assay. All fifty-four *BRCA1* amplicons were re-optimised individually at ramping temperature (Table 7.3). This appeared to be the optimal mutation detection assay using the melt-MADGE method. The ramping temperature constructed at 5°C window (e.g. 55°C to 60°C) for each amplicon between starting temperature and final one. Many mutations and polymorphisms in different regions of *BRCA1* gene demonstrated the constancy of repeating the procedure under the same conditions. Figure 28.3 shows some of the amplicons tested using ramping temperature. I was able to detect variant alterations in *BRCA1* fragments with relatively good reproducibility, and Table 7.3 lists *BRCA1* amplicons checked by ramping temperature using melt-MADGE.

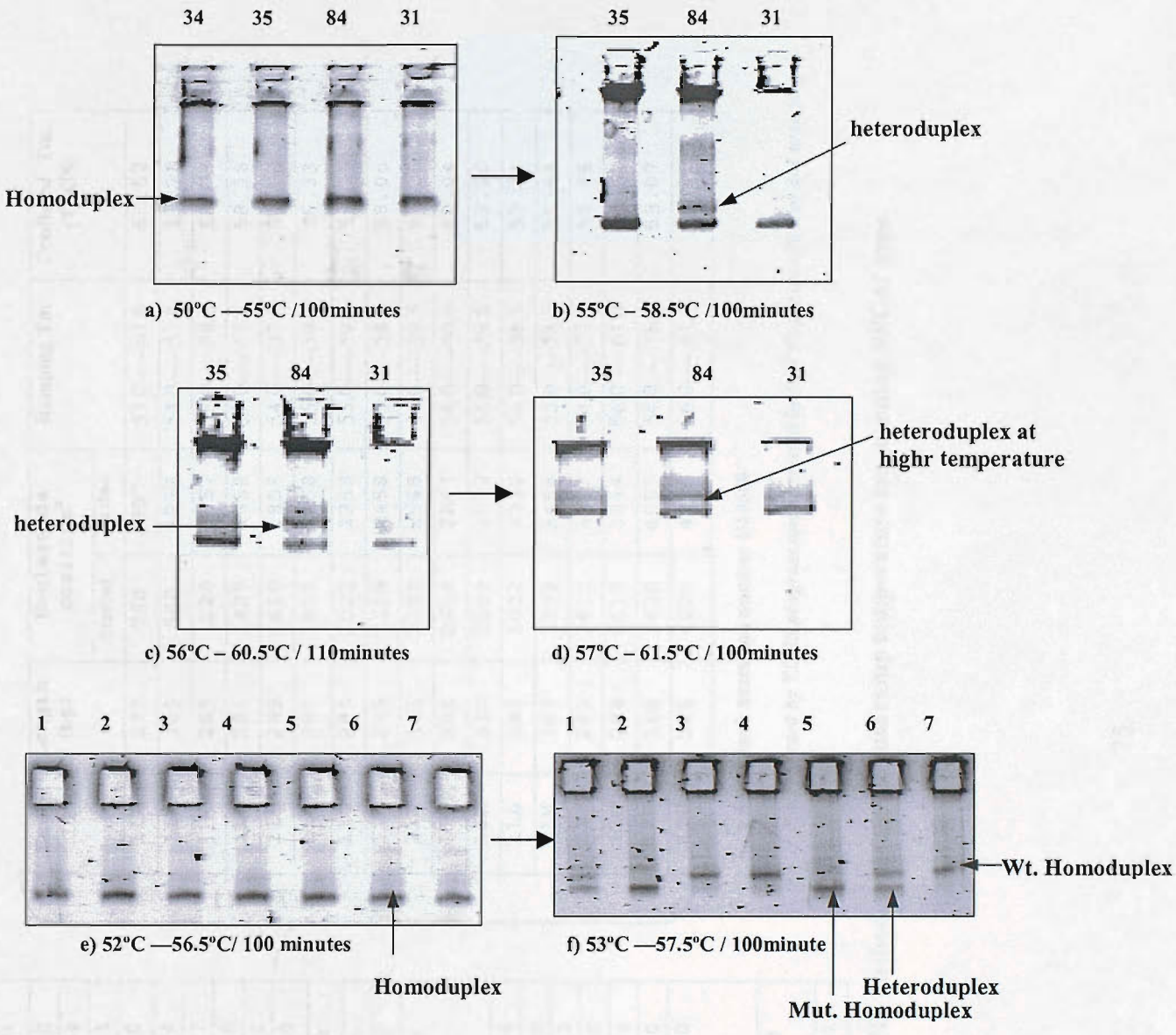


Figure 28.3. Examples of the *BRCA1* amplicons optimisation using thermal ramp electrophoresis. Panels a-d show the cascade of thermal ramp for amplicon 11-C, and it reveals the mutation in sample 84 (1447insCC), sample 35 has 1445T>A, which is a mutation sited near the primer location, and samples 34 and 31 are wild-type control samples. Whereas panels e and f of exon 11-L fragment show two step approach to the optimal thermal ramp electrophoresis. The variant in amplicon exon11-L shows the polymorphism 3232A>G. The samples employed in this fragment are control samples.

Exon #	Length (bp)	Ramping temp. (°C) Start-----Final	Predicted* Tm. (TIXIS)
2	180	54.0-----57.5	58.10
3	305	54.0-----58.3	57.94
5	290	51.0-----55.5	55.33
6	270	56.0-----60.5	58.80
7	310	54.0-----57.3	57.24
8	207	58.0-----61.5	60.27
9	227	53.0-----56.5	55.86
10	193	58.0-----63.0	59.11
12	295	57.0-----62.0	62.30
13	325	57.0-----61.5	61.9
14	300	55.0-----58.5	57.6
15	355	57.0-----61.5	61.0
16	445	61.0-----63.5 +3.75Urea / 5% PAG gel	62.3
17	255	55.0-----58.5	57.44
18	353	55.0-----58.5	58.44
19	200	55.0-----58.5	56.29
20	285	58.0-----63.5	61.86
21	254	58.0-----63.5	62.79
22	244	57.0-----61.0	61.70
23	191	57.0--60.0 + 4.5M Urea/ 6% PAG gel	62.00
24A	318	56.0-----59.8 +7M Urea	65.3
24G	234	57.0-----60.0	60.97

Exon 11

Exon 11	Length (bp)	Nucleotide position[a]		Ramping Tm	Predicted* Tm. (TIXIS)
		Started	Ended		
11A	277	768	997	57.0-----61.0	60.52
11B	345	960	1258	54.0-----59.5	59.78
11C	285	1220	1457	53.0-----59.5	58.73
11D	285	1429	1658	53.0-----57.5	58.18
11E	285	1619	1858	54.0-----57.5	57.56
11F	285	1819	2058	55.0-----59.5	58.33
11G	285	2021	2258	55.0-----59.5	59.79
11H	285	2219	2458	55.0-----58.5	58.06
11I	305	2399	2648	54.0-----59.5	59.25
11J	285	2634	2867	56.0-----60.0	60.04
11K	330	2802	3079	54.0-----59.5	59.20
11L	287	3022	3299	54.0-----58.5	57.34
11M	307	3203	3458	55.0-----58.5	57.64
11N	285	3419	3657	54.0-----57.0	57.46
11O	289	3619	3844	56.0-----61.0	60.67
11P	285	3820	4057	54.0-----58.0	59.07
11Q	265	4020	4239	56.0-----61.5	60.37

a = GenBank accession number U14680

**** = Predicted by TIXIS program and corrected by (-10°C) for an effect of 4M urea in gel***

Table 7.3 The empirical condition of melt-MADGE method applying thermal ramp temperature for scanning *BRCA1* gene.

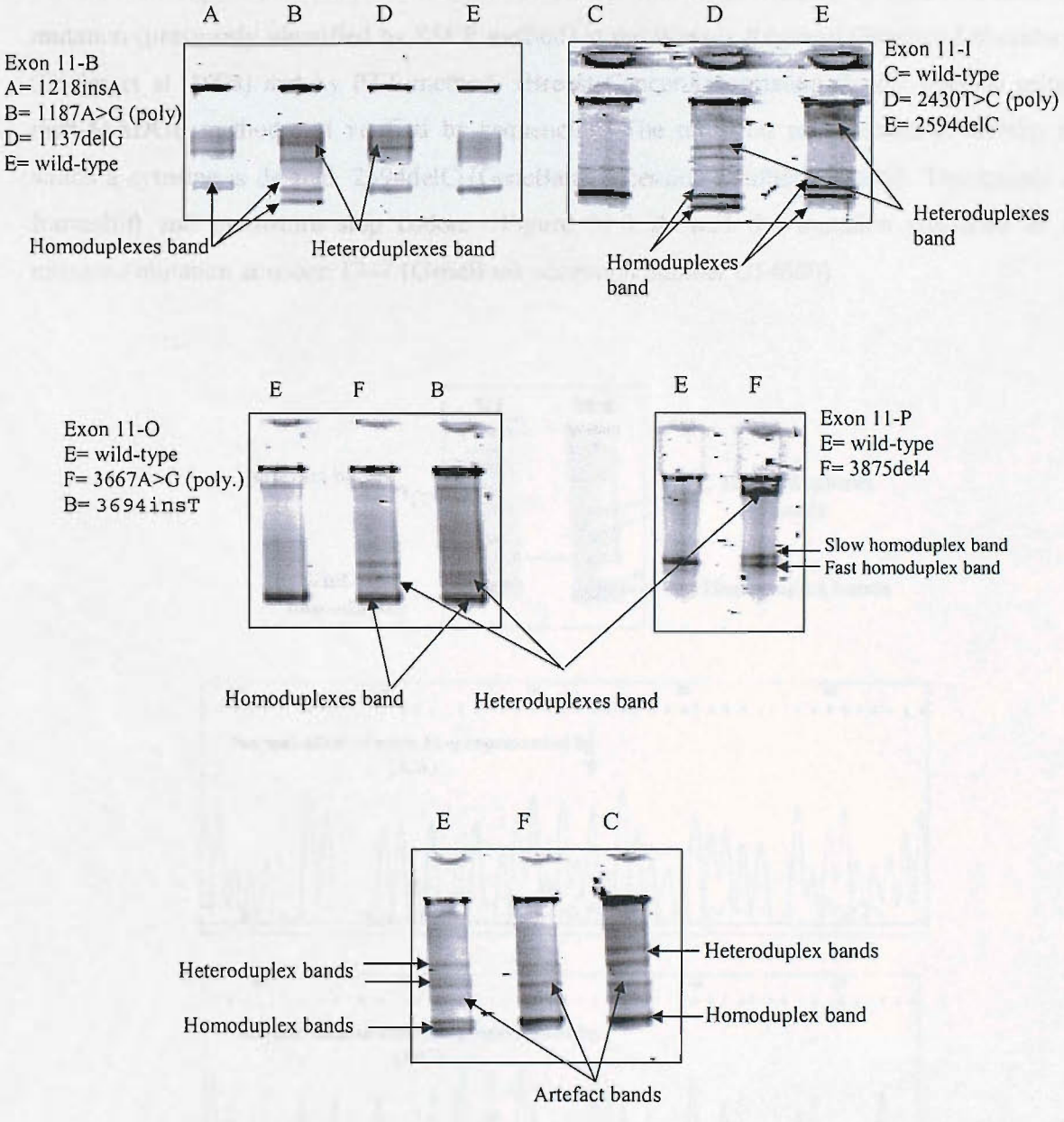
3.6 Melt-MADGE evaluations

3.6.1 Test of positive controls

Six positive DNA samples (sample A-F) obtained from different patients, in the previously identified mutations in the *BRCA1* gene, were tested for melting assay using melt-MADGE method (Table 8.3). All the six samples had shown different genotypes in the melt-MADGE assay experiment. Nucleotide sequencing for the samples showing variants revealed the same mutation previously identified at the Wessex Regional Genetics laboratory, Salisbury District Hospital (Figure 29.3).

Sample ID	Mutation type	Exon 11 fragment	Melt-MADGE result
A	1218insA	11-B	+
B	3694insT	11-O	+
C	4158delAG	11-Q	+
D	1137delG	11-B	+
E	2594delC	11-I	+
F	3875del4	11-P	+

Table 8.3. Represents some of the *BRCA1* mutations detected in positive sample of patients with breast cancer.



Exon 11-Q, E = 4158A>G, F= wild-type, C = 4158delAG

Figure 29.3. Shows different mutations identified in the positive control samples in exon 11 of *BRCA1* gene as detected by Melt-MADGE at constent melting temperature, 59°C.
(Poly = Polymorphism)

One example of DNA sequencing, which is illustrated in Figure 30.3, presented the 4158A>G mutation identified at the 3` end of exon 11 (4158A>G, Arg to Gly), which was previously detected and reported in the literature (Breast Cancer Information Core). In addition, another mutation (previously identified by SSCP method) at the Wessex Regional Genetics Laboratory (Eccles et al. 1998) and by PTT method (Breast Cancer Information Core), detected using melt-MADGE method and verified by sequencing. The mutation was located at 2594bp at which a cytosine is deleted, 2594delC (GeneBank accession number U14680). This causes a frameshift and premature stop codon. Figure 31.3 showed the mutation classified as a missense mutation at codon 1347 (GeneBank accession number U14680).

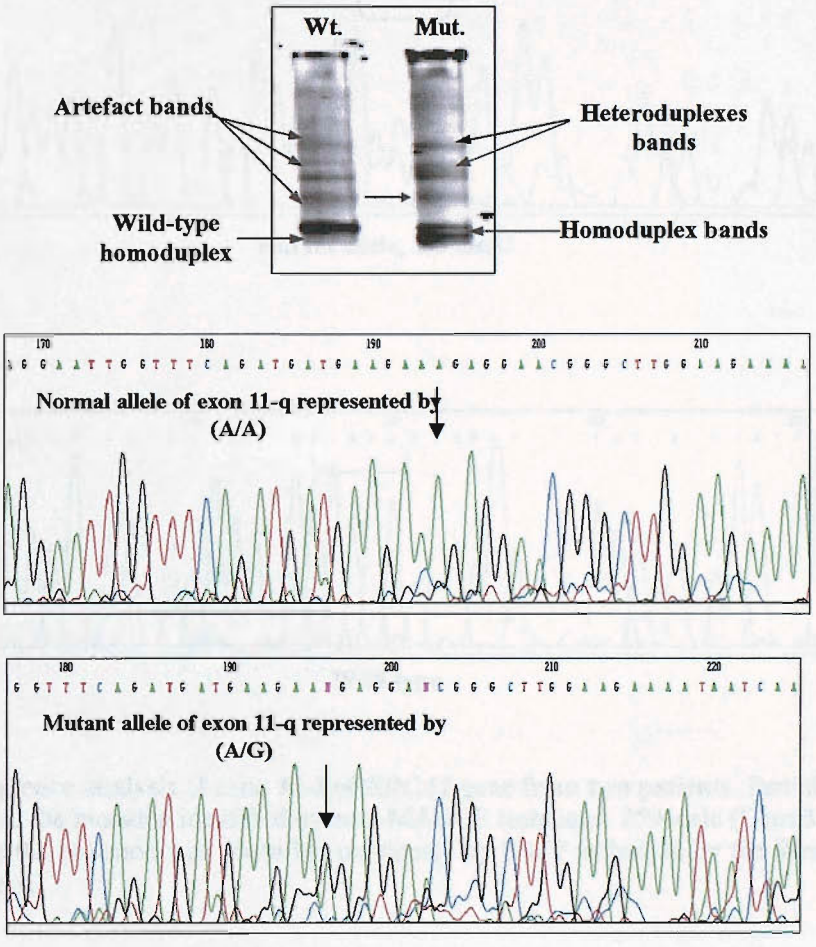
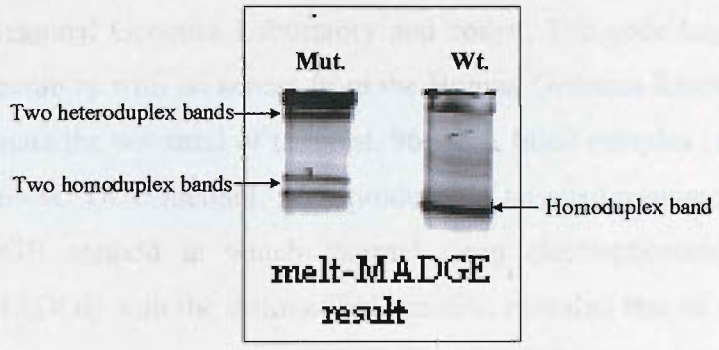
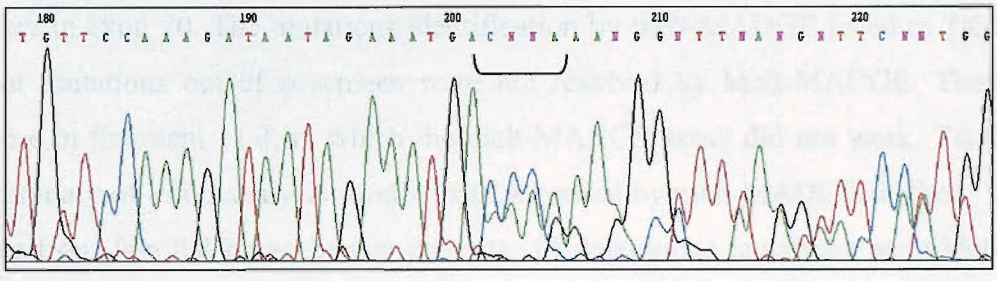
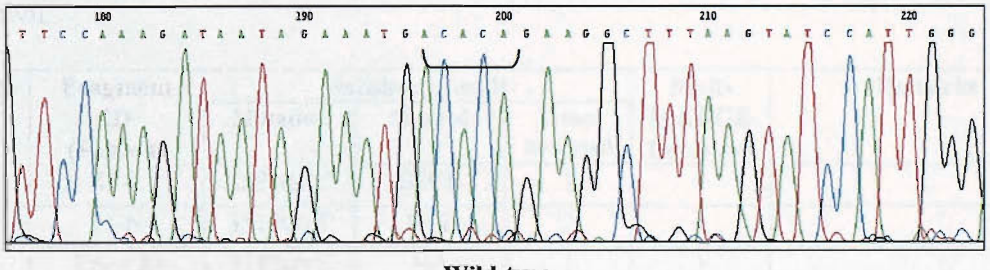


Figure 30.3. Sequence and melt-MADGE result of the exon 11-q mutation in *BRCA1* gene. The variant 4158A>G (GeneBank accession number U14680) detected by melt-MADGE technique and confirmed by ABI 377 DNA sequencer.





Mutant allele, 2594delC



Wild-type

Figure 31.3. Sequence analysis of exon 11-I of *BRCA1* gene from two patients. Partial sequences of exon 11 confirmed the mutation identified by melt-MADGE technique. 2594delc (GeneBank accession number U14680) the mutation was identified previously by SSCP technique by the Wessex Regional Genetics laboratory.

3.6.2 Blind test

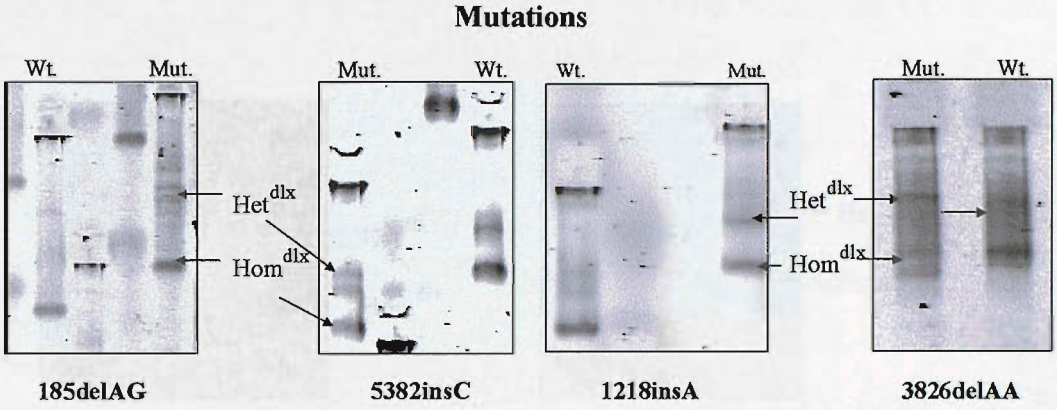
Samples were previously screened for mutations in the *BRCA1* gene using SSCP / HA and PTT at the Wessex Regional Genetics Laboratory and coded. The code kept by the Wessex Regional Genetics Laboratory with no access from the Human Genetics Research laboratory in Southampton. To evaluate the potential of this test, 96 DNA blind samples (anonymous) were re-screened by the Melt-MADGE method. PCR products of targeted sequences were analysed using the melt-MADGE method in which thermal ramp electrophoresis was employed. Comparison of melt-MADGE with the decoded information revealed that of the 17 previously identified mutations, two mutations were in exon 2, 13 mutations were in exon 11 and 2 mutations were in exon 20. The mutations identification by melt-MADGE listed in Table 9.3. Two different mutations out of seventeen were not resolved by Melt-MADGE. These two mutations were in fragment 11-J, in which the melt-MADGE assay did not work. Table 10.3 shows the summary of blind analysis results that I screened by melt-MADGE method. Within these samples from familial breast cancer patients, 13 deleterious mutations were identified. Eight patients had five frameshift mutations, and five missense mutations. In addition to the mutations and variants described above, a number of polymorphic variants were detected in the 96 samples. Examples of heterozygous mutations (Figure 32.3) and polymorphisms (Figure 33.3) are shown.

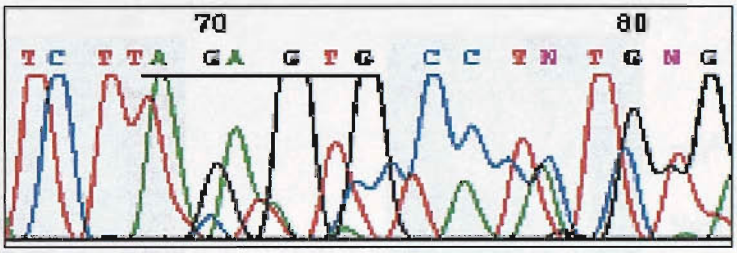
Sample #	Fragment ID (Region)	Salisbury result			Melt-MADGE Detection	Remarks
		Mutation	Method	Times analysed		
7	11-I	2594delC	SSCP	1	+	
10	11-N	3519G>T	PTT	2	+	
30	Exon 20	5382insC	HA	1	+	
34	Exon 20	5382insC	HA	1	+	
35	Exon 2	185delAG	HA	1	+	
52	11-D	1445T>A	Seq.	1	+	
60	11-O	3695insT	HA	1	+	
70	11-G	2187delA	PTT	1	+	
71	11-B	1138delG	HA	1	+	
72	11-P	3882delAA	HA	1	+	
79	Exon 2	185delAG	HA	1	+	
81	11-Q	4184del4	SSCP/HA	1	+	
83	11-B	1218insA	Seq.	1	+	
87	11-B	1131A>T	PTT	1	+	
88	11-J	2773delTC	PTT	1	-	Melting assay did not work
90	11-J	2804delAA	HA	1	-	Melting assay did not work
95	11-O	3826delAA	PTT	1	+	

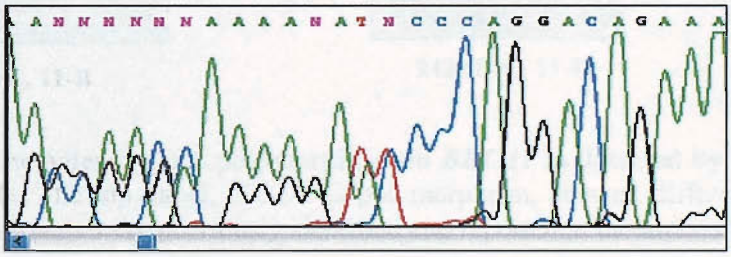
Table 9.3. Comparison of melt-MADGE results for mutations previously identified by other techniques.

Sample No.	Fragment ID	Blind test		Mutation position	Mutation effect	Re-checking the array		Remarks
		+Ve	-Ve			+Ve	-Ve	
35, 79	Exon 2	+		185delAG	FS			
71	11-B		-	1137delG	FS	+		Genomic DNA, which gives less background in the track
83	11-B	+		1218insA	T367N			
87	11-B		-			+		Genomic DNA, which gives less background in the track
52	11-D	+		1445T>A	FS			
70	11-G	+		2190delG	FS			
57, 73	11-I		-	2441T>A	G774G	+		Error created during transferring the ARMS data. Whole array were re-checked individually.
	11-J		-					The melting assay did not work
10	11-N	+		3519G>T	nonsense			
95	11-O	+		3825delAA	FS			
72	11-P	+		3881delGA	FS			
81	11-Q	+		4176G>T	nonsense			
30, 34	Exon 20	+		5385insC	FS			
		Other findings from blind test not reported by Salisbury lab. or new finding						
		Mutation	Polymorphisms					
Many	11-B		+	1187A>G	Q358R			
86	11-F	+		1967T>G	S616C			Not in literatures
11	11-H	+		2413A>G	E765A			Not in literatures
57, 73	11-L, M		+	3238G>A	S1040N			
16, 49	11-O		+	3667A>G	K1183R			
20	11-Q	+		4158A>G	N1354S			Known
11, 12, 67	Exon 17	+		5143C>T	T175I			Not in literatures

The 10.3 Melt-MADGE results of blind tests, 96 samples analysed for mutations in three exons, 2, 11, and exon 20 of *BRCA1* gene. FS; Frameshift.







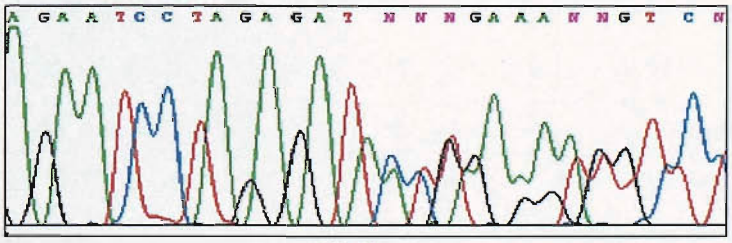


Figure 32.3. Examples of different mutations identified in *BRCA1* by Melt-MADGE. The top panels represent mutations observed in exon 2, 11-B, 11-O and exon 20, previously identified and reported in the literatures. In the second panel, some of the sequencing results confirm the nature of the genotype identified by using DNA sequencing.

Het^{dlx} = heteroduplex
Hom^{dlx} = homoduplex

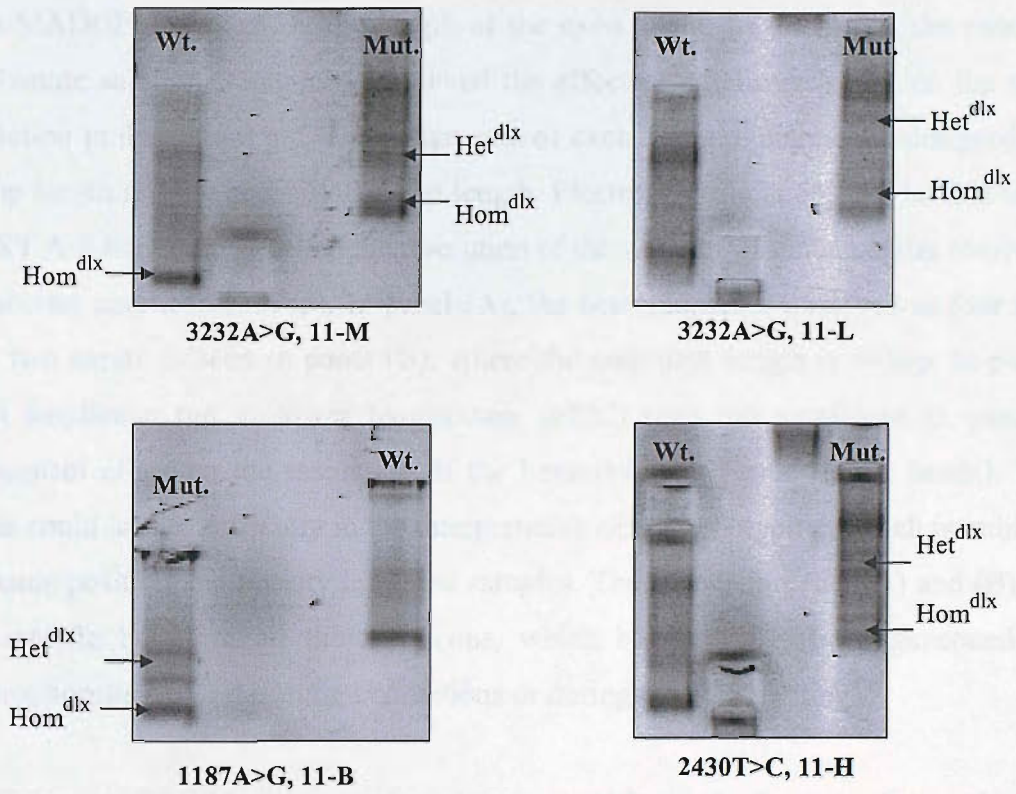


Figure 33.3. Example of the heterozygous polymorphism in *BRCA1* as detected by Melt-MADGE in blind test of 96-sample. The top panel, 3232A>G polymorphism, showed different pattern with different amplicon. These variants were previously identified and reported in the literature.

Het^{dlx} = heteroduplex
Hom^{dlx} = homoduplex

3.7 Bands patterns and high resolution of the fragments

The amplification with high frequencies of common polymorphisms in the BRCA1 gene (exon 11, 13, and 16) was chosen for the initial analysis. However, the band patterns displayed by Melt-MADGE depended on the length of the exon or the fragment and the concentration of acrylamide solution. Figure 34.3 showed the effect of amplimer length on the mobility and resolution in denaturant gel. Two fragments of exon 16 amplimers were designed, the first of 445bp length and the second of 337bp length. Electrophoresis in 5% acrylamide and 4-M urea in 1XT.A.E buffer showed that the resolution of the variants was much better resolved by using the shorter amplicon (337bp). In panel (A), the heteroduplexes observed as four bands rather than two bands as seen in panel (B), where the amplicon length is 445bp. In panel (C), the short amplicons run at lower temperature (62°C) than the amplicons in panel (A) with consequent effect on the resolution of the homoduplexes bands (lower bands). The artefact bands could lead to difficulty in the interpretation of the genotyping, which is unlikely to be at the same position and density in all the samples. The arrows in panel (A) and (B) showed the non-specific bands in all the amplicons, which had probably been generated during the electrophoresis under denaturant conditions or during the PCR cycling.

Heterozygotes should generate four bands; two heteroduplexes and two homoduplexes, depending on whether the four bands are sufficiently separated during the electrophoresis. Because the heteroduplexes are less stable than homoduplexes and display slower mobility under denaturing conditions than homoduplexes, heteroduplexes tend to be retarded in mobility and probably will disappear during the electrophoresis as an over melting band. The DNA heteroduplexes have reduced mobility compared with DNA homoduplexes. Figure 35.3 illustrates the PCR-Melt-MADGE pattern of 12 control DNA samples, containing a polymorphism in the non-coding region of exon 24 that was used to determine the optimum band resolution. The upper two bands represent the heteroduplex bands, caused by alterations in duplex stability due to base mismatch, in this case T-C and G-A. The two lower bands (C-G/T-A) represent the homoduplex bands, one band represents the mutant allele and the other the normal allele compared with the bands in track 2 and 3. The fragment runs in denaturant gel containing 6% polyacrylamide/4M Urea. The fragment length is 234bp.

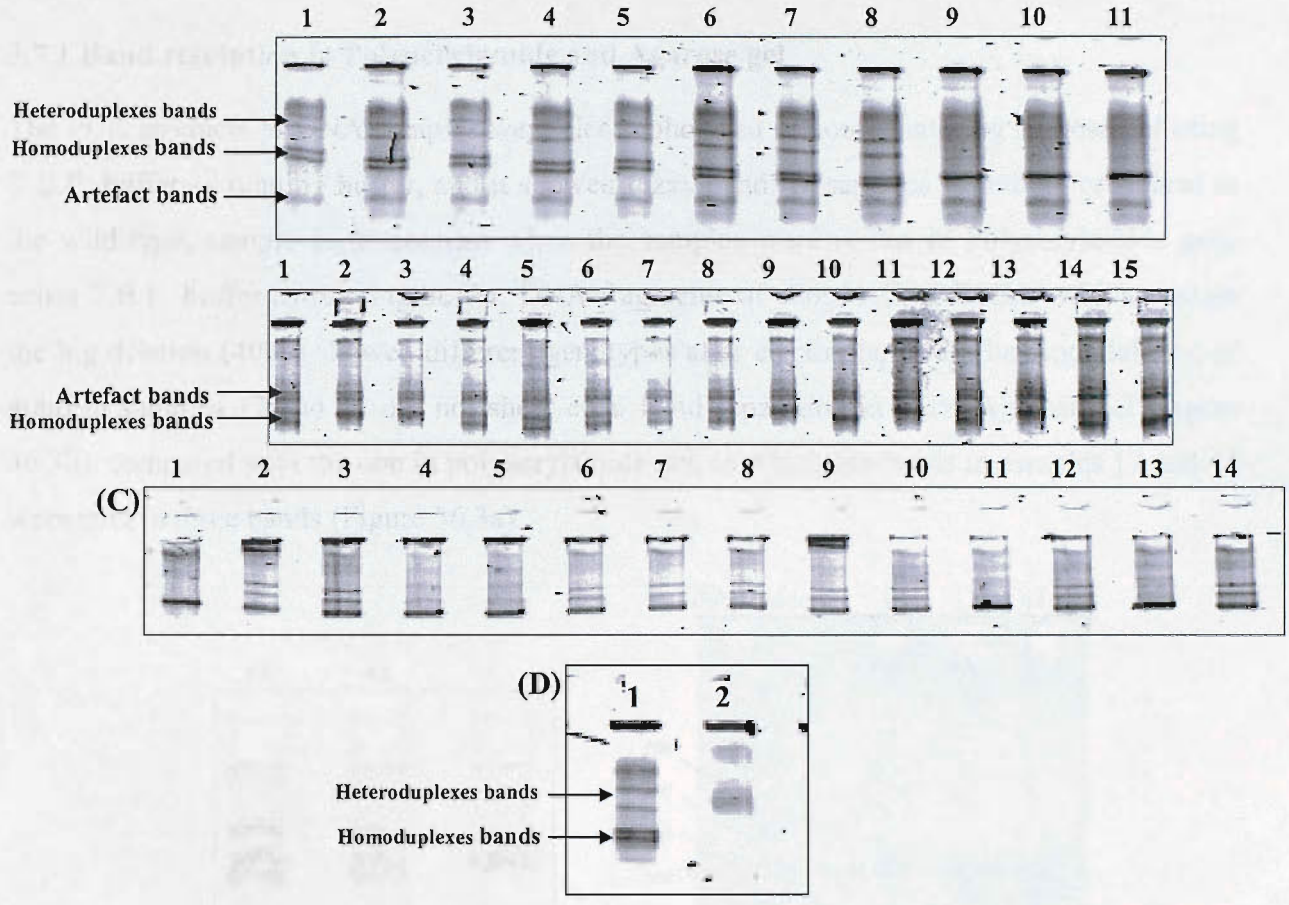


Figure 34.3. melt-MADGE analysis of exon 16. The fragments were amplified then electrophoresed in 5% acrylamide gel/ 4M Urea. Three conditions were performed for exon 16. In panel (A), the amplicons were run at constent melting temperature (63.3 °C for 35 minutes. The length of this fragment was 337bp, where the best resolution was seen. In panel (B), the fragments were run under the same conditions as the conditions in panel (A), the length of the amplicon is longer, 445bp. In panel (C), the same amplicons as in panel (A) were run in 6% acrylamide gel / 4M Urea at a lower temperature, 62.0°C, for 30 minutes. Panel (D) shows the comparison between two lengths for the same amplicon running in the same gel under the same conditions as it in panel (A). Lane (1) is the short amplicon (337bp) and lane (2) is the long amplicon (445bp).

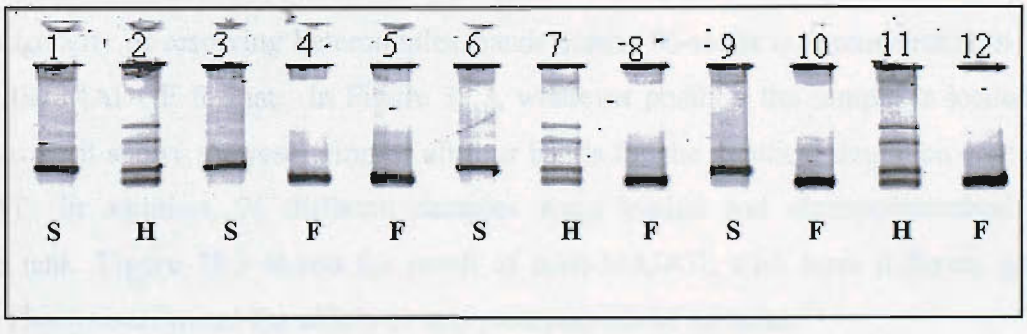


Figure 35.3. Shows the result of Melt-MADGE analysis of exon 24 (fragment g) and demonstrates the variant genotypes within 12 control DNA samples. (Amplicons were loaded in a single row to allow genotype comparison). Samples 4,5,8,10 and 12 showed the fast allele (C/C), 3,6, and 9 showed the slow allele (T/T) and 2,7 and 11 showed the heterozygote allele (C/T). S= slow, F= fast, H= heterozygote.

3.7.1 Band resolution in Polyacrylamide and Agarose gel

The PCR products of DNA samples were electrophoresed in non-denaturing agarose gel using T.B.E. buffer as running buffer, and it showed fuzzy bands in samples 17 and 44 compared to the wild-type, sample 1. In contrast when the samples were re-run in Polyacrylamide gels, using T.B.E. buffer as running buffer, DNA fragments of exon11-C of *BRCA1* which contain the big deletion (40bp) showed different genotypes after electrophoresis. The large deletion of 40bp in samples 17 and 44 did not show clear band separation in a 3% Agarose gel (Figure 36.3b), compared with the one in polyacrylamide gel, in which the bands in samples 17 and 44 were split in three bands (Figure 36.3a).

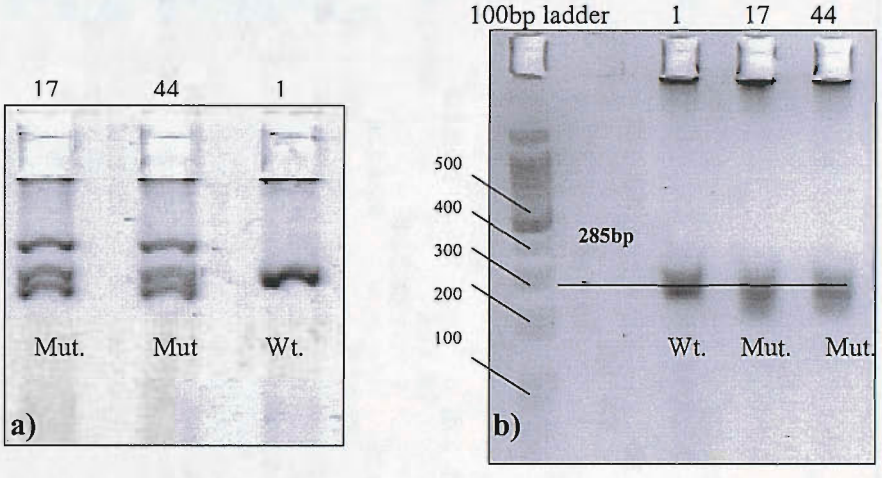


Figure 36.3. Shows the comparison between mobility in non-denaturant of Polyacrylamide and mobility in Agarose gels of DNA samples with big deletion (40bp). In (a) demonstrates the resolution of separating the bands in 6% polyacrylamide gel prepared in TBE buffer, while in (b) presenting the fuzzy band in 3% Agarose gel prepared in TBE buffer.

3.7.2 Homogeneity of Melt-MADGE method

The homogeneity of resolving heteroduplex bands across 96-wells is demonstrated in one gel by using the MADGE format. In Figure 37.3, whatever position the sample is loaded in the 96-well array, it shows the resolution of all four bands for the identical amplicon contains the exact SNP. In addition, 96 different samples were loaded and electrophoresised in our prototype tank. Figure 38.3 shows the result of melt-MADGE with three different genotype patterns. This demonstrated the ability to analyse hundreds of samples.

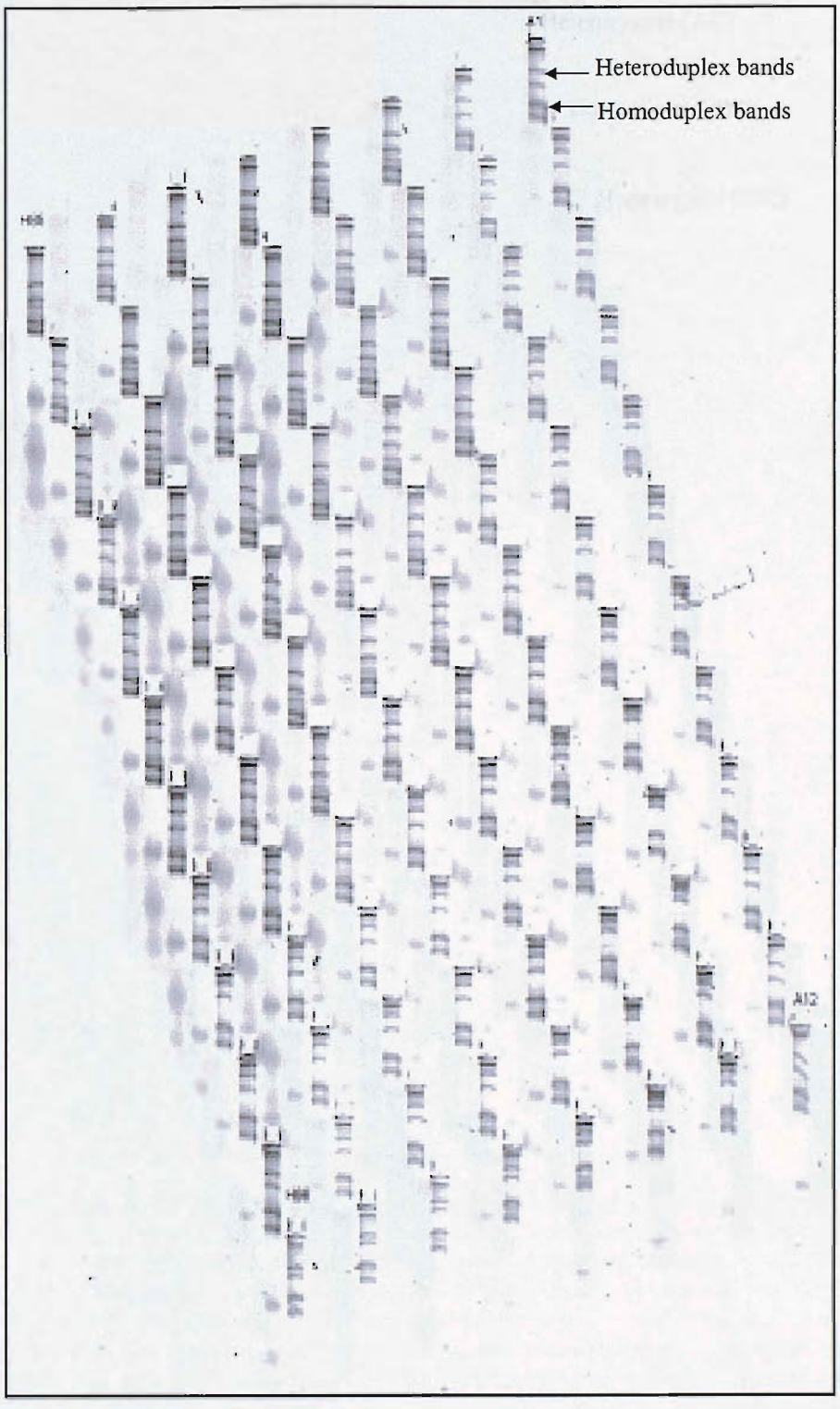


Figure 37.3. Melt-MADGE analyses. Represent amplified DNA from a single individual. DNA loaded across a 96-well plate shows equally clear resolution of heterozygous bands at all positions.

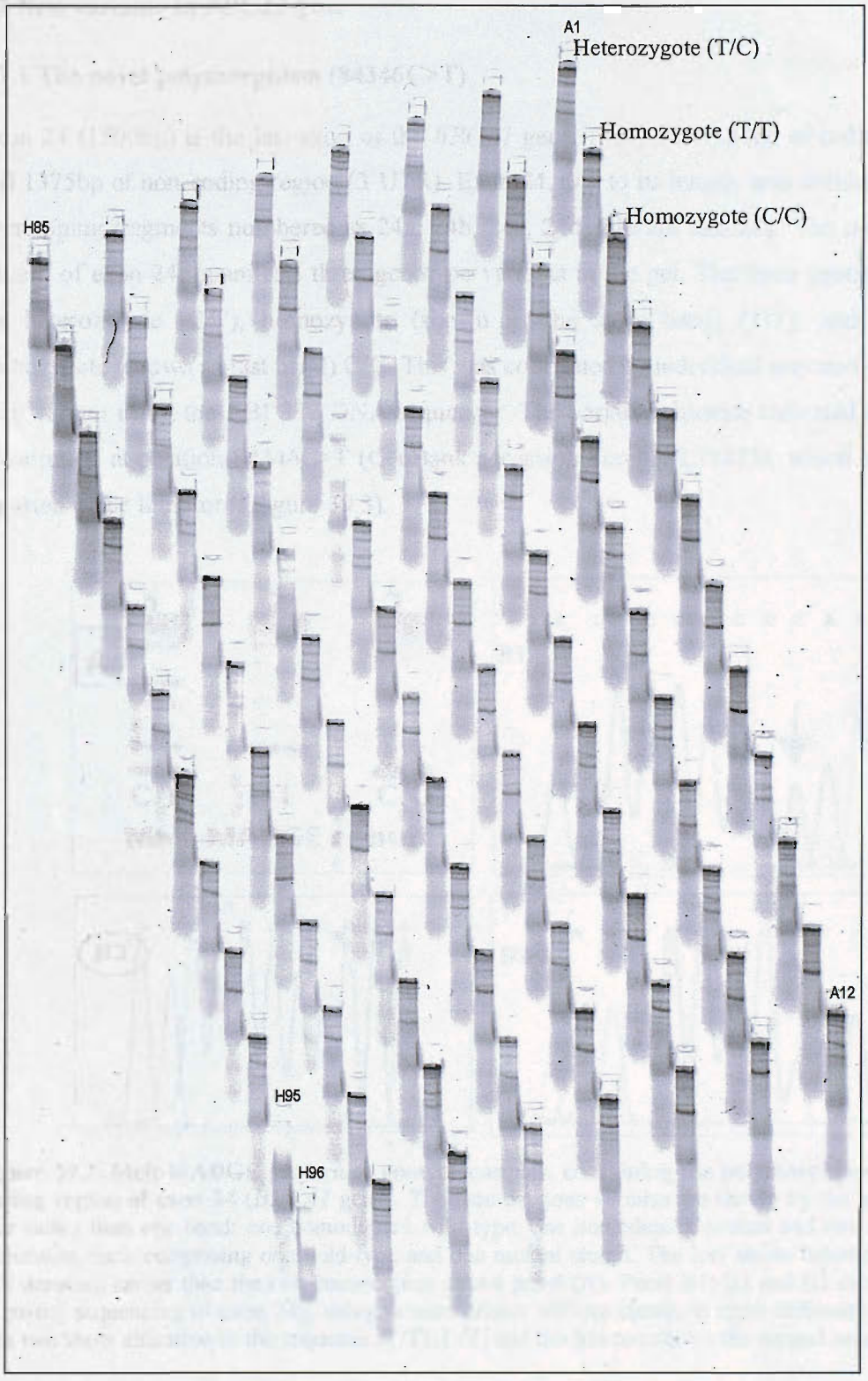


Figure 38.3. Melt-MADGE analyses. Variant bands pattern across the gel, with three different genotypes resolved. DNA loaded across a 96-well plate shows equally clear resolution of heterozygous and homozygous bands at all positions.

3.7 New variants in *BRCA1* gene

3.7.1 The novel polymorphism (84346C>T)

Exon 24 (1500bp) is the last exon of the *BRCA1* gene. It contains 125bp of coding sequence and 1375bp of non-coding region (3`UTR). Exon 24, due to its length, was divided into seven overlapping fragments numbered as 24a, 24b, 24c, 24d, 24e 24f and 24g. The melt-MADGE pattern of exon 24g manifests three genotype variants in the gel. The three genotype variants are heterozygote (C/T), homozygote (shown as the slow band) (T/T), and the second homozygote (shown as fast band) C/C. This was confirmed by individual sequence analysis for each variant using the ABI 377 DNA sequencer. The variant sequence indicated a nucleotide substitution at position 84346C>T (GenBank accession number L78833), which has not been reported in the literature (Figure 39.3).

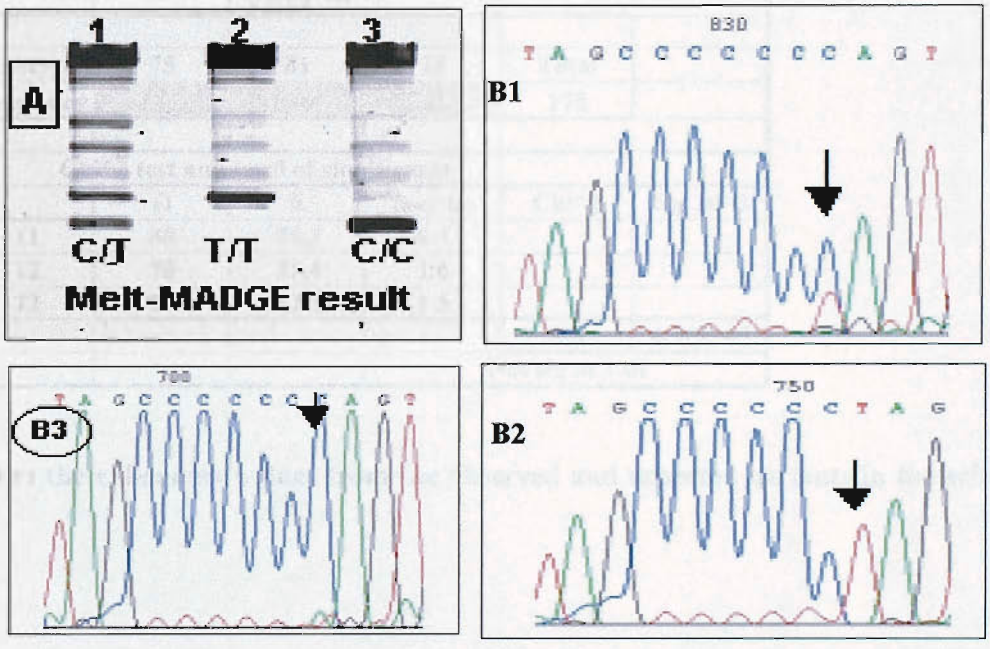


Figure 39.3. Melt-MADGE pattern of control samples, containing the polymorphism in the non-coding region of exon 24 (*BRCA1* gene). The heterozygous variants are shown by the appearance of four rather than one band: one homoduplex wild type, one homoduplex mutant and two heteroduplex molecules, each comprising one wild-type and one mutant strand. The less stable heteroduplex strand will denature earlier than the two-homoduplex strand panel (A). Panel B1, B2 and B3 shows the result of partial sequencing of exon 24g, using reverse primer without clamp, in three different subjects. The first two show alteration in the sequence (T/T), C/T) and the last one shows the normal sequence (C/C).

Statistical analysis of the polymorphism

178 samples from anonymous breast cancer patients were examined for the exon 24G variant. Hardy-Weinberg equilibrium is usually fulfilled in unselected random population samples and often in other sets, the exon 24-g variant showed an approximate agreement between the observed and expected values in all genotype patterns (Table 11.3).

		Observed			
Genotypes	11	12	22	Total	(Dropouts)
Frequency	80	70	28	178	0
	p	q	p+q		
	0.65	0.35	1		
		Expected			
Genotypes	75	81	22	Total	
Frequency	74.3	81.4	22.3	178	
Chi^2 test and level of significance					
	O	E	(o-e)^2/e	Chi^2	Sig level
11	80	74.3	0.4		
12	70	81.4	1.6		
22	28	22.3	1.5		
				3.5	6.16E-02
			Not sig at 1 df		

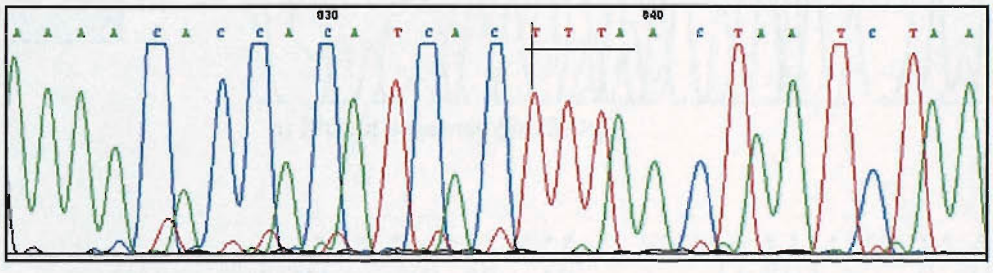
Table 11.3 Shows the chi-square values from the observed and expected variants in the selected population.

3.7.2 New *BRCA1* mutations

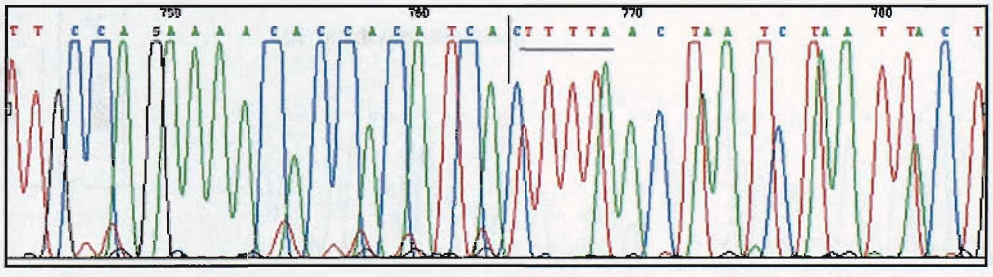
Three single nucleotide substitutions were found in our analyses that have not been reported in breast cancer information core (BIC) nor in the literature. By Melt-MADGE method two mutations were identified in exon 11 and one was identified in exon 17.

a) 5143C>T

Three different samples were observed to have this alteration in the expected sequence from the first 100 samples that had been received from the Wessex Regional Genetics laboratory. Using melt-MADGE method, different patterns of band velocity where displayed in the denaturant gel. These samples were confirmed by DNA sequencing. Single nucleotide substitutions 5143C>T (T1675I) in exon 17 was observed in three samples 11, 12 and 67 (Figure 40.3).



Wild-type sequence of exon 17



Mutant sequence of exon 17

Figure 40.3 Sequence analysis for exon 17 in *BRCA1* gene from patients with 5143C>T. the upper sequence shows the normal DNA sequence for the portion of exon 17, in which the alteration in single nucleotide occurred. The lower sequence shows the mutant DNA sequence.

b) 1967T>C and 2413A>C

Different genotypes were identified in exon 11 of the *BRCA1* gene. Nucleotide sequencing revealed two new variant sequences in two fragments of exon 11, fragment F and H. Both were missense mutations in exon 11 that result in an amino acid substitution. The two mutations in the patients are 2413A>C (E765A) and 1967T>G (S616C) Figure 41.3.

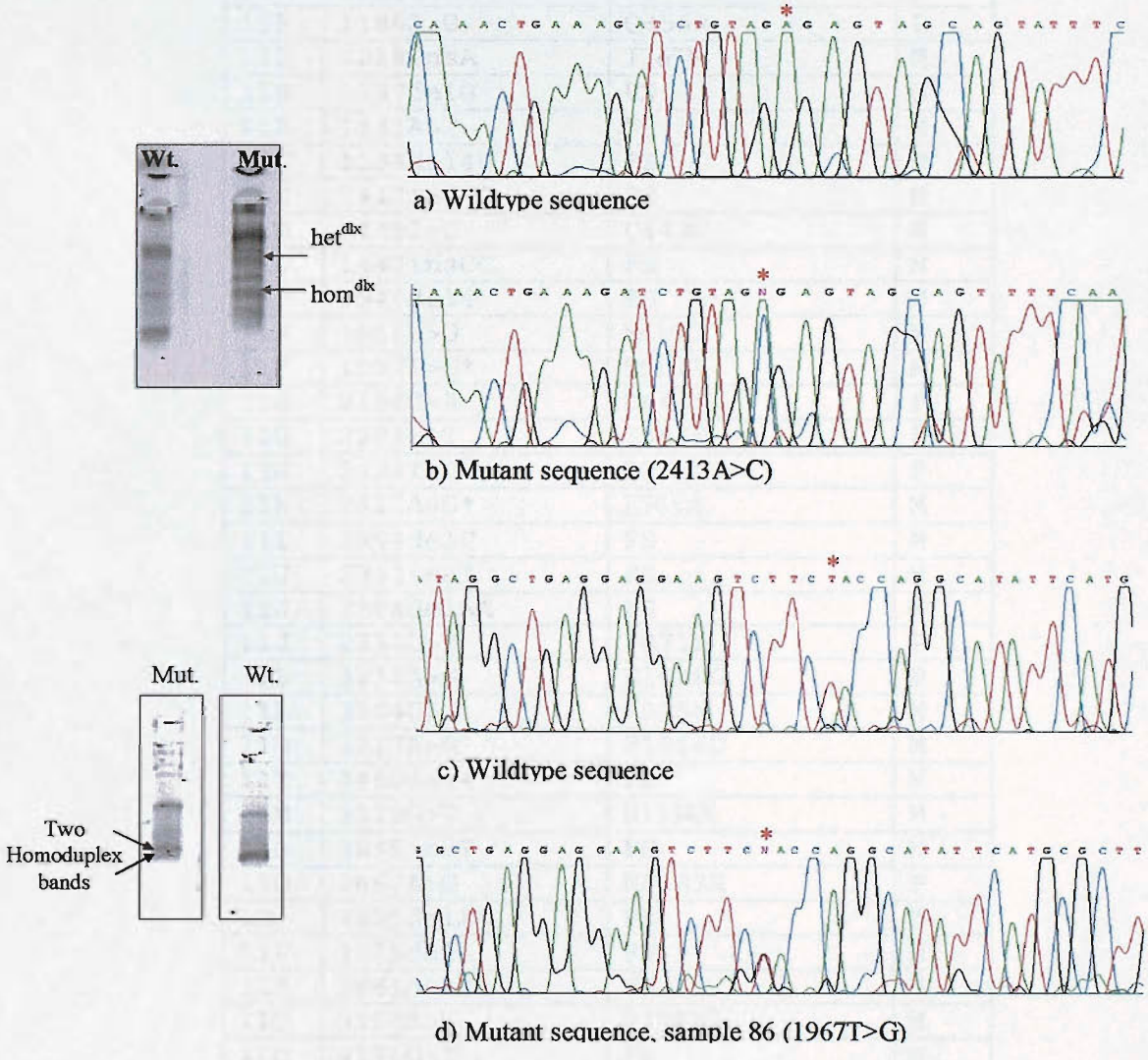


Figure 41.3 Identification of the new missense mutations 2413A>C and 1967T>G. The gel tracks on the left panel presented melt-MADGE results of the new variants. Wildtype and mutant chromatogram sequences illustrated to confirm the different genotypes detected by melt-MADGE method. The arrows on the tracks pointed to the split bands by denaturant conditions and compared to the wildtype band in the same gel.
het^{dlx} = heteroduplex, hom^{dlx} = homoduplex

3.8 Summary of the mutations and SNPs detected by Melt-MADGE method

Thirty-eight variants were detected by melt-MADGE and confirmed by DNA sequencing. Many of these variants were previously identified and reported in the BIC site and few new variants were identified in my study summarised in Table 12.3.

Exon	Nucleotide change‡	Amino acid change	Type
2	185delAG	FS	M
11B	1186A>G	Q356R	P
11B	1218insA	T367N	M
11B	1137delG	FS	M
11B	1131A>T	FS	M
11C	1294del40	FS	M
11C	1417insCC	FS	M
11D	1445T>C	C442C	M
11D	1447insCC	FS	M
11F	1940del4	FS	M
11F	1966C>G	S616C	M
11F	1967T>G*	S616W	M
11G	2196G>A	D693N	P
11G	2201C>T	S694S	P
11H	2430T>C	L771L	P
11H	2413A>G*	E765A	M
11I	2594delC	FS	M
11J	2731insT	FS	M
11J	2804delAA	FS	M
11J	2731C>T	P871L	P
11L	3232A>G	E1038G	P
11L	3104G>C	K995N	M
11M	3317A>C	E1066D	M
11N	3450del4	FS	M
11N	3519G>T	E1134X	M
11O	3695insT	FS	M
11O	3667A>G	K1183R	P
1O	3825delAA	FS	M
11P	3875del4	FS	M
11P	3881delGA	FS	M
11Q	4158A>G	R1347G	M
11Q	4176G>T	FS	M
11Q	4158delAG	FS	M
13	4427T>C	S1436S	P
16	4956A>G	S1613G	P
17	5143C>T*	T1675I	M
20	5382insC	FS	M
20	5396+47ins12	Intronic site	M
24G	5711+1286C>T	3`UTR	P

Table 12.3 Mutations and polymorphisms identified in breast cancer patients using melt-MADGE.

‡ = GeneBank accession number U14680, *= New variants; M =mutation; P= polymorphism; FS = Frame-Shift.

3.9 The Amplification Refractory Mutation System (ARMS)

PCR primers for five common Single Nucleotide Polymorphisms (SNPs) of the *BRCA1* gene, 2201C>T, 2430C>T, 2731C>T, 3232A>T and 3667G>A, were used in this study for the amplification refractory mutation system (ARMS) method. Two separated PCR amplification reactions were performed to allow identification of homozygous and heterozygous samples. The primers used in this analysis specified the amplification of different regions in exon 11 of the human *BRCA1* gene. Prior to screening, the samples for the mismatches and optimisation for $MgCl_2$, upper control primer, and annealing temperature were achieved. Figure 42.3 showed some representative results obtained with gradient annealing temperature from 50°C to 65°C.

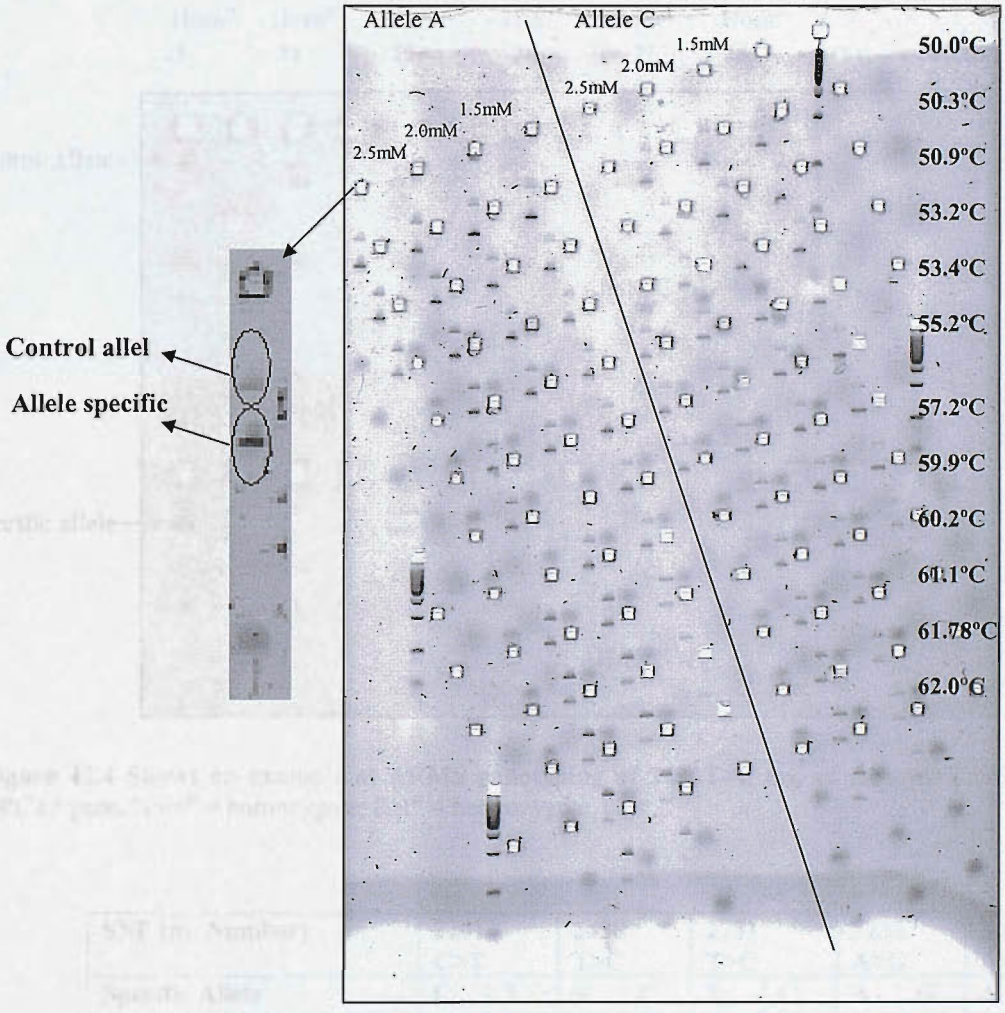


Figure 42.3 Optimisation of results for (2430A>C) SNPs. In this gel, two alleles (allele A and allele C) were run in 5% PAG gel after PCR optimisation with annealing temperature starting at 50°C and ending at 62°C. PCR product was amplified at different Mg^{+2} ion concentrations ranging from 1.5mM to 2.5mM.

Annealing temperatures and titration of $MgCl_2$ were used for the desired fragments of DNA molecule. PCR amplifications were satisfactory without detectable non-specific bands. PCR amplification of individual DNA samples for 2430T>C SNP, as an example, were separated on a 5% non-denaturing acrylamide gel and visualised with Ethidium bromide (Figure 43.3). The homozygous and heterozygous samples were easily distinguishable. Sample 23 represents wild type homozygous while sample 24 represents the homozygous abnormal allele. The heterozygous sample appeared as two bands in each allele and that can be seen in sample 25 and 26. The results obtained with all five mismatches were at the same resolution as in Figure 43.3 at different annealing temperatures and also $MgCl_2$ concentration and upper control primer concentration (Table 13.3).

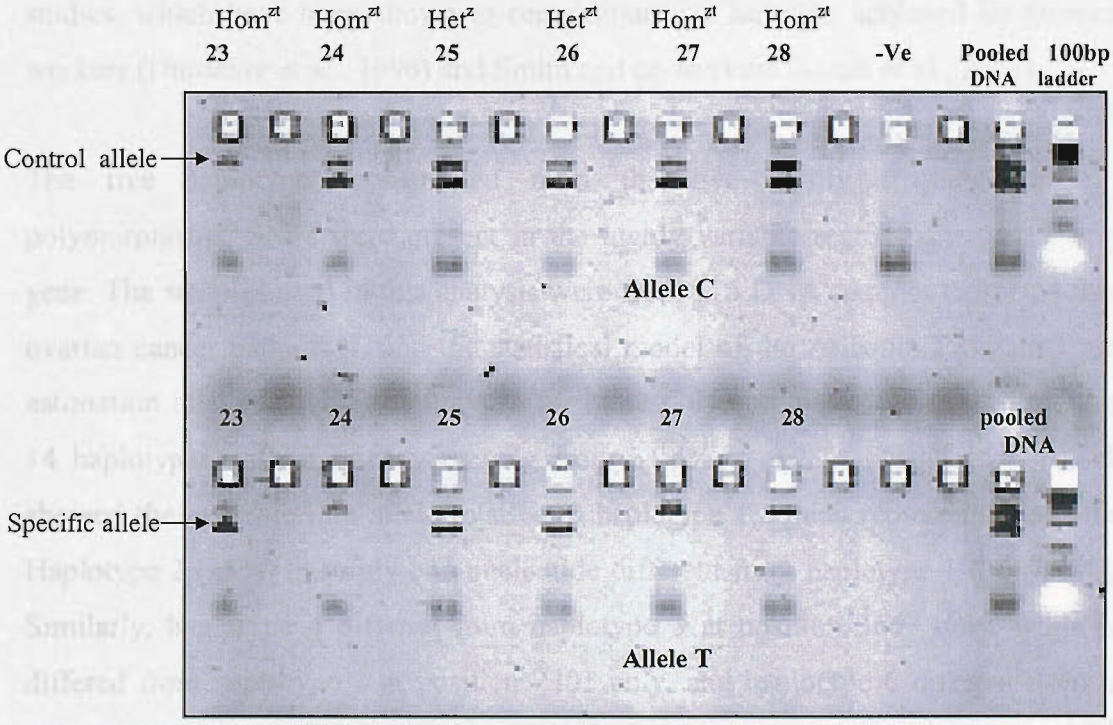


Figure 43.4 Shows an example of ARMS genotyping of 2430T>C one of the common SNP in exon 11 of *BRCA1* gene. Hom^{zt} = homozygote, Het^{zt} = heterozygote

SNP (nt. Number)	2201 C>T		2430 T>C		2731 T>C		3232 A>G		3667 A>G	
Specific Allele	C	T	T	C	T	C	A	G	A	G
Annealing temp. (°C)	56	56	57	57	59	59	56	56	62	62
$MgCl_2$ (mM)	2.5	2.5	2	2	2.5	2.5	2.5	2.5	2	2
Upper control volume (µl)	2		2		1		4	8	3	

Table 13.3. Optimised ARMS result for the common SNPs in exon 11 of *BRCA1* gene.

3.10 Statistical analyses for common polymorphisms: genotype analyses

176 samples were genotyped by melt-MADGE and ARMS methods for five common polymorphisms in exon 11 of the *BRCA1* gene. Three polymorphisms resulted in amino acid change, 2732C>T, 3232A>G and 3667A>G corresponding to Pro871Leu, Glu1038Gly and Lys1183Arg respectively, and two polymorphisms showed no amino acid change, 2201C>T and 2430T>C (silent mutation). Hardy-Weinberg law was applied to calculate the frequencies of these polymorphisms. In Table 14.3, allele frequencies for common polymorphisms observed in my study were compared with allele frequencies described in previous studies (Smith et al., 2001; El-harith et al., 2002; Markoff A. et al., 1998; Durocher et al., 1996). Polymorphism frequencies in this study are in agreement with the frequency observed in other studies, which have been shown to run similar size samples, achieved by Durocher and co-workers (Durocher et al., 1996) and Smith and co-workers (Smith et al., 2001).

The five haplotypes constructed from the five highly frequent single nucleotide polymorphisms, SNPs were present in the highly variable region of exon 11 of the *BRCA1* gene. The samples used in this analysis were from 176 DNA samples extracted from breast or ovarian cancer patients. Using the statistical model of the Arlequin Program I performed an estimation of the different haplotypes for these polymorphisms. Arlequin Program generated 14 haplotypes with frequency ranging from 31.4% to 0.28% (Table 15.3, *a*). Haplotype 3 showed the opposite rare allele relative to haplotype 1, which represents the common allele. Haplotype 2 (25%) has only one nucleotide different from haplotype 1 (3667A>G, K1183R). Similarly, haplotype 4 differed from haplotype 3 at position 3667 only, while haplotype 5 differed from haplotype 2 at position 2201 only, and haplotype 6 differed from haplotype 3 only at positions 2201 and 3667. Together haplotypes 1 to 6 account for 95.28% of all haplotypes.

Polymorphism ID	Allele frequency (%)					
	in our study (Case-based study) 176 Ov /BC patients	Durocher et al. Ov/BC patients	Smith et al. (Unselected clinical-based study) 258 USA women of OV. CA	Markoff A. et al. (Based on family history) 20 Bulgarian women of BC	El-harith etal. (Unselected cohort with BC) 29 Arab and 11 Asian	
					Arab	Asian
2201C>T	34	111/364 (32%)	-----	10	31	32
2430T>C	33	46/144 (32%)	30.6	10	31	32
2731T>C	50	90/214 (42%)	-----	10	31	32
3232A>G	32	80/234 (34%)	31.4	10	31	32
3667A>G	34	57/182 (32%)	28.7	10	31	32

Table 14.3 Represents frequency of the common polymorphisms in exon 11 of the *BRCA1* gene.
Ov. Ovarian cancer; BC; breast cancer.

However, the haplotypes produced by Arlequin method were compared with the haplotypes results obtained with the Phase method (Table 15.3,b). There were noticeable mild discrepancies between the two-statistical analyses for the same SNP. The dissimilarity comparison could be distinguished in haplotypes 1 (31.4%) and 2 (25.0%) in the Arlequin method and haplotype 1 (37.1%) and 3 (20.8%) in the Phase method. This discrepancy might relate to the specification of the two methods, in which the samples size is one of the important parameters.

a- Arlequin analysis

Haplotype No.	2201	2430	2731	3232	3667	Frequency (%)
1	1	1	1	1	1	31.4
2	1	1	1	1	2	25.0
3	2	2	2	2	2	14.68
4	2	2	2	2	1	8.38
5	2	1	1	1	2	8.08
6	1	2	2	2	1	7.74
7	1	1	1	2	2	1.41
8	2	2	2	1	1	1.18
9	2	2	1	2	2	0.46
10	1	2	1	2	1	0.38
11	2	1	1	2	2	0.34
12	1	2	1	1	1	0.29
13	2	1	2	2	2	0.29
14	2	1	1	2	1	0.28
Total						100

b-Phase analysis

Haplotype No.	2201	2430	2731	3232	3667	Frequency (Absolute)	Frequency (%)
1	1	1	1	1	1	130	37.1
2	2	2	2	2	2	73	20.8
3	1	1	1	1	2	71	20.3
6	2	1	1	1	2	25	7.1
5	1	2	2	2	1	22	6.3
4	2	2	2	2	1	13	3.7
8	1	1	1	2	2	4	1.1
12	2	2	2	1	1	4	1.1
7	1	2	1	2	1	3	0.9
10	2	1	1	2	2	2	0.6
9	2	1	2	2	2	1	0.3
11	1	1	1	2	1	1	0.3
13	1	2	1	1	1	1	0.3
Total						350	100

Table 15.3 Analysis of haplotype of five SNPs in exon 11 of *BRCA1* gene by use of the Arlequin method and phase method.

3.11 cost-effectiveness estimation

Table 16.3 represents Melt-MADGE cost analysis for known or unknown variants compared with the other techniques that have been recently studied by Sevilla (Sevilla et al. 2002). The highest number of fragments analysed by the current method best performing (dHPLC) is 25,606 fragments screened per year. However, our cost estimation demonstrates that melt-MADGE shows the lowest cost screening method with at least 10 000 fragments per week and 210000 fragments per year, at a cost of only 1.35€ per sample.

Cost Factor	DS	DHPLC	SSCP	DGGE	HA	FAMA	PTT	Melt-MADGE
Consumable reagents	16.7 €	1.4 €	3.6 €	1.3 €	1.9 €	7.90 €	7.2 €	1.12€
Equipment	5.20 €	1.6 €	0.3 €	0.2 €	0.1 €	6.70 €	0.2 €	0.06€
Personal	7.20 €	0.9 €	2.3 €	3.0 €	1.0 €	12.6 €	1.7 €	0.17€
Total cost	29.1 €	3.9 €	6.2 €	4.5 €	3.0 €	27.2 €	9.1 €	1.35€

Table 16.3 Shows cost comparison of different techniques for mutation analysis of individual fragment of *BRCA1* gene Adapted from Sevilla C. 2002 (Sevilla et al. 2002) to which we compared our approach, melt-MADGE, in cost-effectiveness context.

CHAPTER FOUR

DISCUSSION

Melt-MADGE technique extends the capabilities of MADGE to *de novo* scanning, whether for common or for rare variations, and accesses variation for which no easy direct typing method can be, or has been, devised. The combination of temperature control with MADGE, therefore, offers a low technology platform for the study of several categories of genetic marker. It also enhances our ability to explore single nucleotide variations at large population scale where parallel analysis of many samples is required. The ultimate goal of this research is to develop a new method for a high throughput low cost screening strategy to enhance the detection of variation in DNA sequence. Certain prevailing circumstances encouraged the pursuance of this project:

1) There is high demand for mutation detection of genetic variation. The human genome project revealed approximately 30,000 – 40,000 genes in the human genome (Lander et al. 2001), more than 1000 genes implicated in inherited human diseases have been identified and reported in the Human Gene Mutation Database (Krawczak et al. 2000),
2) Growing knowledge of the genetic contribution to complex diseases (Schork 1997) requires large scale population studies for mutation detection.
3) There is lack of an ideal method for screening large populations and sample collections at the lowest level of expense and complexity and with the highest level of sensitivity and specificity.

4.1 Development of melt-MADGE method for high throughput mutation screening

Determining the genetic basis of complex diseases with DNA sequence variations is likely to be an important task. Complex diseases such as breast cancer require the screening of large samples of the susceptibility genes for all possible individuals. In this study, I introduced a new method for scanning for all possible point mutations in a large gene, *BRCA1*. However, this technique will not reveal the large deletions or insertions in which a qualitative method such as Multiplex Ligation dependant Probe Amplification (MLPA) (Slater et al. 2003; Hogervorst et al. 2003), or by the conventional method, linkage studies, will do. Various techniques have been developed and improved for mutation detection, e.g. combined SSCP/duplex analysis by

capillary electrophoresis has been modified into an automated combined CE-SSCP/duplex method, which provides a good high throughput analysis, in which 48 samples can be analysed in an ABI 310 within 20 hours. This could be increased up to 1000 or 6000 samples per day if an ABI 3100 or ABI 3700 were used (Kozlowski and Krzyzosiak 2001). Furthermore, the modified TDGS method presents a good method for screening a large variety of genes for all possible sequence alterations, e. g. *BRCA1* and TP53, however, its performance is 91 % detection rate compared to dHPLC (100%) (Bounpheng et al. 2003), in spite of that it does not seem to be the method of choice in terms of simplicity and high throughput methods. Various methods for mutation detection such as denaturant gradient gel electrophoresis (DGGE), single strand conformation polymorphism (SSCP), heteroduplex analysis (HA), protein truncation test (PTT), denaturing high-performance liquid chromatography (dHPLC) and DNA sequencing have all been developed as an assay for mutation detection. These techniques are useful in research but may be limited to more specialized molecular biology laboratories. The previously mentioned methods differ in specificity, sensitivity and simplicity and are not high throughput, because they are time-consuming and they are not cost-effective. Easier and cheaper methods are needed for clinical and population studies.

Melt-MADGE exchanges the dimension of conventional DGGE from vertical to horizontal and from a denaturant gradient in space to a denaturing thermal ramp in time. Electrophoresis of an array of 96 samples is enhanced by using MADGE format (Day et al. 1995b; Day et al. 1998) which is compatible with the 96-passive replicator and 96-well plates that allows very high throughput with a minimal risk of errors. Melt-MADGE accommodates 10 gels (96-well) per tank, so 960 PCR products can be analyzed within approximately four hours from preparation of the gels. The assay for mutation detection for the *BRCA1* gene requires no incorporated radiolabel or fluorescence materials in PCR amplimers as applied in the modified method, TDGS (Bounpheng et al. 2003). These additional labels are expensive and require special care. The Melt-MADGE technique (Day et al. 1999a) offers a potential method for scanning many samples of DNA sequence alterations and identification of rare or common sequence variants with high throughput. In this study, *BRCA1* gene was used as the model for the application of the melt-MADGE technique. I described a prototype system and the proof of principle through the melt-MADGE experiment and screened many DNA samples.

Long PCR assay was applied in other studies (van Orsouw et al. 1999; Bounpheng et al. 2003) for amplifying short PCR fragments. A multiplex short PCR assay used in a 2-D electrophoresis technique based on size in the first dimension and DGGE in the second

dimension, but that would not possible for a 1-D technique such as melt-MADGE. Our purpose of applying long-PCR procedures is to reduce consumption of genomic DNA and to exclude false mutation detection by generating a cleaner (less sequence complex) template than genomic DNA. Internal primers were designed (Appendix A) to cover all the *BRCA1* coding sequence. Long PCR product was diluted as 1/100, which is utilized as DNA template for the second PCR reaction. Long PCR in the *BRCA1* regions were successfully optimised and amplified for further internal short PCR amplification. The average fragment length of nucleotide sequence for screening by melt-MADGE is 300bp, and if the average coding sequence is approximately 2000bp per gene (Bounpheng et al. 2003) that could be split into approximately 6 fragments for melt-MADGE, unlike the recently modified TDGS method which requires around 20 fragments for analysis (Bounpheng et al. 2003). This approach will enhance the scanning of large populations with minimal genomic DNA. This may be useful for small diagnostic samples but will be most important in accessing research DNA banks (e.g. to understate 'reference range' studies) where template is much more limited.

The purpose of this work was to develop improved basic concepts of the melt-MADGE method. However, an assay for any small alteration in the DNA sequence in the *BRCA1* gene has been developed. This assay demonstrated reliable results for detection of multiple variants in the DNA sequence. There are polymorphic changes which can be successfully detected by Melt-MADGE analysis, e.g. Q356R and 2201T>C in exon 11. In addition, this technique is capable of detecting changes in the DNA sequence that result from deletion/insertion alterations. Such as the 185AGdel and 5382insC in the *BRCA1*. A large deletion in a given gene affects the melt-MADGE assay. That changes the conformation of DNA strands, which would change the stability of them, i.e. heteroduplexes are less stable compared to homoduplexes. This can be easily detected at lower melting temperature than the actual melting temperature for a small deletion. Nevertheless, the large deletion can also be detected by non-denaturing gel, as I detected the 1417del40bp in exon 11, Figure 32.3. This leads to underestimate the optimal electrophoresis temperature for that amplicon and missing detection of mutation, if there is any. This method can be easily applied for identification of rare or common variation for any targeted gene under optimised conditions. The simplicity of melt-MADGE makes it a suitable method to set up in any molecular biology laboratory. The melt-MADGE method is a powerful screening test for thousands of samples per day.

4.2 Melting profile and temperature

To examine a nucleotide sequence variation in DNA molecules with a highly sensitive rate of detection using a denaturant method, an artificial GC-rich sequence is introduced into the targeted sequence. A melting map is frequently calculated by using special software to predict the stability distribution of a sequence and the location of the low melting domains as with the conventional Melt87 software (Lerman and Silverstein 1987). The conditions for successful mutation detection using the TIXIS program were determined by temperature analysis, which was based on the melt87 program. The TIXIS program has a similar function to melt87 calculating the melting temperature of the entered nucleotide sequences. TIXIS uses melt87 in its core, with a lot of useful annotation tools and sequence editing and melt profiling tools. It contains more features than the DOS-based melt87 program. Table 3.5 presents the close agreement between the predicted Tm (constant) by TIXIS based on the statistical theory for sequence specific melting evolution and the mobility of individual DNA fragments on denaturant polyacrylamide gel through electrophoresis in melt-MADGE. This correspondence showed the accuracy of this software for predicting the Tm for DNA sequence. The aim of planning a GC clamp is to achieve two melting domains: the first with a high melting temperature that is generated by GC-rich sequence attached to one end of the DNA sequence; and the second homogenous for AT/GC nucleotides (Figure 16.3), where any single nucleotide variation in a DNA sequence should gain a representative influence within the flat melting profile (Myers et al. 1985b). Theoretically, this will be identified with greater sensitivity in the lower melting domain (Myers et al. 1985a), because this will be the domain exerting the larger effect on transition of mobility in the gel.

A DNA sequence that has natural multiple repeats of GCs followed by relatively long sequences of ATs in one analysed domain will generate more than one melting domain (Myers et al. 1985b). Therefore, there are two reasons to divide nucleotide sequences into more than one fragment; firstly, the section of a DNA sequence in which AT-rich has less stability than a sequence of GC-rich, in which the alterations in GC-rich regions might lead to misidentified variants (Lerman and Silverstein 1987; Lerman 1987; Myers, Maniatis, and Lerman 1987). Thus, long nucleotide sequences with unequal distribution of AT/GC sequence in the coding region must be divided into different fragments to generate only two melting profiles (Figure 16.3) incorporating the defined melting profile. Secondly, some of the coding regions are very long nucleotide sequences, such as exon 11 of the *BRCA1* gene. That needed to be divided into more than one domain as overlapping fragments to ensure all coding sequences were screened and to be the appropriate sequence length that is compatible with the melt-MADGE method for

better resolution (see gene map in Appendix F). All the 54 amplicons of the *BRCA1* gene designed by TIXIS software showed two domains in the melting profile map. That maximized the chance of identification of any small alterations as electrophoretic mobility shifts

4.3 Testing and validation of melt-MADGE assay

Melt-MADGE assay was tested for five common SNPs in *BRCA1* gene (table 3-10) on 100 different DNA samples. Results were compared to the genotypes determined by ARMS assay. I had tested the assay on the six different previously known mutations in the *BRCA1* gene (table 3-5). These tests confirm the application of this method for screening large populations for unknown mutations. The result of this evaluation indicates the potential of this assay in screening large samples for diagnostic backlog and for association studies.The identification of *BRCA1* SNP 84346C>T represents the first *de novo* SNP identification by melt-MADGE and validates the application of the approach to the *BRCA1* gene specifically. The samples showed a novel polymorphism in exon 24. This is a single nucleotide substitution in the 3`-non-coding region. Alterations in the 3`-region are predicted to be potentially important in the stability of the mRNA (Conne, Stutz, and Vassalli 2000). The *BRCA1* 84346C>T SNP was observed in a sample of 20 random normal anonymous volunteers of mixed ethnicity and nationality. At present, it seems likely that it is a common SNP without pathogenic significance or functional effect, but it could plausibly belong to a haplotype with small functional consequence. This could be examined in the future by large-scale case-control or family-based association studies. Deviation from Hardy-Weinberg equilibrium may be due to laboratory technical artefacts or factors that disrupt Hardy-Weinberg equilibrium such as non-random mating, gene flow, natural selection and mutation. In addition, there are some factors that could result in an increased number of heterozygotes such as heterozygous advantage. In this study, there is no significant departure from Hardy-Weinberg equilibrium although slightly more heterozygotes were observed than expected. Further studies will be necessary to determine the role of the polymorphism at position 84346C>T, particularly in relation to the association between the *BRCA1* haplotype and late onset breast cancer. The identification of this common SNP is incidental to the development but validatory to my developments to facilitate rare mutation scanning in large cohorts of breast cancer subgroups.

4.4 DNA melting in denaturing gel electrophoresis at constant and ramping temperature

Using denaturant conditions with a gel electrophoresis-based method enhanced the ability to detect small alterations in DNA molecules. Identifying single nucleotide changes of DNA fragments in denaturing gel is achieved by applying methods such as DGGE and CDGE that are based on thermodynamic properties of DNA duplexes. Changing the physical shape of DNA molecules during electrophoresis is enhanced by partial melting in the presence of urea and relatively high temperatures (55°C – 70°C).

The melting temperature for a DNA molecule predicted by the TIXIS program will be reduced during the electrophoresis run by 2.5°C/1M of urea (Lerman and Silverstein 1987). The results show that the melt-MADGE method can be successfully used for variant detection of alterations in a horizontal gel. A number of known variants were identified in the 54 fragments, in the *BRCA1* gene running at constant temperature. DNA mobility in denaturing polyacrylamide gel does not rely on the molecular weight of the DNA molecules, but it depends on the constitution of the nucleotide sequence. Thus, a melt-MADGE experiment separates DNA fragments differing by a single nucleotide substitution or by a mismatch. The buffer temperature in the melt-MADGE tank and gel-containing urea was homogenous for each individual amplimer that had been optimised. The GC-clamp introduced by the PCR primer at one end of the DNA fragment prevented complete dissociation of the DNA molecule. The amplicon started to separate at the beginning of the electrophoresis. Each molecule begins its mobility at the top of the gel, migrating faster until the intact DNA double helix starts to partially denature and then falls to low mobility as the first domain melts. Sequence variation in different DNA molecules is detected by the difference in the final position of the bands of the DNA molecules running under similar conditions in one denaturing gel.

The experiments that were run at the constant predicted temperature showed good band resolution once the right melting temperature was identified. Nevertheless, repeating the experiment under similar conditions will often fail to produce a similar band pattern because even 0.5°C in the achieved run condition exerts a large impact on relative mobilities of heteroduplexes and homoduplexes. This divergency might result from uncontrolled effect on the DNA molecule, in which the DNA starts to melt in the loading well at the start of electrophoresis temperature. This, perhaps, allows the strands to tangle with each other and /or the gel. This similar phenomenon has also been observed in an early study by Lyamichev V. I. et al., 1982 (Lyamichev, Panyutin, and Lyubchenko 1982) in which they created denaturing regions in the DNA molecules that retard the DNA fragment at the site of the local denaturation in denaturing gel. These interpretations of the early melting of the DNA molecule at a constant temperature in denaturant polyacrylamide gel explain the plateau spot on the top of the gel.

The target of my investigation was aimed specifically to improve the utility of the method for the detection of unknown mutations. The primary consideration was the optimisation of detection conditions. After PCR amplification, mutation scanning requires an initial optimisation of an individual amplicon electrophoresis temperature. This takes three to four runs to check or confirm the predicted optimal condition, e.g. urea concentration and temperature, for each amplicon. Therefore, application of strictly controlled temperature for melt-MADGE is critical. In some fragments, a 1°C change would greatly alter the melting profile. This was observed in amplicon 11O (3619-3844), which has a common polymorphism at nucleotide 3667A>G. By contrast, a temperature ramping method gives more consistent band separation in denaturing gel. Theoretical illustration in Figure 44.4 presents the melting curve of temperature ramp (even if 0.5°C error for a ramp range of 4-6°C) collects an integrated mobility, profile for all moieties compared with the 'hit or miss' effect where a constant temperature 'snapshot' is run.

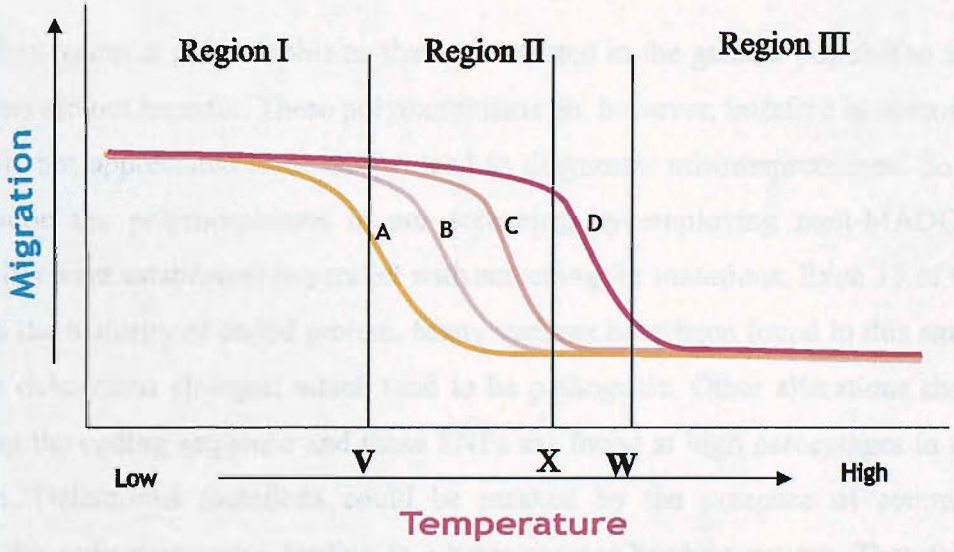


Figure 44.4 Illustrates the theoretical mobility curve of the mutant type using a temperature ramp in time in melt-MADGE method. A and B curves represent the first separated two bands of heteroduplex, and C and D curves represent the last separated two homoduplexes bands that required higher temperature than the heteroduplexes. In region III, if the temperature is higher than expected then the amplicon will be completely denatured, but if the temperature hits the optimal melting temperature the good resolution of the band will be displayed in the gel, region II, in contrast, if the amplicon runs at lower temperature the amplicon will migrate faster and will present a single band, region I. Application of constant temperature for screening mutations in DNA sequence will only detect the changes if the optimal melting temperature is hit directly and will then be very high resolution. In contrast, employing ramp temperature will help resolve all moieties A-D. For example, at temperature X, C and D would be very well resolved but A and B would not be resolved. At temperature W, no sufficient resolution would be achieved. By contrast, a ramp from V to W would capture (driving same part of the ramp period) each of the differential mobilities of moieties A-D for a part of the time, maximising the chance of resolving at least some of them.

Ramping temperature has been applied to methods used for screening point mutations, e.g. TGGE (Vijg and van Orsouw 1999). Ramping temperature enhances melt-MADGE electrophoresis and improves reproducibility of the results. Although the experimental procedure is still similar to the procedure that I used at constant temperature, including the gel set up and preparation of the electrophoresis buffer, the difference was that I converted the temperature in the tank from a snapshot to a ramped temperature with time (gradient). Using this refinement, I successfully obtained similar results by running the same amplicon in different times during melt-MADGE experiments. The ramping method simplified the technique by improvement of reproducibility of the melt-MADGE assay. The method utilised a ramping temperature between the initial and the final temperature at the rate of approximately 0.05°C / minute.

4.5 Combined ARMS assay with melt-MADGE method

The *BRCA1* gene has common polymorphisms that are detected in the general population and these polymorphisms are not harmful. These polymorphisms do, however, interfere in scanning assays and could (if not appreciated to be SNPs) lead to diagnostic misinterpretations. So to diminish and exclude the polymorphisms in our screening by employing melt-MADGE, ARMS tests for SNPs were established in parallel with screening for mutations. Exon 11 of the *BRCA1* gene forms the majority of coded protein. Many variants have been found in this exon, some of which are deleterious changes, which tend to be pathogenic. Other alterations show mild or no effect on the coding sequence and these SNPs are found at high percentages in the general population. Deleterious mutations could be masked by the presence of common polymorphisms in the coding sequence leading to a heterozygous banding pattern. Therefore, to pursue the aim of high throughput and low cost for analysis of hundreds of samples, I established ARMS assays for five SNPs in exon 11 to exclude the subjects that show these SNPs. The combination of Melt-MADGE and ARMS techniques improves the overall evaluation in regions harbouring SNPs. However, SNPs mapped by ARMS assay had come out positively with the melt-MADGE assay. The assay had high accuracy (100%) correlation between the two methods, melt-MADGE and ARMS.

4.6 Sensitivity of melt-MADGE to heterozygous sequence variation

The sensitivity of the melt-MADGE method was tested in two types of variation detection in human genomic DNA samples. First, the assay was analysed by performing blind analysis of 96 samples previously screened for pathogenic mutations at the Wessex Regional Genetics laboratory. Melt-MADGE identified 14 out of 17 mutations (table 3-6). Three mutations were not resolved by melt-MADGE, 1445T>A, 2773delTC and 2804delAA which had been identified by DNA sequencing, PTT and HA methods, respectively. For the region of 1445T>A alteration, 1429-1658, I lacked a positive control during assay optimisation. However, melting temperature for this region was determined by running the amplicon several times in different ramping temperatures to identify a suitable ramp range. This single nucleotide substitution was located at seven nucleotides (+7) from the clamped primer and this may have limited its potential to influence melting of the assay domain. Failing to identify the variant may result from the influence on stability of specific non-Watson-Crick base pairs and the effect of mismatches on the stability of the formation of spontaneous mutation. Ke and Wartell (Ke and Wartell 1993) had shown that G.T, G.G and G.A are the most stable mismatches, therefore, these mismatches reduce the stability of the first melting domain by 1 to 4°C. An unpaired base identical to one of its adjacent bases caused less destabilisation than an unpaired base with an identity differing from its neighbours (Ke and Wartell 1995). However, I was able to identify the failed amplicon (tested sample) when re-run in parallel with positive amplicon by melt-MADGE. Lack of sensitivity to two mutations in the region 2634-2867 base pair (exon 11 fragment J) of this blind analysis was unexpected because this region was optimised correctly and a common polymorphism, 2731C>T was readily detected. These undetected mutations might refer to the failure of the melt-MADGE assay for this individual amplicon for some reason. The estimated sensitivity of this method by taking into account these failure mutations being missed, 1445T>A, 2773delTC and 2804delAA, will be 82.4%. But if we considered only 1445T>A the mutation that was unresolved, the sensitivity becomes 94.1%. A second indicator of the utility of melt-MADGE is discovering a new polymorphism in exon 24, 84346C>T, as explained above in Section 4.5. This was confirmed subsequently by sequencing the relevant region. It would be feasible, if many amplicons (from any genes) of similar %GC richness were examined in parallel, to extend the technique for efficient SNP scanning; it would also be an economical and rapid approach to scan for frequent SNP. In addition, twenty-three mutations and nine polymorphisms (table 12.3), which have been recorded in the Breast Cancer Information Core (BIC), were identified by the melt-MADGE method.

It is clear that factors such as base composition, sequence context, position of the base change in the analysed sequence, and fragment length all influence the sensitivity of the conformation-based methods (Markoff et al. 1998). However, melt-MADGE presents good resolution for most amplicons, which were analysed. Length of the amplicon and the concentration of the gel were observed to affect the resolution of the bands on denaturant acrylamide gel. Exon 16 is relatively long to run as one fragment in melt-MADGE technique. Therefore, it was split into two fragments of 445bp and 337. These two fragments electrophoresed in 5% and 6% polyacrylamide gel, respectively. The result showed quite different band patterns mobility (Figure 29.3). The resolution of the band in the long sequence (445bp) is demonstrated as two homoduplex bands separated from each other at the bottom of the lanes. The two-heteroduplex bands were not evident and probably were masked by an artefact band that co-migrated in the gel. In contrast, the 337bp fragment appeared to resolve as three bands, two heteroduplex bands and one homoduplex band in 5% polyacrylamide gel, which the resolution of the heteroduplex is distinguishable in two gel concentrations, 5% and 6% polyacrylamide gel.
The artefact band observed in fragment 445bp results from the effect of non-specific bands co-migrated with the amplicon band in denaturant gel. These artefact bands might be generated during amplification by an error in *Taq* DNA polymerase or position effect of a base alteration in the sequence.

4.7 Correlation between mutation in *BRCA1* gene and Breast Cancer

Different mutations have been seen in many cases with expression of the phenotype. Mutations such as 185delAG and 5382insC in *BRCA1* gene were seen commonly in individuals of Ashkenazi Jewish origin (Ghaderi et al. 2001; Backe et al. 1999). However, not all patients with breast cancer exhibit a mutation in either *BRCA1* or *BRCA2*. In a study screening 38 cases with either ovarian or breast cancer and a family history of breast or ovarian cancer, 50% of these patients had no mutation identified by sequencing the coding region of the two genes (Schoumacher et al. 2001). Furthermore, in the same study it was shown that an individual harbouring a known mutation, 185delAG and 6174delT in the *BRCA1* and *BRCA2* genes respectively, exhibited no breast cancer and remained unaffected to age 36 years, while her sister with 185delAG developed breast cancer at age 30 years. A male with both mutations did not develop cancer and died at age 70 (Moslehi et al. 2000). From these observations, a mutation in a single gene such as *BRCA1* may not be sufficient to cause cancer. Other genes than *BRCA1* or *BRCA2* are also likely involved in breast cancer. From the functional point of view, *BRCA1* shows a distinct domain structure comprising an N-terminal RING finger domain and two BRCT domains at the C-terminus. The RING finger domain has been shown to

mediate interactions with both BARD1 and ubiquitin C-terminal hydrolase BAP1 (Jensen and Rauscher, III 1999). The BRCT domains have also been shown to mediate a number of other protein-protein interactions. Any of these interacting gene products might be candidates for germline breast cancer predisposing genes or for genes involved in 'sporadic' breast cancer.

4.8 Common polymorphisms and Haplotype analysis

Five common SNPs were genotyped (table 3-10) and compared with other studies (Durocher et al. 1996; Smith et al. 2001; Markoff et al. 1998; El Harith et al. 2002). These studies were in different populations and the sample sizes are not in close agreement. Allele frequency estimates for individual SNPs might differ by ethnicity or because there may be frequency estimation errors in small collections. However, the 2731 SNP was present at the highest allele frequency in our study, 50%, and in the Durocher study with a frequency of 42%, compared to others in similar studies. Furthermore, considering the population of the three studies, my study, Durocher and Smith studies, was in patients with breast and ovarian cancer. By contrast, the populations typed in the Markoff and El-harith studies were mainly composed of breast cancer patients. In the Markoff study all the SNPs showed 10% allele frequency for the 20 Bulgarian women only expressing breast cancer, and the El-harith study with 29 patients with breast cancer. The possible explanations for observed these SNPs frequencies in those populations are: 1) the high frequency of these SNPs is correlated to breast and ovarian cancer, 2) the low frequency of the SNPs in the Bulgarian population might reflect population-specific mutations, for example the 185delAG mutation amongst Ashkenazi Jews.

The five SNPs were used to construct haplotypes in the population that I studied. Two programs were used; the expectation-maximisation algorithm EMHAPERE (Excoffier and Slatkin 1995) was applied to predict haplotypes from genotype data on 176 individual from the same ethnic group. Fourteen haplotypes were estimated by the Arlequin method for the total population (table 3-9). The first five haplotypes were considered, which have a frequency of more than 8%. There are a few haplotype studies in the literature; each haplotype study used different SNPs, which make comparison difficult. However, partial comparison to the study by Van Orsouw is possible (van Orsouw et al. 1999) table 4-1. In my study, haplotype 3 corresponds to Van Orsouw B+C and my haplotype 1 corresponds to Van Orsouw A. However, my haplotype 2, with only one cSNP of 3667A>G, has no correspondent in the Van Orsouw study. All subjects were included in my analysis if there was a single common mutation always on the same haplotype background (as is the situation for 185delAG in

Ashkenazi Jews), then that haplotype would be found in excess in my case set relative to controls. Similarly, a haplotype itself with a modest effect would be more frequent in cases than controls. My project lacked sufficient case (and control) numbers for this goal, but the analysis were undertaken for sake of completeness.

	Exon 11						exon 13	exon 16	
Type	2196	2201	2430	2731	3232	3667	4427	4956	%
A	G	C	T	C	A	A	T	A	63
B	G	T	C	T	G	G	C	G	28
C	A	T	C	T	G	G	C	G	7
D	G	C	T	T	A	A	T	A	2

Table 17.4. Haplotypes analysis of the common SNPs in *BRCA1* gene. Taken from Van Orsouw et al., 1999.

The common haplotypes observed might relate to each other in the following way, Figure 46.4.

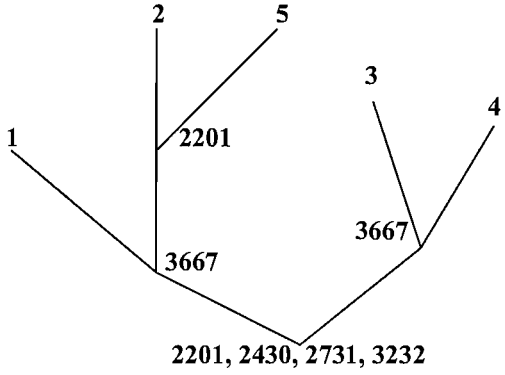


Figure 45.5 Haplotype tree illustrates the association of the five haplotypes with each other.

Haplotypes 3 and 4 share similar SNPs with only one SNP different, 3667A>G. In the similar scenario is true for haplotype 2 and 5, only 2201 SNP is found in haplotype 5. Furthermore, a 3667 SNP was the only difference in the two haplotypes, 1 and 2. Sharing of the 3667 SNP in most of the haplotypes suggests this is a neutral common SNP found in breast cancer and control samples (Li et al. 2002).

I have compared the haplotype-prediction algorithm that was produced by the Arlequin method by using another statistical method called PHASE method developed by Stephens M. in 2001 (Stephens, Smith, and Donnelly 2001). There were minor discrepancies in the results of this comparison. These differences refer to the specification of the two methods. Stephens M. and his group (Stephens, Smith, and Donnelly 2001) stated that the PHASE model is more accurate over the EM algorithm (Arlequin method) as EM will not generate accurate estimations with the large number of samples and number of linked SNPs employed to calculate haplotype

frequencies, moreover, PHASE will perform well when there is clustering in the true haplotype configuration. Although the sample size that I used for this analysis is not large enough to have good power, there was reasonable concordance of the haplotype frequency estimates by different methods. Unsurprisingly, there is no uniquely prevalent haplotype present in the Wessex region high risk patients. A large set of suitably mutated controls would be needed to search for subtle haplotypic differences indicating either risk modifier haplotypes or associated rare founder mutations.

4.9 Cost-effectiveness

Melt-MADGE was designed to be suitable for large-scale analysis of most human genes in population-based studies. Therefore, a very small number of samples planned for full gene scanning might be more economical and faster if the analysis were performed by direct DNA sequencing, especially for the long coding sequence genes, such as RB or *BRCA1* that required longer procedures.

The acceleration of identification of new genes causing disease in the human genome has increased the need for methods with high throughput, simplicity and low cost for screening large genes with many mutations scattered through the gene in large numbers of cases. None of the current methods for mutation detection is ideal to produce high throughput data on many genes from thousands of individuals either in a population-based study for association studies or in a diagnostic laboratory to examine the new accrued backlog of patient samples. The method of choice for screening large populations does not rely on sensitivity only, but should also be at high throughput and cost-effectiveness. Screening programs for testing genetic predisposition to cancer would have to meet the conventional criteria for population screening. Certain criteria were initially proposed by Wilson and Jungner in 1968 (Wilson JMG. and Jungner G 1968) for evaluating screening tests.

The Wilson-Jungner criteria for validation of a screening programme;

1. The condition being screened for should be an important health problem
2. The natural history of the condition should be well understood
3. There should be a detectable early stage
4. Treatment at an early stage should be of more benefit than at a later stage
5. A suitable test should be available to detect early stage disease
6. The test should be acceptable
7. Intervals for repeating the test should be determined

8. Adequate health service provision should be made for the extra clinical workload resulting from screening
9. The risks, both physical and psychological, should be less than the benefits
10. The costs should be balanced against the benefit

Further criteria based on Wilson and Jungner's criteria have been defined for evaluation of new technology for population screening. In the context of cost-effectiveness, the majority of the measurements necessitated the balance between the benefit of the screening test and the screening cost (Goel 2001).

The cost for screening large populations by melt-MADGE was estimated for running the program for one year. Our approach is certainly much more cost- effectiveness than any other mutation detection method with a more rapid rate of sample throughput. For example, the cost of PCR for 10,000 amplicons will be £400, and to run these amplicons by melt-MADGE will cost £50 and will take one week. The cost of the DNA analysis along with technical considerations will determine the choice of the applied method. The cost of direct DNA sequencing for mutation detection will be greater than any electrophoresis-based method and will depend in part on the number of mutations in the gene. However, it may be favourable for some diseases that are caused by a limited number of mutations, single gene disorders, such as cystic fibrosis, where the presence of a genetic mutation results in disease (Wang and Freedman 2002). However, in some diseases such as breast cancer, disease susceptibility may be due to one of many mutations in the gene, *BRCA1* and *BRCA2*, or in some cases, more than one gene may be involved. The technical efficiency will determine the type of technology that will be most effective to detect the genetic abnormality at the lowest cost in large populations. However, Vijg and Orsouw (Vijg and van Orsouw 1999) reviewed the cost for screening the full coding sequence of the *BRCA1* gene and others such as RB1 and TP53 genes by TDGS method, the total cost is $66.40 , $39.31 and $45.40 per sample, respectively. The number screened per week is 60 samples. TDGS is the method of choice compared with DNA sequencing or dHPLC, and it has similar sensitivity as DGGE for mutation detection, between 80 and 100%.

Melt-MADGE is expected to reduce the cost of scanning population samples for rare and common mutations in large populations (Day INM et al. 2003). However, the design features of melt-MADGE are such that its final cost is likely to be much less than TDGS (Bounpheng et al. 2003; Vijg and van Orsouw 1999) or any other DNA molecular technique (at the present) for testing a large population for mutations (Table 16.3).

FUTURE WORK

The establishment of the melt-MADGE assay for the coding region of *BRCA1* and completing the screening of 100 familial breast cancer patients enhanced our potential to expand our screening to cover large sample sizes and many genes that might predispose to breast cancer. This development will enable us to carry out the screening for all known and unknown variants. This will be plausible for many genes in population studies. Detection of a new polymorphisms and/or rare variant in the *BRCA1* gene would probably define the correlation between the genotype and phenotype. However, determining the risk of specific mutations as well as mutations that have high or low risk in women who developed breast cancer require a large sample size, probably thousands of samples for association studies.

In the future, studies to extend this research will involve the following:

- In this project, I have compared cost of melt-MADGE method with the standard mutation detection methods such as dHPLC and SSCP. In future, a blinded comparison of speed and sensitivity between melt-MADGE and standard methods would clearly reinforce the value of the melt-MADGE assay.

- In future, we need to determine the reproducibility of the technique in other laboratories.

- High throughput analysis and optimisation of the technique for the *BRCA1* gene could be applied to the large backlog of genetic samples submitted for BRCA1 and BRCA2 analysis to UK diagnostic laboratories.

- Population based association studies could use this technique to provide full-scale scanning of each amplimer in the *BRCA1* gene (often several thousands cases and controls).

- As previously mentioned many molecules are co-localising together at sites of DNA damage. Therefore, developing this technique to screen genes such as, *BRCA2*, Rad51 and BARD1 in large sets of breast cancer patients may help develop a better understanding of the association between different types of mutations in different molecules and their interaction in causing susceptibility to breast cancer.

CHAPTER FIVE

APPENDIXES

A) Internal primers sequence for melt-MADGE

Melt-MADGE internal primers for all coding regions of *BRCA1* gene, GC-rich sequence "GC-clamp" were attached with primer sequence. The primers without GC-clamp that used for sequencing in all the amplicons are in **bold letters.**

Forward	*BRCA1*x25'	5'-CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCGTATGTTTTTCTAATGTGT-3
Reverse	*BRCA1*x23'	5'-ACTCTTGTGCTGACTTAC-3'
Forward	*BRCA1*x35'	5`-CGCGGATTTATTTTCTTTTTCTCCC-3`
Reverse	*BRCA1*x33'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCCAACCAACTTTTGATAACTA-3`
Forward	*BRCA1*x55'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCAAAGGAAGTAAATTAAATTG-3`
Reverse	*BRCA1*x53'	5`-GCCGCGCCCCAACTCTTTTATAAATTTTTC-3`
Forward	*BRCA1*x65'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCATCACTTGCTGAGTGTGTTT-3`
Reverse	*BRCA1*x63'	5'-CGCGGAGGTCTTATCACCACGTCAT-3'
Forward	*BRCA1*x75'	5`-CGCGGCGGAGAACTGCAAACATAATGTTTT-3`
Reverse	*BRCA1*x73'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCACAGAGCAAGACTCCATCTC-3`
Forward	*BRCA1*x85'	5'-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCAATAATTGCTTGACTGT-3
Reverse	*BRCA1*x83'	5'-CGCGGAAAAAACCTGACCCT-3'
Forward	*BRCA1*x95'	5`-CGCGGCGGAGAGAAAACTTTTATTGATTTA-3`
Reverse	*BRCA1*x93'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCTATTTTTAAAAAGAGAGAAA-3`
Forward	*BRCA1*x105'	5`-CGCGGCGCCGGTGTTTCTTATTAGGACTCT-3`
Reverse	*BRCA1*x103'	5`-GCGCCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCCGTGCCTGTTAAGTTGGCAAA-3`
Forward	*BRCA1*x11A5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCTTGTTATTTTTGTATATTTT-3`
Reverse	*BRCA1*x11A3'	5`-CGCGGAGTGAGTAATAAACTGCTGT-3`
Forward	*BRCA1*x11B5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCCAGCTCATTACAGCATGAGA-3`
Reverse	*BRCA1*x11B3'	5`-CGCGGTGAATGCTGCTATTTAGTGT-3`
Forward	*BRCA1*x11C5'	5`-CGCGGCTGAAGATGTTCCTTGGATA-3`
Reverse	*BRCA1*x11C3'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCCTCTTTCACTTTTACATATT-3`
Forward	*BRCA1*x11D5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCAGTGATCCTCATGAGGCTTT-3`
Reverse	*BRCA1*x11D3'	5`-CGCGGATGAAGGCCTGATGTAGGTC-3`
Forward	*BRCA1*x11E5'	5`-CGCGGTAAATTAAAGCGTAAAAGGA-3`
Reverse	*BRCA1*x11E3'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCAAAGCAGATTCTTTTTCGAG-3`
Forward	*BRCA1*x11F5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCATCCTAACCCAATAGAATCA-3`
Reverse	*BRCA1*x11F3'	5`-CGCGGTGCTAGAACAACTATCAATT-3`
Forward	*BRCA1*x11G5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCCCTAATTGTACTGAATTGCA-3`
Reverse	*BRCA1*x11G3'	5`-CGCGGTGAACACTTAGTAAAAGAAC-3`
Forward	*BRCA1*x11H5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCGAAGTTAACAAATGCACCTG-3`
Reverse	*BRCA1*x11H3'	5`-CGCGGTGAGTGCCATAATCAGTACC-3`
Forward	*BRCA1*x11I5'	5`-CGCGGTGAAAGATCTGTAGAGAGTA-3`
Reverse	*BRCA1*x11I3'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCTTTCTATGCTTGTTTCCCGA-3`
Forward	*BRCA1*x11j5'	5`-CGCGGGGACATGAAGTTAACCACAG-3`
Reverse	*BRCA1*x11j3'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCCCCATTCTTTCCTTGATTTTCTT-3`
Forward	*BRCA1*x11k5'	5'-CGCGGGGCCAGTCCAAAAGTCACTTTTGA-3'

Reverse	*BRCA1*x11k3'	5`-GGGCCCGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCCTTGATGGGAAAAAGTGGTG-3'
Forward	*BRCA1*x11L5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCCTCCAAATAAACATGGACTT-3`
Reverse	*BRCA1*x11L3'	5`-GGCGCGGCTTCATTAATATTGCTTGAG-3`
Forward	*BRCA1*x11M5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCGCCGTAATAACATTAGAGAA-3`
Reverse	*BRCA1*x11M3'	5`-GGCGCGGATATTCTTGCTTTTTTATTT-3`
Forward	*BRCA1*x11N5'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCAAGTAATTGTAAGCATCCTG-3`
Reverse	*BRCA1*x11N3'	5`-CGCGGCTTTTGCTAAAAACAGCAGA-3`
Forward	*BRCA1*x11O5'	5`-CGCGTCCCGAAAATGACATTAAGGAAAGT-3`
Reverse	*BRCA1*x11O3'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCCGGTGCTATGCCTAGTAGAC-3`
Forward	*BRCA1*x11P5'	5`-CGCGGGTAAACAATATACCTTCTCA-3`
Reverse	*BRCA1*x11P3'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCCTGGGTGTTTGTATTTGCAG-3`
Forward	*BRCA1*x11Q5'	5`-CGCGGCAGTGAATTGGAAGACTTGA-3`
Reverse	*BRCA1*x11Q3'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCACACAAAAACCTGGTTCCAA-3`
Forward	*BRCA1*x125'	5`-CGCGGGGGCCGGGTCTGCTTTTACATCTGAAC-3'
Reverse	*BRCA1*x123'	5'-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCTTGTGCCATTAATTCAAAGAGATGA-3'
Forward	*BRCA1*x135'	5`-CGCGGGCCATTTATCGTTTTTGAAG-3`
Reverse	*BRCA1*x133'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCTGGAGCTAGGTCCTTACTCT-3`
Forward	*BRCA1*x145'	5'-CGCGGGCCATCAGAACAAAGCAGTAAAG-3'
Reverse	*BRCA1*x143'	5'-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCGAAAAAAATTAACAATCAGA-3'
Forward	*BRCA1*x155'	5'-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCTTTGTCCTTTCACAATTGGT-3'
Reverse	*BRCA1*x153'	5'-CGGCCGGCCGCGGTGTAGGATTCAGAGTAAAAT-3'
Forward	*BRCA1*x165'	5-`CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCTCAACATTCATCGTTGTGTA-3`
Reverse	*BRCA1*x163'	5`-CGCGGCGGAGCGAGGCCCGCTTAGGGAGATACATATGGAT-3`
Forward	*BRCA1*x175'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCAATAACTAGTATTCTGAGCT-3`
Reverse	*BRCA1*x173'	5`-CGCGGTCATGTGGTTTTATGCAGCA-3`
Forward	*BRCA1*x185'	5`-CGCGGGGGAGTGTAAAAAACTGAGG-3`
Reverse	*BRCA1*x183'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCTTTAGTATTACAATTAAAGA-3`
Forward	*BRCA1*x195'	5`-CGCGGCTTTTCTATGATCTCTTTAG-3`
Reverse	*BRCA1*x193'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCAGGAAAGTGGTGCATTGATG-3`
Forward	*BRCA1*x205'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCTCCACTTCCATTGAAGGAAG-3`
Reverse	*BRCA1*x203'	5`-CGCGGTTACAAAATGAAGCGGCCCA-3`
Forward	*BRCA1*x215'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCTCTTTTATAACTAGATTTTC-3`
Reverse	*BRCA1*x213'	5`-CGCGGCGGAGCCCATCGTGGGATCTTGCTT-3`
Forward	*BRCA1*x225'	5'-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCAGAGGGCCTGGGTTAAGTAT-3
Reverse	*BRCA1*x223'	5'-GACAGGTGCCAGTCTTGCTC-3'
Forward	*BRCA1*x235'	5`-CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCGACAGTTCCAGTAGTCCTAC-3`
Reverse	*BRCA1*x233'	5'-CATATAGCACAGGTACATG-3'
Forward	*BRCA1*x24A5'	5'-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCTGGAGTCGATTGATTAGA-3'
Reverse	*BRCA1*x24A3'	5'-CGCGGTCCCAGGGCCTGGAAAGGCC-3
Forward	*BRCA1*x24G5'	5`-CGCGGTTTATAGCTGTTGGAAGGAC-3`
Reverse	*BRCA1*x24G3'	5`-CGCGGCGGAGCGAGGCCCGCGGGCCCGCCCGCCGCGCGCCCCAAAGGCTCTGAGAAAGTCGG-3`

B) Primers for ARMS assay

Primers design for ARMS assay of the fife common polymorphisms sites in the exon 11 of the *BRCA1* gene.

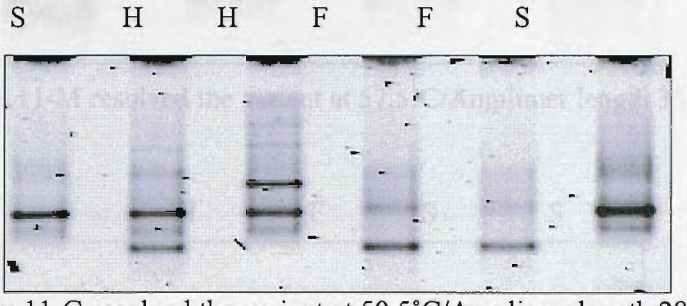
Primer ID	Oligo's sequence
2201FC	5'-CAAGTAAAAGACATGACCGC-3'
2201FT	5'-CAAGTAAAAGACATGACCGT-3'
2201U/C	5'-AAGAAAAAAAAGTACAACCAAATGC-3'
2201D/C	5'-GGAAGGCTAGGATTGACAAAT-3'
2430FC	5'-GATCTGTAGAGAGTAGCAGTATTTAAC-3'
2430FT	5'-GATCTGTAGAGAGTAGCAGTATTTAAT-3'
2430U/C	5'-AATCCTAGCCTTCCAAGAGAAGAA-3'
2430D/C	5'-AACAACCATGAATTAGTCCCTTGG-3'
2731FT	5'-GCGCCAGTCATTTGCGCT-3'
2731FC	5'-GCGCCAGTCATTTGCGCC-3'
2731U/C	5'-GGCTTTAAGTATCCATTGGGACATGAAGTTAA-3'
2731D/C	5'-TGTACAGGCTTGATATTAGACTCATTCTTTCC-3'
3232FA	5'-CATTAGAGAAAATGTTTTTAACGA-3'
3232FG	5'-CATTAGAGAAAATGTTTTTAACG-3'
3232U/C	5'-AAGAGAAATGGGAAATGAGA-3'
3232D/C	5'-CCTAATCTAAGCATAGCATTC-3'
3667FA	5'-CTGTTTTTAGCAAAAGCGTCCATAA-3'
3667FG	5'-CTGTTTTTAGCAAAAGCGTCCATAG-3'
3667U/C	5'-GATGGTGAAATAAAGGAAGATACTAGTTTTG-3'
3667D/C	5'-AGCTCTTCATCCTCACTAGATAAGTT-3'

C) Primers for amplifying long sequence

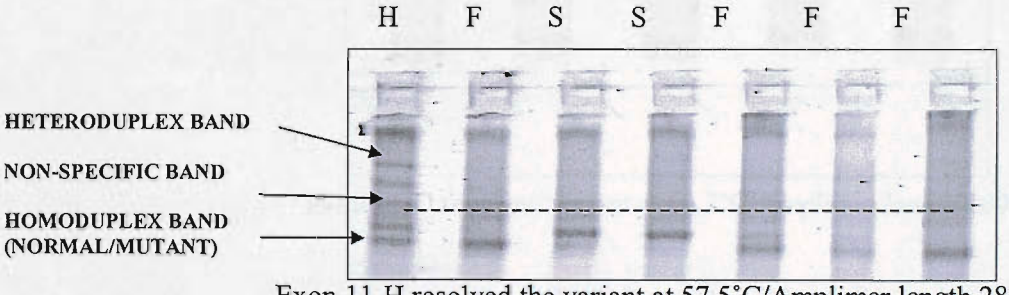
Primer ID.	Covered EXON	Forward	Reverse	Length
*BRCA1*L1	1a, 1b, 2	5-GACTCTCCAACTCTCCGGCGCTTTTC-3	5-TTGCCATTTCAAGTGATGGAGCTTG-3	3kbp
*BRCA1*L2	X3	5-GGGGTAGGGGTGGATATGGGTGAAACAG-3	5-GGTGTTTCCTGGGTTATGAAGGACAAAA-3	700bp
*BRCA1*LN 10-11	10 and 11	5-GAGAGCAGCTTTCACTAACTAAATAAGATTGGTCAGCTTTCTGT-3	5-TCAAGTTTAAGAAGCAGTTCCTTTAACTATACTTGGAAATTTGT-3	4.8kbp
*BRCA1*LN 12-13	12 and 13	5-GCTAGGACGTCATCTTTGACTGAATGAGCTTTA-3	5-GCGATAATTACCCATGTGCTGAGCAAGGATCA-3	9.0kbp
BRCA1 14-16	14, 14 and 16	5-TTTTTCAGCCTTGTCTCAGCTGGGTGT-3	5-CCATTTCTGGCATTAAGGACCCAAGGT-3	5.8kbp
BRCA1L 18-20	18, 19 and 20	5-TCTTAACTTCATATCAGCCTCCCCTAGACTTCCAAATATCC-3	5-CATCTCTGCAAAGGGGAGTGGAATACAGAGTG-3	7.2kbp
BRCA1L 21-24	21, 22, 23 and 24	5-CACTCTTCCATCCCAACCACATAAATAAGTATTGTCTC-3	5-GCATAGCCAGAAGTCCTTTTCAGGCTGATGTAC-3	11.4kbp

D) Examples of amplicon run at constant temperature

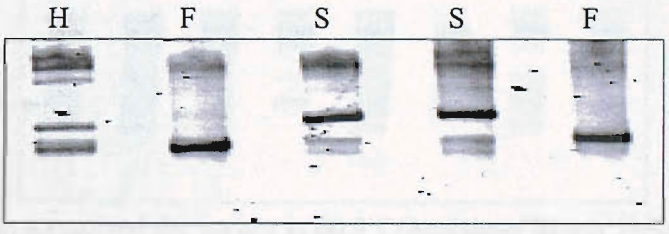
Examples of heterozygous polymorphisms and mutations that were detected in *BRCA1* gene by Melt-MADGE method at constant temperature. These Figures represented the amplicons in table 3-1, results section.



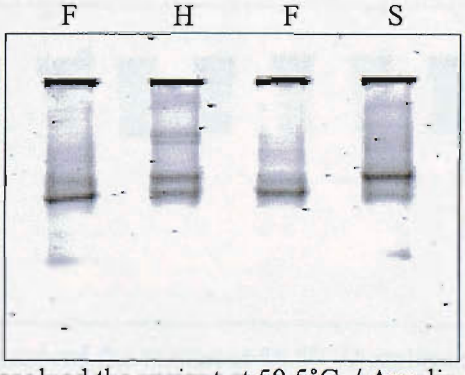
Exon 11-G resolved the variant at 59.5°C/Amplimer length 285bp.



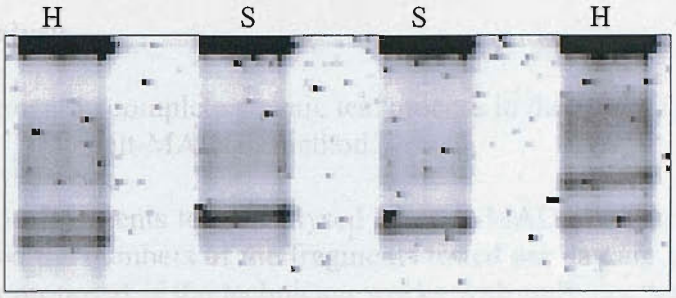
Exon 11-H resolved the variant at 57.5°C/Amplimer length 285bp.



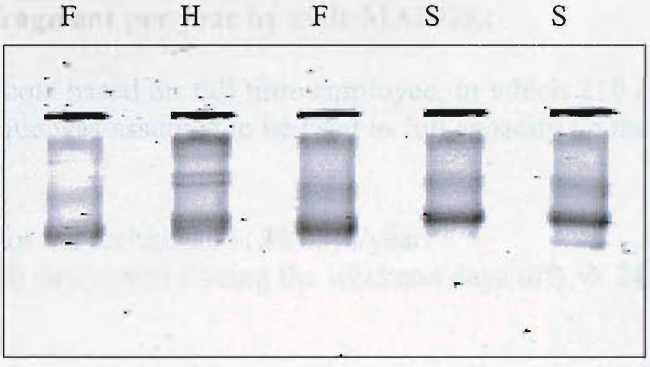
Exon 11-L resolved the variant at 57.5°C / Amplimer length 287bp.



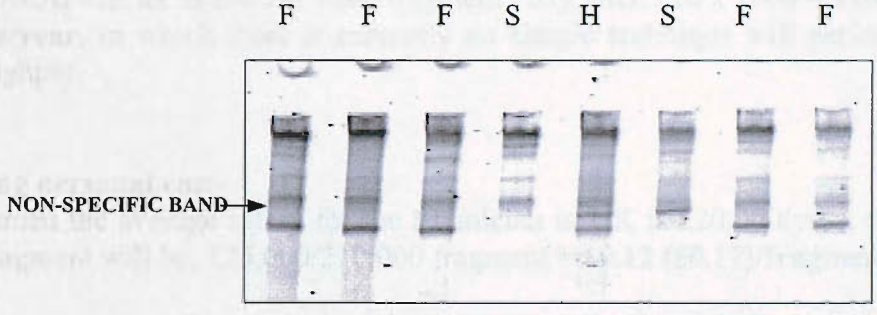
Exon 11-I resolved the variant at 59.5°C / Amplimer length 305bp.



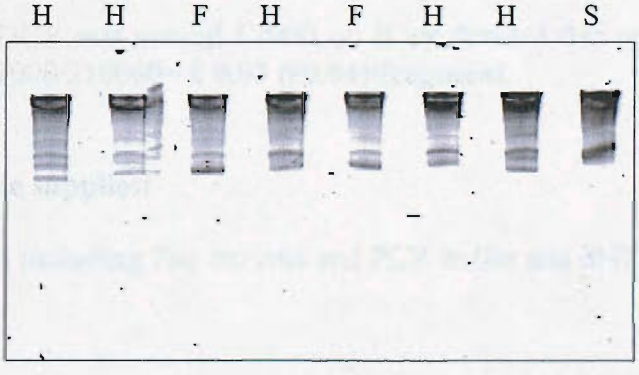
Exon 11-M resolved the variant at 57.5°C/Amplimer length 307bp



Exon 11-O resolved variant at 60.7°C/Amplimer length 289bp.



Exon 8 resolved the variant at 60.9°C/Amplimer length 338bp.



Exon 18 resolved the variant at 58.5°C/Amplimer length 353bp.

S = Slow Homoduplex amplicon.
F = Fast Homoduplex amplicon.
H= Heteroduplex amplicon.

E) Cost-effectiveness analysis

The estimate cost per fragment in complete genetic test process in the *BRCA1* gene by melt-MADGE method

BRCA1 gene divided into 54 fragments to be analysed by melt-MADGE. Therefore, in the melt-MADGE method the numbers of the fragments tested per day are 10 000 amplicons. This number is proposed if the technician works with only one tank per day.

Calculating the analysed fragment per year by melt-MADGE:

Calculation of the personal cost based on full time employee, in which 210 days per year. In addition, the technique was assumed to be used to full capacity on the annual basis.

I assumed the annual leave for the technician is **30** days/year.
Working day per month →20 day/month (taking the weekend days off) → 240 / year (20X12)

240-**30 (annual leave)** = 210 full working days in UK for a technician.

With melt-MADGE let us assume 1000 fragment / day, then 210 x 1000 = **210,000 fragments/year**, in which there is currently no simple technique will perform this high throughput.

Calculating personal cost:

If we assumed the average salary for the technician in UK is £20,000/year, then the cost per fragment will be; £25,000/210,000 fragment = **£0.12 (€0.17)/fragment**

Calculating the equipment cost:

The full system for melt-MADGE cost around £7000, so if we divided this amount on the total fragment will be; 7000/210000= **£ 0.03 (€0.04)/fragment**.

Calculating PCR Consumable supplies:

The total cost of PCR reagents, including *Taq* enzyme and PCR buffer and dNTPs, in 96-plate is estimated at £3.84.

Estimating the cost of the reagents that used for melt-MADGE electrophoresis.

1. **The cost for preparing the gel:**

- Urea = £71.71/ 5kg
  £0.014/1g

I generally used 4M Urea in which equal to 12g of one gel
The cost for one gel = 12x0.014 = £ 0.17/ (96-well)
The cost per well = £0.002

- PAG = £31.42/ 1000ml

It needs 10ml per gel for preparing 6%PAG
£0.31/10ml
The cost per well = **£0.003**

2. **The cost for preparing T.A.E. buffer:**

200ml of 10xTAE prepared and diluted to 2ltrs with dH_2O for electrophoresis.

- Na-EDTA > £111.60/ 2.5kg
  £0.045/1g

For 200ml, It needs 0.74g → **0.74x 0.045= £0.03**

- Tris, =>£16.0/kg---→£0.016/g

9.68g /200ml--------→ 0.016x9.68 = £0.15

- Glacial acidic acid, => £6.4/500ml

2.28ml of A.A. / 200ml-------→ 2.28 x £0.013 = £0.03

Summery;

Urea	=£0.17/gel (96-well)----------→£0.002/ well	
PAG	=£0.31/gel (96-well)----------→£0.003/well	
Na-EDTA	=£0.03/ 200ml----------→£0.0003/well	
Tris	=£0.15/ 200ml----------→£0.002 /well	
Glacial A.A.	=£0.03/ 200ml----------→£0.0003/ well	
Total	= £0.69/one gel	£0.73/96-amplicon
	= £6.9 /10 gel	£7.30/960-amplicon

F) Annotated BRCA1 sequence.

Map of the *BRCA1* gene showed the construction of the coding and non-coding sequence based on GenBank accession Number L87733 for full *BRCA1* gene sequence. The alignment of translated cDNA to Amino Acids was based on the GenBank accession U14680. The promoter and 5`UTR sequence were based on GenBank accession Number U37574. The arrows are represent the PCR primers position and marked as F (Forward) and R (Reverse) letter combined with the exon or segment name. Red colour represents the exon sequence and dark blue represents the 5`UTR and 3`UTR.

f*; denoted for the frequency as reported by Markoff A. et al.,1998
f; denoted for frequency observed in my study**

Promoter sequence Started at 1763●

●

GCG GGC CGG CAG TGC TGG GGG ACC CGG CGC ACC CTC TGC AG[C TGC TGG CCC GGG
TGC TAG GCC CCT GAC TGC CCG GGG CCG GGG GTG CGG GGC CCG CTG AGC CCG
CGC CCA CCT GGA ACT CGC GCT GGC TGG CGA GCG CTG CGC GCA GCC CCA GTT CCC
ACA CCC GCC TCT CCC TCC ACA CTT CCC CGC AAG CAG AGG GAG CCG GCT CTG GCT
TCG GCC AGC CCA GAG AGG GGC CCC CAC AGC GCA GTG GCG GGC TGA AGG GCT CCT
CCA GCA CGG CCA GAA TGG ACG CCA AGG CCG AGG AGG CGC CGA GAG CGA GCG AGG
GCT GCT AGC ACG TTG TCA CCT CGC ATT CTG AAC CAC AGA CTC TCC AAC TCT CCG
GCG CTT TTC GCC CAC TCG GTC CCT CAG AAC ACG AAG GGC TCT CTC ATC CTG TCA
CTA AAA CGA TTA GCT GTC CGG AGA CAC GGA AAA AGT CGC CCC TCT TCT TTG CAG
GAT TCC TCC CTT GAA CTT CTC CAA ACC CTC TTA GTG TGA CGT GAC CCC ACC CCT
AGC TAA CCC AGG CTG CTT CCT TAC CAG CTT CCC GCC CCC TGG GGA GGC GGC AAT
GCA AAG ACC GTC CGC TGC CAG CTC TGC CGC TAT CTC TGT GGG GTG AAT CTA ACA
TGG CGG ACA AAG ACA GTA ACT AGT CCC GTT TCT CCG CGT TTT CGC CAA GAA GAT
TGG CTC TTA CCA CTT GTC CCT CAA AAC GAC CAC CCC ATT GAC TGG TGG CGA TTG
CGT CGA CGG AGA CGG GGC AAA AGC AAG CTG AAC CCG AAA AAT AAC AAA CAC TGG
GGC TGA GGG GTG GAA CTA CGA GTG CGC AGA CAT GGG CCA GAG CGC ATT TCC CCT
GCC CCA GGC AAA TTC GGC GCT CAC TGC GTC CCC GCA GGC CAC TGA CCT TAC AAG
ACT ACT TGC CCC AGA CTC CTG GGG CTG GAT GGG AAT TGT AGT CTC CCT AAA GAG
TTG TAC GTA TCT TTT TAA GGC CTA GTT TCT GCT TTC AAA ATA CGA AAA CAT AAC
ACT CCA GTC CAT AAC TGT TGA CAA GTA CAA GCG CGC ACA GGT CTC CAA TCT ATC
CAC TGG ATT TCC GTG AGA ATT GTG CCC GCT CTG GTA TTG GAT GTT CCT CTC CAT
AAG ACT ACA GTT TCT AAG GAA CAC TGT GGC GAA GAC CTT TCA TTC CGC AAC GCA
TGC TGG AAA TAA TTA TTT CCC TCC ACC CCC CCA ACA ATC CTT ATT ACT TAT ATT
TAC CGA AAC TGG AGA CCT CCA TTA GGG CGG AAA GAG TGG GGG ATT GGG ACC TCT
TCT TAC GAC TGC TTT GGA CAA TAG GTA GCG ATT CTG ACC TTC GTA CAG CAA TTA
CTG TGA TGC AAT AAG CCG CAA CTG GAA GAG TAG AGG CTA GAG GGC AGG CAC TTT
ATG GCA AAC TCA GGT AGA ATT CTT CCT CTT CCG TCT CTT TCC TTT TAC GTC ATC
CGG GGG CAG ACT GGG TGG CCA ATC CAG AGC CCC GAG AGA CGC TTG GCT CTT TCT
GTC CCT CCC ATC CTC TGA TTG TAC CTT GAT TTC GTA TTC TGA GAG GCT GCT GCT
TAG CGG TAG CCC CTT GGT TTC CGT GGC AAC GGA AAA **GCG CGG GAA TTA CAG ATA**

F1a

● ●

AAT *T]* ● End of the promoter sequence is at Nucleotide No.3343 bp.

(5`UTR region) Exon 1a

*(AA AAC TGC GAC TGC GCG GCG TGA GCT CGC TGA GAC TTC CTG GAC GGG GGA CAG GCT GTG GGG TTT CTC AGA TAA CTG GGC CCC TGC GCT CAG GAG GCC TTC ACC CTC TGC TCT GGG TAA AG)*G TAG TAG AGT CCC GGG AAA GGG ACA GGG GGC CCA AGT GAT
GCT CTG GGG TAC TGG CGT GGG AGA GTG GAT TTC CGA AGC **TGA CAG ATG GGT ATT**

R1a (at 3561bp)

CTT TGA CGG GGG GTA GGG GCG GAA CCT GAG AGG CGT AAG GCG

(5`UTR region)

TTG TGA ACC CTG GGG AGG GG*[G GCA GTT TGT AGG TCG CGA GGG AAG CGC TGA GGA*
TCA GGA AGG GGG CAC TGA GTG TCC GTG GGG GAA TCC TCG TGA TAG GAA CTG GAA
TAT GCC TTG AGG GGG ACA CTA TGT CTT TAA AAA CGT CGG CTG GTC ATG AGG TCA
GGA GTT CCA GAC CAG CCT GAC CAA CGT GGT GAA ACT CCG TCT CTA CTA AAA ATA
CAA AAA TTA GCC GGG CGT GGT GCC GCT CCA GCT ACT CAG GAG GCT GAG GCA GGA
GAA TCG CTA GAA CCC GGG AGG CGG AGG TTG CAG TGA GCC GAG ATC GCG CCA TTG
CAC TCC AGC CTG GGC GAC AGA GCG AGA CTG TCT CAA AAC AAA ACA AAA CAA AAC
*AAA ACA AAA AAC ACC GGC TG]*G TAT GTA TGA GAG GAT GGG ACC TTG TGG AAG AAG
AGG TGC CAG GAA TAT GTC TGG GAA GGG GAG GAG ACA GGA TTT TGT GGG AGG GAG
AAC TTA AGA ACT GGA TCC ATT TGC GCC ATT GAG AAA GCG CAA GAG GGA AGT AGA
GGA GCG TCA GTA GTA ACA GAT GCT GCC GGC AGG GAT GTG CTT GAG GAG GAT CCA
GAG ATG AGA GCA GGT CAC TGG GAA AGG TTA GGG GCG GGG AGG CCT TGA TTG GTG
TTG GTT TGG TCG TTG TTG ATT TTG GTT TTA TGC AAG AAA AAG AAA ACA ACC AGA
AAC ATT GGA GAA AGC TAA GGC TAC CAC CAC CTA CCC GGT CAG TCA CTC CTC TGT
AGC TTT CTC TTT CTT GGA GAA AGG AAA AGA CCC AAG GGG TTG GCA GCA ATA TGT
GAA AAA ATT CAG AAT TTA TGT TGT CTA ATT ACA AAA AGC AAC TTC TAG AAT CTT
TAA AAA TAA AGG ACG TTG TCA TTA GTT CTT TGG TTT GTA TTA TTC TAA AAC CTT
CCA AAT CTT AAA TTT ACT TTA TTT TAA AAT GAT AAA ATG AAG TTG TCA TTT TAT
AAA CCT TTT AAA AAG ATA TAT ATA **TAT GTT TTT CTA ATG TGT** TAA A

F2

Exon #2

(5`UTR region) *

[GT TCA TTG GAA CAG AAA GAA] ATG GAT TTA TCT GCT CTT CGC GTT

M D L S A L R V
1 2 3 4 5 6 7 8

* **Coding gene started at Nucleotide #120 is Initiation codon**

185Agdel↓

GAA GAA GTA CAA AAT GTC ATT AAT GCT ATG CAG AAA ATC TTA G**AG** TGT CCC ATC

E E V Q N V I N A M Q K I L E C P I
9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

TG]*G* ***TAA GTC AGC ACA AGA GTG*** TAT TAA TTT GGG ATT CCT ATG ATT ATC TCC TAT

R2

GCA AAT GAA CAG AAT TGA CCT TAC ATA CTA GGG AAG AAA AGA CAT GTC TAG TAA
GAT TAG GCT ATT GTA ATT GCT TTC TCA GTT CCT GAC ACA GCA GAC ATT TAA TAA
ATA TTG AAC GAA CTT GAG GCC TTA TGT TGA CTC AGT CAT AAC AGC TCA AAG

(Exon # 3)

TTG AAC TTA TTC ACT AAG AAT AGC TTT ATT TTT AAA TAA ATT ATT GAG CCT C**AT**
TTA TTT TCT TTT TCT CCC CCC CTA CCC TGC T*AG* [T CTG GAG TTG ATC AAG GAA

F3c

C L E L I K E
27 28 29 30 31 32 33

CCT GTC TCC ACA AAG TGT GAC CAC ATA TTT TGC **AA]** ***GT***A AGT TTG AAT GTG TTA

P V S T K C D H I F C L
34 35 36 37 38 39 40 41 42 43 44 45

TGT GGC TCC ATT ATT AGC TTT TGT TTT TGT CCT TCA TAA CCC AGG AAA CAC CTA
ACT TTA TAG AAG CTT TAC TTT CTT CAA TTA AGT GAG AAC GAA AAA TCC AAC TCC
ATT TCA TTC TTT CTC AGA GAG TAT **ATA GTT ATC AAA AGT TGG TTG** TAA TCA TAG

R3 (at 13177bp)

TTC CTG GTA AAG TTT TGA CAT ATA TTA TCT TTT TTT TTT TTT TTG AGA CAA AGT
CTC GCT CTG TCG CCC AGG CTG GAG TGC AGT GGC ATG ATC TTG GCT CAC TGC AAC
CTC CGC CCC CCG AGT TCA AGC GAT TCT TCT ACC TCA GCC TCC CAG GTA TAT GCC
(21924)

(Exon #5)

AGT ATT CTT TCT ACA **AAA GGA AGT AAA TTA AAT TGT** 22176 TCT TTC TTT CTT

F5 at 22154

TAT AAT TTA TAG [A TTT TGC ATG CTG AAA CTT CTC AAC CAG AAG AAA GGG CCT

F M M K L L N Q K K G P S

46 47 48 49 50 51 52 53 54 55 56 57 58

TCA CAG TGT CCT TTA TGT AAG AAT GAT ATA ACC AAA AG] GTA TAT AAT TTG GTA

Q C P K L C K N D I T K R

59 60 61 62 63 64 65 66 67 68 69 70 71

ATG ATG CTA GGT TGG AAG CAA CCA CAG TAG GAA AAA GTA GAA ATT ATT TAA TAA

CAT AGC GTT CCT ATA AAA CCA TTC ATC A**GA AAA ATT TAT AAA AGA GTT** TTT AGC

R5 (at 22395 bp)

(Exon #6)

AT CAC TTG CTG AGT GTG TTT CTC AAA CAA TTT AAT TTC AG[G AGC CTA CAA GAA

F6 (at 23754)

S L Q E

72 73 74 75

AGT ACG AGA TTT AGT CAA CTT GTT GAA GAG CTA TTG AAA ATC ATT TGT GCT TTT

S T R F S Q L V E E L L K I I C A F

76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93

CAG CTT GAC ACA GGT TTG GAG T]GT AAG TGT TGA ATA TCC CAA GAA TGC AAC TCA

Q L D T G L E

94 95 96 97 98 99 100

AGT GCT GTC CAT GAA AAC TCA GGA AGT TTG CAC AAT TAC TTT CT**A TGA CGT GGT**

R6 (at 23996 bp)

GAT AAG ACC TTT TAG TCT AGG TTA ATT TTA GTT CTG TAT CTG TAA TCT ATT ATC

R6

GTA AGG TGC ACA TTT TTC ACA TCT TAA CAT CTC TGA AAT TGG GAA CAT TTT ACT

ATT GAG GGT

GTG TCA TTT GTT TAA TTT GTG TGC TTT CTT TCT TAG TGA TAC ACG AAA TTT GAT

TTT TAA AAA AAA TCA CAG GTA ACC TTA ATG CAT TGT CTT AAC ACA ACA AAG AGC

ATA CAT AGG GTT TCT CTT GGT TTC TTT GAT TAT AAT TCA TAC ATT TTT CTC T**AA**

(Exon #7)

CTG CAA ACA TAA TGT TTT CCC TTG TAT TTT ACA G[AT GCA AAC AGC TAT AAT

F7 (at 24440 bp)

Y A N S Y N

101 102 103 104 105 106

TTT GCA AAA AAG GAA AAT AAC TCT CCT GAA CAT CTA AAA GAT GAA GTT TCT ATC

F A K K E N N S P E H L K D E V S I

107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124

ATC CAA AGT ATG GGC TAC AGA AAC CGT GCC AAA AGA CTT CTA CAG AGT GAA CCC

I Q S M G Y R N R A K R L L Q S E P

125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142

GAA AAT CCT TCC TTG] GTA AAA CCA TTT GTT TTC TTC TTC TTC TTC TTC TTT

E N P S L

143 144 145 146 147

TCT TTT TTT TTT CTT TTT TTT TTT T**GA GAT GGA GTC TTG CTC TGT** GGC CCA GGC

R7 (at 24965 bp)

TAG AAG CAG TCC TCC TGC CTT AGC CCC CTT AGT AGC TGG GAT TAC AGGCAC GCG
CCA CCA TGC CAG GCT AAT TTT TGT ATT TTT AGT AGA GAC GGG GTT TCA TCA TGT
TGG CCA GGC(24821)

(Exon #8)

(28602) GGT CTT GAA CTC CTG GCC TCA AGC AGT CCT GCT CCA GCC TCC CAA AGT
GCT GGG ATT ATA GGC ATG AGC TAC CGC TCC CAG CCC CAG ACA TTT TAG TGT GTA

AAT TCC TGG GCA TTT TTT CCA GGC ATC ATA CAT GTT AGC TGA CTG ATG ATG GTC
AAT TTA TTT TGT CCA TGG TGT CAA GTT TCT CTT CAG GAG GAA AAG CAC AGA ACT

GGC C**AA TAA TTG CTT GAC TGT** TCT TTA CCA TAC TGT TT***AG***[CAG GAA ACC AGT CTC
F8 →
Q E T S L
148 149 150 151 152

AGT GTC CAA CTC TCT AAC CTT GGA ACT GTG AGA ACT CTG AGG ACA AAG CAG CGG
S V Q L S N L G T V R T L R T K Q R
153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170

ATA CAA CCT CAA AAG ACG TCT GTC TAC ATT GAA TTG G]*GTA***A GGG TCT CAG GTT**
← R8
I Q P Q K T S V Y I E L
171 172 173 174 175 176 177 178 179 180 181 182

TTT TAA GTA TTT AAT AAT AAT TGC TGG ATT CCT TAT CTT ATA GTT TTG CCA AAA
←
ATC TTG GTC ATA ATT TGT ATT TGT GGT AGG CAG CTT TGG GAA GTG AAT TTT ATG
AGC CCT ATG GTG AGT TAT AAA AAA TGT AAA AGA CGC AGT TCC CAC CTT GAA GAA
TCT TAC TTT AAA AAG GGA GCA AAA GAG GCC AGG(29168)

(Exon #9)

(31311) ATT GTA CCT GCC ACA GTA GAT GCT CAG TAA ATA TTT CTA GTT GAA
TAT CTG TTT TTC AAC AAG 31374TAC ATT TTT TTA ACC CTT TTA ATT **AAG AAA**
→
F9 (at 31399bp)

ACT TTT ATT GAT TTA TTT TTT GGG GGG AAA TTT **TTT** AG [GA TCT GAT TCT TCT
→
F9
G S D S S
183 184 185 186 187

GAA GAT ACC GTT AAT AAG GCA ACT TAT TGC AG] GTG AGT CAA AGA GAA CCT TTG
E D T V N K A T Y C S
188 189 190 191 192 193 194 195 196 197 198

TCT ATG AAG CTG GTA TTT TCC TAT TTA GTT AAT ATT AAG GAT TGA TGT **TTC TCT**
←
R9
CTT TTT AAA AAT ATT TTA ACT TTT ATT TTA GGT TCA GGG ATG TAT GTG CAG TTT
←
R9 (at 31575)

GTT ATA TAG (31625)

(Exon# 10)

(32634)GCA ACC ATT TCA TTT CAA CTA GAA GTT TCT AAA GGA GAG AGC AGC TTT
CAC TAA CTA AAT AAG ATT GGT CAG CTT TCT GTA ATC GAA AGA GCT AAA ATG TTT
GAT CTT GGT CAT TTG ACA GTT CTG CAT ACA TGT AAC TAG **TGT TTC TTA TTA GGA**

F10

CTC TGT CTT TTC CCT ATA G [T GTG GGA GAT CAA GAA TTG TTA CAA ATC ACC CCT
V G D Q E L L Q I T P
199 200 201 202 203 204 205 206 207 208 209

CAA GGA ACC AGG GAT GAA ATC AGT TTG GAT TCT GCA AAA AAG G] GTA ATG GCA
Q G T R D E I S L D S A K K
210 211 212 213 214 215 216 217 218 219 220 221 222 223

AAG **TTT GCC AAC TTA ACA GGC** ACT GAAAAG AGA GTG GGT AGA TAC AGT ACT GTA

R10

ATT AGA TTA TTC TGA AGA CCA TTT GGG ACC TTT ACA ACC CAC AAA ATC TCT TGG
CAG AGT TAG AGT ATC ATT CTC TGT CAA ATG TCG TGG TAT GGT CTG ATA GAT TTA
AAT GGT ACT AGA CTA ATG TAC CTA TAA TAA GAC CTT CTG TAA CTG ATT GTT GCC
CTT TCG TTT TTT TTT TTG TTT GTT TGT (33138)

(Exon # 11)

(33705)TAA ATG AAA GAG TAT GAG CTA CAT CTT CAG TAT ACT TGG TAG TTT ATG
AGG TTA GTT TCT CTA ATA TAG CCA GTT GGT TGA TTT CCA CCT CCA AGG TGT ATG
AAG TAT GTA TTT TTT TAA TGA CAA TTC AGT TTT TGA GTA CCT **TGT TAT TTT TGT**

F11a (at33848)

ATA TTT TCA G[CT GCT TGT GAA TTT TCT GAG ACG GAT GTA ACA AAT ACT GAA CAT

A A C E F S E T D V T N T E H
224 225 226 227 228 229 230 231 232 233 234 235 236 237 238

CAT CAA CCC AGT AAT AAT GAT TTG AAC ACC ACT GAG AAG CGT GCA GCT GAG AGG
H Q P S N N D L N T T E K R A A E R
239 240 241 242 243 244 246 246 247 249 249 250 251 252 253 254 255 256
CAT CCA GAA AAG TAT CAG GGT AGT TCT GTT TCA AAC TTG CAT GTG GAG CCA TGT
H P E K Y Q G S S V S N L H V E P C
257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274
GGC ACA AAT ACT CAT GCC AGC TCA TTA CAG CAT GAG AAC AGC AGT TTA TTA CTC
G T N T H A S S L Q H E N S S L L L
275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292

F11b **R11a**

ACT AAA GAC AGA ATG AAT GTA GAA AAG GCT GAA TTC TGT AAT AAA AGC AAA CAG
T K D R M N V E K A E F C N K S K Q
293 294 295 296 297 298 399 300 301 302 303 304 305 306 307 308 309 310
CCT GGC TTA GCA AGG AGC CAA CAT AAC AGA TGG GCT GGA AGT AAG GAA ACA TGT
P G L A R S Q H N R W A G S K E T C
311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328

1131A>T 1137delG

AAT GAT AGG CGG ACT CCC AGC ACA GAA AAA AAG GTA GAT CTG AAT GCT GAT CCC
N D R R T P S T E K K V D L N A D P
329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346

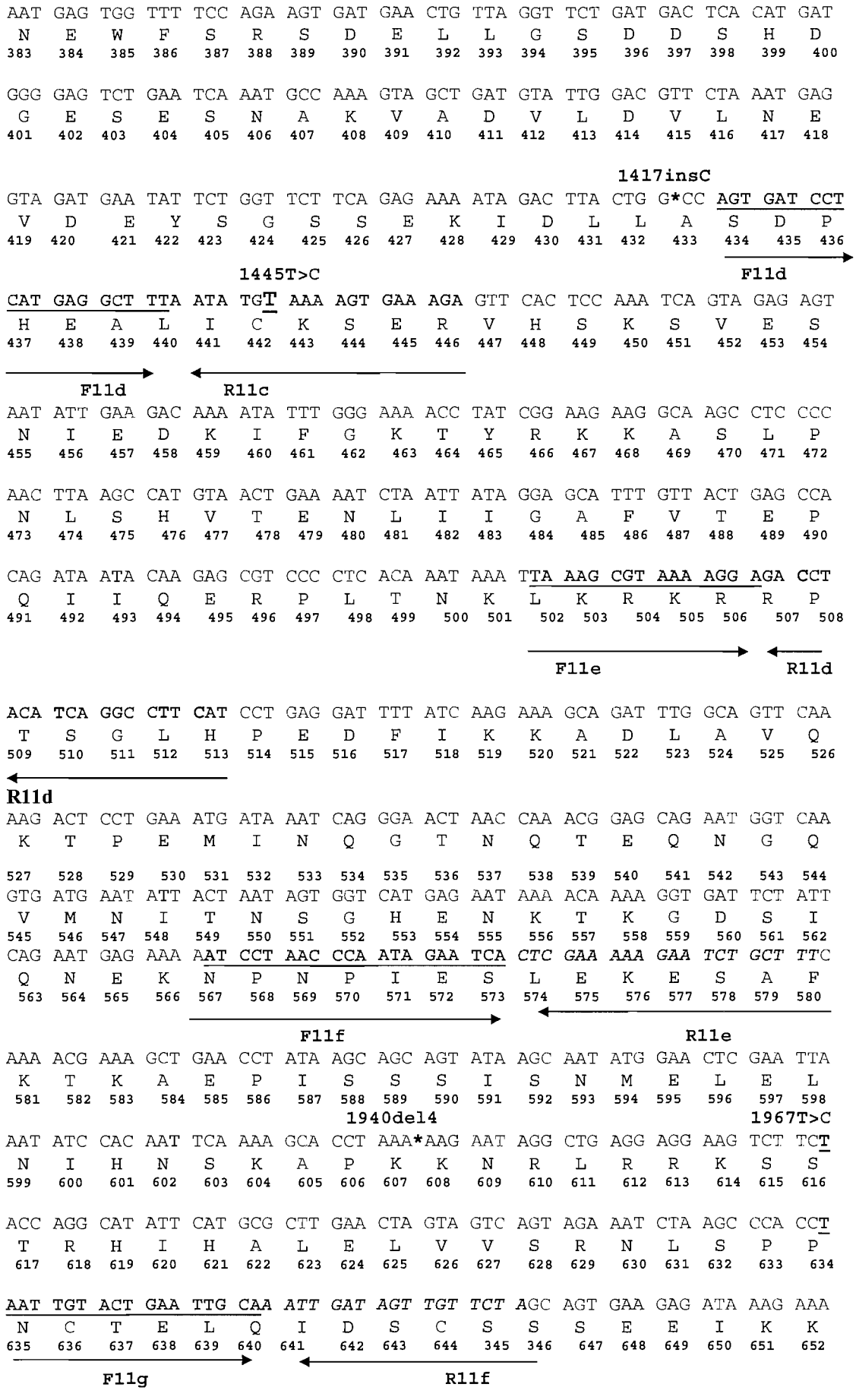
1187G>A (f=5%)

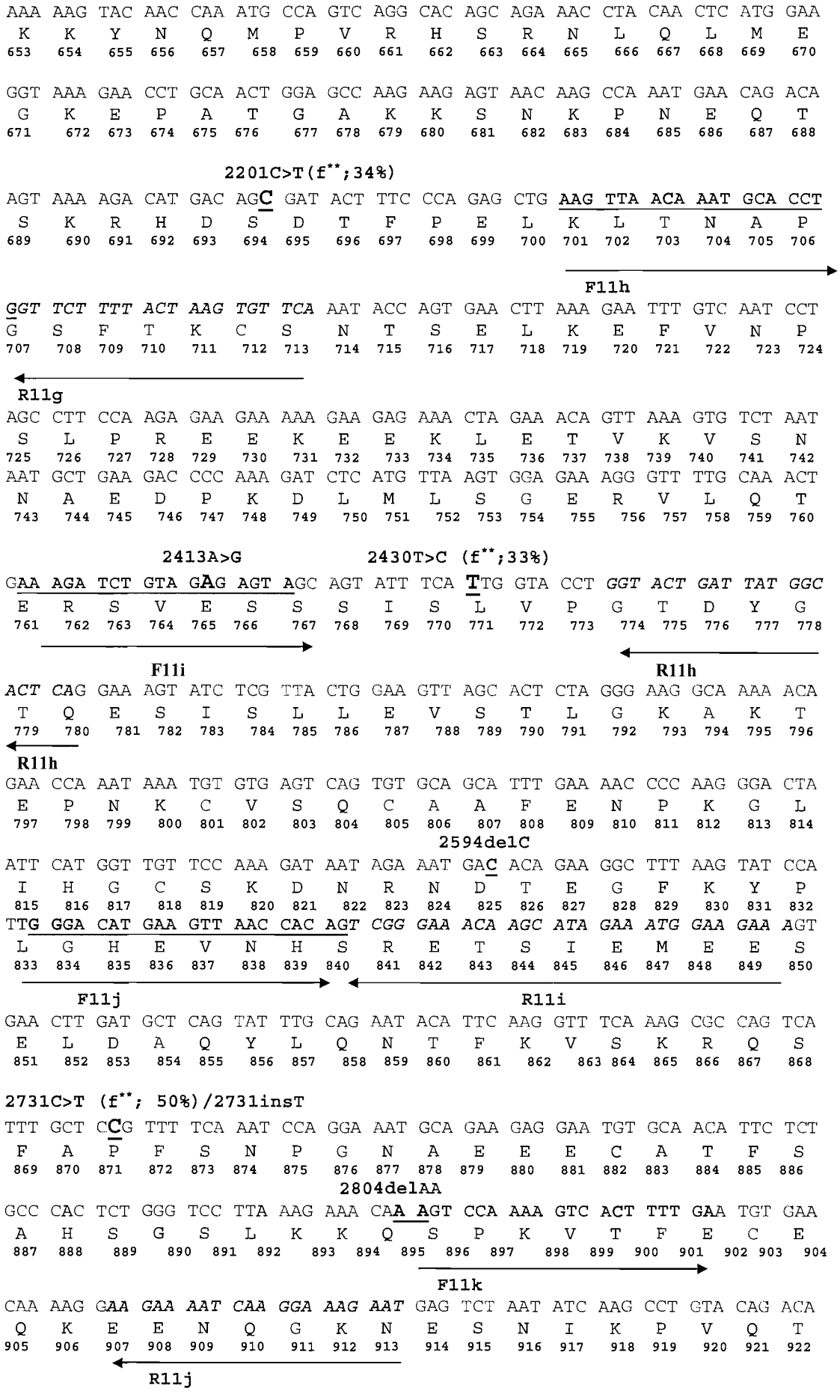
CTG TGT GAG AGA AAA GAA TGG AAT AAG CAG AAA CTG CCA TGC TCA GAG AAT CCT
L C E R K E W N K Q K L P C S E N P
347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364

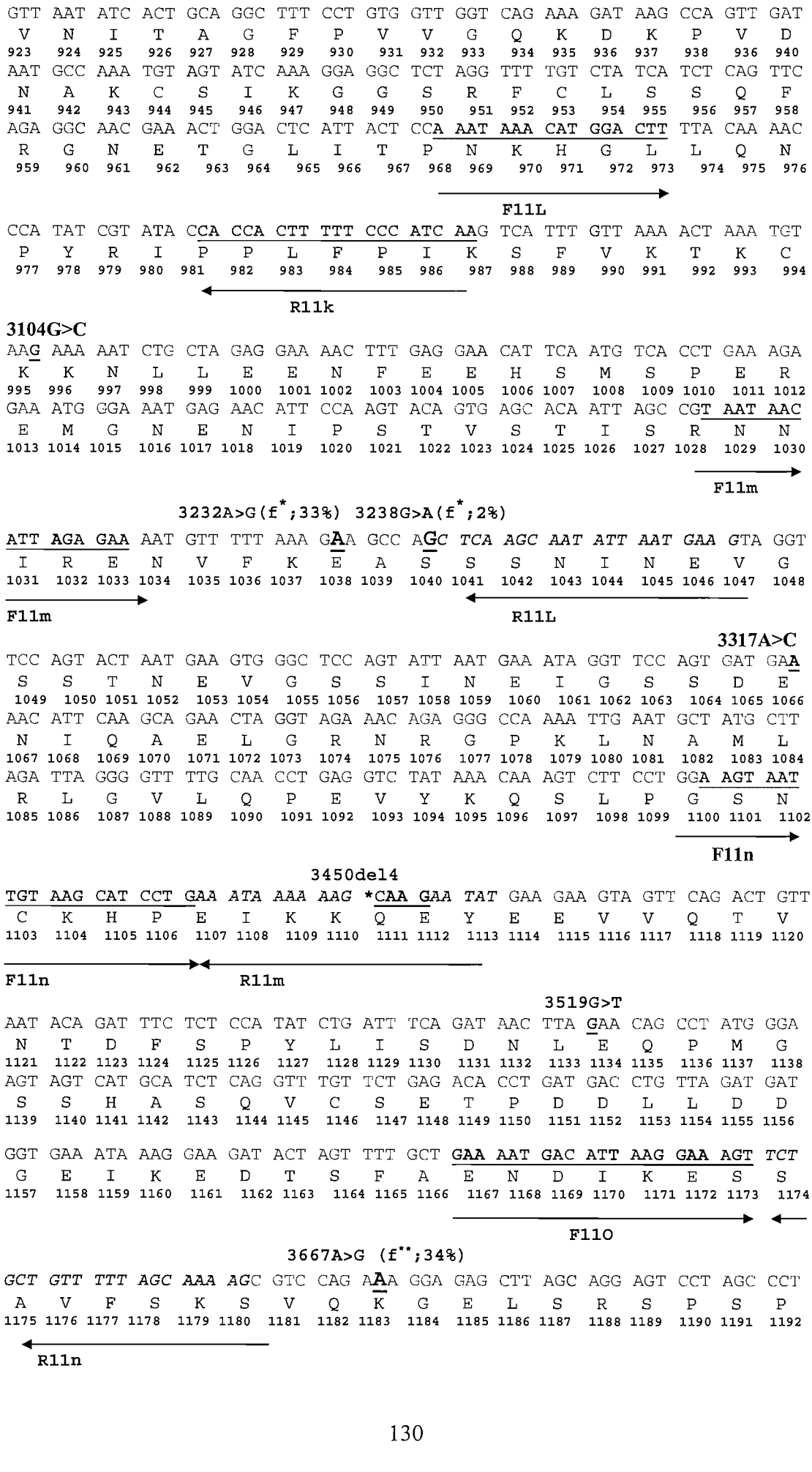
1218insA

AGA GAT *ACT GAA GAT GTT CCT TGG ATA ACA CTA AAT AGC AGC ATT CAG AAA GTT
R D T E D V P W I T L N S S I Q K V
365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382

F11c **R11b**


1417insC
1445T>C
F11d
F11d
R11c
F11e
R11d
R11d
F11f
R11e
1940del4
1967T>C
F11g
R11f


AAA AAG TAC AAC CAA ATG CCA GTC AGG CAC AGC AGA AAC CTA CAA CTC ATG GAA
K K Y N Q M P V R H S R N L Q L M E
653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670
GGT AAA GAA CCT GCA ACT GGA GCC AAG AAG AGT AAC AAG CCA AAT GAA CAG ACA
G K E P A T G A K K S N K P N E Q T
671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688
2201C>T(f**;34%)
AGT AAA AGA CAT GAC AGC GAT ACT TTC CCA GAG CTG AAG TTA ACA AAT GCA CCT
S K R H D S D T F P E L K L T N A P
689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706
F11h
GGT TCT TTT ACT AAG TGT TCA AAT ACC AGT GAA CTT AAA GAA TTT GTC AAT CCT
G S F T K C S N T S E L K E F V N P
707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724
R11g
AGC CTT CCA AGA GAA GAA AAA GAA GAG AAA CTA GAA ACA GTT AAA GTG TCT AAT
S L P R E E K E E K L E T V K V S N
725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742
AAT GCT GAA GAC CCC AAA GAT CTC ATG TTA AGT GGA GAA AGG GTT TTG CAA ACT
N A E D P K D L M L S G E R V L Q T
743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
2413A>G
2430T>C (f**;33%)
GAA AGA TCT GTA GAG AGT AGC AGT ATT TCA TTG GTA CCT GGT ACT GAT TAT GGC
E R S V E S S S I S L V P G T D Y G
761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778
F11i
R11h
ACT CAG GAA AGT ATC TCG TTA CTG GAA GTT AGC ACT CTA GGG AAG GCA AAA ACA
T Q E S I S L L E V S T L G K A K T
779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796
R11h
GAA CCA AAT AAA TGT GTG AGT CAG TGT GCA GCA TTT GAA AAC CCC AAG GGA CTA
E P N K C V S Q C A A F E N P K G L
797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814
2594delC
ATT CAT GGT TGT TCC AAA GAT AAT AGA AAT GAC ACA GAA GGC TTT AAG TAT CCA
I H G C S K D N R N D T E G F K Y P
815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832
TTG GGA CAT GAA GTT AAC CAC AGT CGG GAA ACA AGC ATA GAA ATG GAA GAA AGT
L G H E V N H S R E T S I E M E E S
833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850
F11j
R11i
GAA CTT GAT GCT CAG TAT TTG CAG AAT ACA TTC AAG GTT TCA AAG CGC CAG TCA
E L D A Q Y L Q N T F K V S K R Q S
851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868
2731C>T (f**; 50%)/2731insT
TTT GCT CCG TTT TCA AAT CCA GGA AAT GCA GAA GAG GAA TGT GCA ACA TTC TCT
F A P F S N P G N A E E E C A T F S
869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886
2804delAA
GCC CAC TCT GGG TCC TTA AAG AAA CAA AGT CCA AAA GTC ACT TTT GAA TGT GAA
A H S G S L K K Q S P K V T F E C E
887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904
F11k
CAA AAG GAA GAA AAT CAA GGA AAG AAT GAG TCT AAT ATC AAG CCT GTA CAG ACA
Q K E E N Q G K N E S N I K P V Q T
905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922
R11j


GTT AAT ATC ACT GCA GGC TTT CCT GTG GTT GGT CAG AAA GAT AAG CCA GTT GAT
V N I T A G F P V V G Q K D K P V D
923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 936 940
AAT GCC AAA TGT AGT ATC AAA GGA GGC TCT AGG TTT TGT CTA TCA TCT CAG TTC
N A K C S I K G G S R F C L S S Q F
941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958
AGA GGC AAC GAA ACT GGA CTC ATT ACT CCA AAT AAA CAT GGA CTT TTA CAA AAC
R G N E T G L I T P N K H G L L Q N
959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976
F11L
CCA TAT CGT ATA CCA CCA CTT TTT CCC ATC AAG TCA TTT GTT AAA ACT AAA TGT
P Y R I P P L F P I K S F V K T K C
977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994
R11k
3104G>C
AAG AAA AAT CTG CTA GAG GAA AAC TTT GAG GAA CAT TCA ATG TCA CCT GAA AGA
K K N L L E E N F E E H S M S P E R
995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012
GAA ATG GGA AAT GAG AAC ATT CCA AGT ACA GTG AGC ACA ATT AGC CGT AAT AAC
E M G N E N I P S T V S T I S R N N
1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030
F11m
3232A>G(f*;33%) 3238G>A(f*;2%)
ATT AGA GAA AAT GTT TTT AAA GAA GCC AGC TCA AGC AAT ATT AAT GAA GTA GGT
I R E N V F K E A S S S N I N E V G
1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048
F11m
R11L
3317A>C
TCC AGT ACT AAT GAA GTG GGC TCC AGT ATT AAT GAA ATA GGT TCC AGT GAT GAA
S S T N E V G S S I N E I G S S D E
1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066
AAC ATT CAA GCA GAA CTA GGT AGA AAC AGA GGG CCA AAA TTG AAT GCT ATG CTT
N I Q A E L G R N R G P K L N A M L
1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084
AGA TTA GGG GTT TTG CAA CCT GAG GTC TAT AAA CAA AGT CTT CCT GGA AGT AAT
R L G V L Q P E V Y K Q S L P G S N
1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102
F11n
3450del4
TGT AAG CAT CCT GAA ATA AAA AAG *CAA GAA TAT GAA GAA GTA GTT CAG ACT GTT
C K H P E I K K Q E Y E E V V Q T V
1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120
F11n
R11m
3519G>T
AAT ACA GAT TTC TCT CCA TAT CTG ATT TCA GAT AAC TTA GAA CAG CCT ATG GGA
N T D F S P Y L I S D N L E Q P M G
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138
AGT AGT CAT GCA TCT CAG GTT TGT TCT GAG ACA CCT GAT GAC CTG TTA GAT GAT
S S H A S Q V C S E T P D D L L D D
1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156
GGT GAA ATA AAG GAA GAT ACT AGT TTT GCT GAA AAT GAC ATT AAG GAA AGT TCT
G E I K E D T S F A E N D I K E S S
1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174
F11O
3667A>G (f**;34%)
GCT GTT TTT AGC AAA AGC GTC CAG AAA GGA GAG CTT AGC AGG AGT CCT AGC CCT
A V F S K S V Q K G E L S R S P S P
1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192
R11n

TTC ACC CAT ACA CAT TTG GCT CAG GGT TAC CGA AGA GGG GCC AAG AAA TTA GAG
F T H T H L A Q G Y R R G A K K L E
1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210
TCC TCA GAA GAG AAC TTA TCT AGT GAG GAT GAA GAG CTT CCC TGC TTC CAA CAC
S S E E N L S S E D E E L P C F Q H
1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228

3825delAA

TTG TTA TTT GGT AAA GTA AAC AAT ATA CCT TCT CA *G TCT ACT AGG CAT AGC* ACC
L L F G K V N N I P S Q S T R H S T
1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246

F11p R11O

3875del4 3881delAG

GTT GCT ACC GAG TGT CTG TCT AAG AAC ACA GAG GAG AAT TTA TTA TCA TTG AAG
V A T E C L S K N T E E N L L S L K
1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264

AAT AGC TTA AAT GAC TGC AGT AAC CAG GTA ATA TTG GCA AAG GCA TCT CAG GAA
N S L N D C S N Q V I L A K A S Q E
1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282

CAT CAC CTT AGT GAG GAA ACA AAA TGT TCT GCT AGC TTG TTT TCT TCA CAG TGC
H H L S E E T K C S A S L F S S Q C
1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300

AGT GAA TTG GAA GAC TTG *ACT GCA AAT ACA AAC* ACC CAG GAT CCT TTC TTG ATT
S E L E D L T A N T N T Q D P F L I
1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318

F11q R11p

GGT TCT TCC AAA CAA ATG AGG CAT CAG TCT GAA AGC CAG GGA GTT GGT CTG AGT
G S S K Q M R H Q S E S Q G V G L S
1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336

4158delAG/4158A>G 4176G>T

GAC AAG GAA TTG GTT TCA GAT GAT GAA GAA AGA GGA ACG GGC TTG GAA GAA AAT
D K E L V S D D E E R G T G L E E N
1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354
AAT CAA GAA GAG CAA AGC ATG GAT TCA AAC TTA G]GT *ATT GGA ACC AGG TTT TTG*

R11q

N Q E E Q S M D S N L
1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365

TGT TTG CCC CAG TCT ATT TAT AGA AGT GAG CTA AAT GTT TAT GCT TTT GGG GAG

R11q

CAC ATT TTA CAA ATT TCC AAG TAT AGT TAA AGG AAC TGC TTC TTA AAC TTG AAA
CAT GTT CCT CCT AAG GTG CTT TTC ATA GAA AAA AGT CCT TCA CAC AGC TAG GAC
GTC ATC TTT GAC TGA ATG AGC TTT AAC ATC CTA ATT ACT GGT GGA CTT ACT TCT
GGT TTC ATT TTA TAA (37548)

(Exon # 12)

AAG CAA ATC CAG GTG TCC CAA AGC AAG GAA TTT AAT CAT TTT GTG TGA CAT GA
AGT AAA TCC AGT CCT GCC AAT GAG AAG AAA AAG ACA CAG CAA GTT GCA GCG TTT
ATA GTC TGC TTT TAC ATC TGA ACC TCT GTT TTT GTT ATT TAAG [GT GAA GCA GCA

F12

G E A A
366 1367 1368 1369

TCT GGG TGT GAG AGT GAA ACA AGC GTC TCT GAA GAC TGC TCA GGG CTA TCC TCT
S G C E S E T S V S E D C S G L S S
1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387
CAG AGT GAC ATT TTA ACC ACT CAG] GTA AAA AGC GTG TGT GTG TGT GCA CAT GCG
Q S D I L T T Q
1388 1389 1390 1391 1392 1393 1394 1395
TGT GTG TGG TGT CCT TTG CAT TCA GTA GTA TGT ATC CCA CAT TCT TAG GTT TGC
TGA CAT CAT CTC TTT GAA TTA ATG GCA CAA TTG TTT GTG GTT CAT TGT CTC CTT

R12

(Exon # 13)

AAA TTA GAC TGT AAG CAC CTT GAT GGA ACT CAT ACT ACC TTT TAT TTC ACA CAC
ACG CAC ACG CGC ACA CAC AGC CTA CAC ATA CAC TGC CTA GCT CAT TGT AGC ATA
CTAAATACTGATTTTAAT(38052)
CTA ACC TAA ATG TCT GAT GAA TCT CAC AAC ACC AAG TCT TTG AAA TGT GCC CAT
ATA AAT AAA ATG TTA ACA GAT TCA TGC TAA TTT TAA ATA TCG ATA GTG TTT AAA
TGC CTT AAT TAT TTT TTC ACT CCC TAG CTT TAA AAG AAA ATA ACC AAC TTC AAA
AGG ACA TCA CAA TAA CAT CAA GTC TAT TTG GGG GAA TTT GAG GAT TTT TTC CCT
CAC TAA CAT CAT TTG GAA ATA ATT TCA TGG GCA TTA TTA GAT TAA AAG GTG TTC
AGC TAG AAC TTG TAG TTC CAT ACT AGG TGA TTT CAA TTC CTG TGC TAA AAT TAA
TTT GTA TGA TAT ATT TTC ATT TAA TGG AAA GCT TCT CAA AGT ATT TCA TTT TCT

TGG T**GC CAT TTA TCG TTT TTG AAG** [CAG AGG GAT ACC ATG CAA CAT AAC CTG ATA

F13

Q R D T M Q H N L I
1396 1397 1398 1399 1400 1401 1402 1403 1404 1405

AAG CTC CAG CAG GAA ATG GCT GAA CTA GAA GCT GTG TTA GAA CAG CAT GGG AGC
K L Q Q E M A E L E A V L E Q H G S
1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423

4410T>C **4427T>C**

CAG CCT TCT AAC AGC TAC CCT **T**CC ATC ATA AGT GAC **TC<u>T</u>** TCT GCC CTT GAG GAC
Q P S N S Y P S I I S D S S A L E D
1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441
CTG CGA AAT CCA GAA CAA AGC ACA TCA GAA AAA G]GT GTG TAT TGT TGG CCA AAC
L R N P E Q S T S E K
1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452
ACT GAT ATC TTA AGC AAA ATT CTT TCC TTC CCC TTT ATC TCC TTC TGA **AGA GTA**

R13

AGG ACC TAG CTC CAA CAT TTT ATG ATC CTT GCT CAG CAC ATG GGT AAT TAT GGA

R13

GCC TTG GTT CTT GTC CCT GCT CAC AAC TAA TAT ACC AGT (46494)

(Exon #14)

(51975) AAA ATT CTG CCT GAT ATA CTT GTT TAA AAA CCA ATT TGT GTA TCA TAG
ATT GAT GCT TTT GAA AAA AAT CAG TAT TCT AAC CTG AAT TAT CAC T**AT CAG AAC**

F14

AAA GCA GTA AAG TAG ATT TGT TTT CTC ATT CCA TTT AAA G[CA GTA TTA ACT TCA
A V L T S
F14 1453 1454 1455 1456 1457
CAG AAA AGT AGT GAA TAC CCT ATA AGC CAG AAT CCA GAA GGC CTT TCT GCT GAC
Q K S S E Y P I S Q N P E G L S A D
1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475
AAG TTT GAG GTG TCT GCA GAT AGT TCT ACC AGT AAA AAT AAA GAA CCA GGA GTG
K F E V S A D S S T S K N K E P G V
1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493
GAA AG]G TAA GAA ACA TCA ATG TAA AGA TGC TGT GGT ATC TGA CAT CTT TAT TTA
E R
1494 1495
TAT TGA ACT **CTG ATT GTT AAT TTT TTT** CAC CAT ACT TTC TCC AGT TTT TTG CAT

R14

ACA GGC ATT TAT ACA CTT TTA TTG CTC TAG GAT ACT TCT TTT GTT TAA TCC
TATATA GGT TTT TTG AAC CTA TAA CAT AAG CTA CAA CAT GAG AAA TGT GCG GTT
AGA TAG ATA TGT CCC TTC TGA AGG TCA (52479). (53928) GAT TAC AGG CGC CAG
CCA CCA CAC CCA GCT ACT GAC CTG CTT TTA AAC AGC TGG GAG ATA TGG TGC CTC
AGA CCA ACC CAA CCC CAT GTT ATA TGT CAA CCC TGA CAT ATT GGC AGG CAA CAT
GAA TCC AGA CTT CTA GGC TGT CTT GCG GGC TCT TTT TTG CCA GTC ATT TCT

GAT CTC TCT GAC ATG AGC TGT TTC ATT TAT GCT TTG GCT GCC CAG CAA GTA TGA
TTT GTC CTT TCA CAA TTG

(Exon #15)

GCC CAG CAA GTA TGA **TTT GTC CTT TCA CAA TTG GTG** GCG ATG GTT TTC TCC TTC

F15

CAT TTA TCT TTC TAG[G TCA TCC CCT TCT AAA TGC
S S P S K C
1496 1497 1498 1499 1500 1501
CCA TCA TTA GAT GAT AGG TGG TAC ATG CAC AGT TGC TCT GGG AGT CTT CAG AAT
P S L D D R W Y M H S C S G S L Q N
1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519
AGA AAC TAC CCA TCT CAA GAG GAG CTC ATT AAG GTT GTT GAT GTG GAG GAG CAA
R N Y P S Q E E L I K V V D V E E Q
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537
CAG CTG GAA GAG TCT GGG CCA CAC GAT TTG ACG GAA ACA TCT TAC TTG CCA AGG
Q L E E S G P H D L T E T S Y L P R
1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555
CAA GAT CTA G]G TAA TAT TTC ATC TGC TGT ATT GGA ACA AAC ACT TTG **ATT TTA**
Q D L
1556 1557 1558 **R15**

CTC TGA ATC CTA CAT AAA GAT ATT CTG GTT AAC CAA CTT TTA GAT GTA CTA GTC

R15

TAT CAT GGA CAC TTT TGT TAT ACT TAA TTA AGC CCA CTT TAG AAA AAT AGC
TCAAGT GTT AAT CAA GGT TTA CTT GAA AAT TAT TGA AAC TGT TAA TCC ATC TAT
ATTTTA ATT AAT GGT TTA (54621).(57330)TTT TAG GAA ACT AAG CTA CTT TGG
ATTTCC ACC AAC ACT GTA

(Exon#16)

TTC ATG TAC CCA TTT TTC TCT TAA CCT AAC TTT ATT GGT CTT TTT AAT TCT TAA
CAG AGA CCA GAA CTT TGT AAT **TCA ACA TTC ATC GTT GTG TAA** ATT AAA CTT CTC

CCA TTC **F16 (at 57447)**

CTT TCA G[AG GGA ACC CCT TAC CTG GAA TCT GGA ATC AGC CTC TTC TCT GAT GAC
E G T P Y L E S G I S L F S D D
1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574
CCT GAA TCT GAT CCT TCT GAA GAC AGA GCC CCA GAG TCA GCT CGT GTT GGC AAC
P E S D P S E D R A P E S A R V G N
1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592
ATA CCA TCT TCA ACC TCT GCA TTG AAA GTT CCC CAA TTG AAA GTT GCA GAA TCT
I P S S T S A L K V P Q L K V A E S
1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610
4956A>G
GCC CAG **A**GT CCA GCT GCT GCT CAT ACT ACT GAT ACT GCT GGG TAT AAT GCA ATG
A Q S P A A A H T T D T A G Y N A M
1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628
GAA GAA AGT GTG AGC AGG GAG AAG CCA GAA TTG ACA GCT TCA ACA GAA AGG GTC
E E S V S R E K P E L T A S T E R V
1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646
AAC AAA AGA ATG TCC ATG GTG GTG TCT GGC CTG ACC CCA GAA GAA TTT] GTG AGT
N K R M S M V V S G L T P E E F
1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662
GT**A TCC ATA TGT ATC TCC CTA A**TG ACT AAG ACT TAA CAA CAT TCT GGA AAG AGT

R16 (at 57831)

TTT ATG TAG GTA TTG TCA ATT AAT AAC CTA GAG GAA GAA ATC TAG AAA ACA ATC
ACA GTT CTG TGT AAT TTA ATT TCG ATT ACT AAT TTC TGA AAA (57959)/(60858)
ACA GAG CAA GAC TCT GTC TAA AAA AAA AAA AAA AAA AAT TCA CTT TAA ATA GTT

(Exon#17)

CCA GGA CAC GTG TAG AAC GTG CAG GAT TGC TAC ATA GGT AAA CAT ATG CCA TGG
TGG **AAT AAC TAG TAT TCT GAG CTG** TGT GCT AGA GGT AAC TCA TGA TAA TGG AAT

F17 (at 60969)

ATT TGA TTT AAT TTC AG [ATG CTC GTG TAC AAG TTT GCC AGA AAA CAC CAC ATC
M L V Y K F A R K H H I
1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674

5143C>T

AC̲T TTA ACT AAT CTA ATT ACT GAA GAG ACT ACT CAT GTT GTT ATG AAA ACA G]
T L T N L I T E E T T H V V M K T
1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691
GTATAC CAA GAA CCT TTA CAG AAT ACC TTG CAT **CTG CTG CAT AAA ACC ACA TGA**

R17 (at 61178bp)

GGC GAG GCA CGG TGG CGC ATG CCT GTA ATC GCA GCA CTT TGG GAG GCC GAG GCG
GGC (61236)

(Exon # 18)

CAC CAG GGG TTT TAG AAT CAT AAA TCC AGA TTG ATC TTG GG**A GTG TAA AAA ACT**

F18 (at 646687)

GAG GCT CTT TAG CTT CTT AGG ACA GCA CTT CCT GAT TTT GTT TTC AAC TTC TAA
TCC TTT GAG TGT TTT TCA TTC TGC AG [AT GCT GAG TTT GTG TGT GAA CGG ACA
D A E F V C E R T
1692 1693 1694 1695 1696 1697 1698 1699 1700
CTG AAA TAT TTT CTA GGA ATT GCG GGA GGA AAA TGG GTA GTT AGC TAT TTC T]GT
L K Y F L G I A G G K W V V S Y F
1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717
AA GTA TAA TAC TAT TTC TCC CCT CCT CCC TTT AAC ACC TCA GAA TTG CAT TTT
TAC ACC TAA CGT TTA ACA CCT AAG GTT TTT GCT GAT GCT GAG TCT GAG TTA CCA
AAA GGT **CTT TAA TTG TAA TAC TAA A**CT ACT TTT ATC TTT AAT ATC ACT (65015)

R18 (at 65000bp)

(Exon # 19)

65205 TAC AAG CCT TAT TAA AGG GCT GTG GCT TTA GAG GGA AGG ACC TCT CCT
CTG TCA TTC TTC CTG TGC TCT TTT GTG AAT CGC TGA CCT CTC TAT CTC CGT GAA
AAG AGC ACG TTC TTC TGC TGT ATG TAA CCT GT**C TTT TCT ATG ATC TCT TTA G**

R19 (at 65339bp)

[GG GTG ACC CAG TCT ATT AAA GAA AGA AAA ATG CTG AAT GAG] GTA AGT ACT TGA
W V T Q S I K E R K M L N E
1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731
TGT TAC AAA CTA ACC AGA GAT ATT CAT TCA GTC ATA TAG TTA AAA ATG TAT TTG
CTT CCT TC**C ATC AAT GCA CCA CTT TCC** TTA ACA ATG CAC AAA TTT TCC ATG ATA

R19 (at 65494bp)

ATG AGG ATC ATC AAG AAT TAT GCA GGC CTG CAC TGT GGC TCA TAC CTA TAA TCC
CAG CGC TTT (65583)/ 71316 GCC TCA GCC TCC CAA GTA GCT GGG ATT ACA GGT
GCC TGC CAC CAC GCC CAA CTA ATT TTT TGT ATT TTT AGT AGA GAT GAG GTT TCA
CCA TGT TGG TCA GAC TGG TGT CGA ACT CCT GAC CTC AAG TGA TCT GCC TGC CTC

(Exon # 20)

```
AGT CTC CCA AAG TGC TAG GAT TAC AGG GGT GAG CCA CTG CGC CTG GCC TGA ATG
CCT TAA ATA TGA CGT GTC TGC TCC ACT TCC ATT GAA GGA AGC TTC TCT TTC TCT
                                 ---------->
                                 F20 (at 71532bp)
TAT CCT GATGGG TTG TGT TTG GTT TCT TTC AG[CAT GAT TTT GAA GTC AGA GGA
                                          H   D   F   E   V   R   G
                                       1732 1733 1734 1735 1736 1737 1738
GAT
D
 1739
                                                              5382insC
GTG GTC AAT GGA AGA AAC CAC CAA GGT CCA AAG CGA GCA AGA GAA T*CC CAG GAC AGA
V  V   N   G   R   N   H   Q   G   P   K   R   A   R   E   S   Q   D   R
1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753  1754 1755 1756  1757  1758

AAG] GTA AAGC TCC CTC CCT CAA GTT GAC AAA AAT CTC ACC CCA CCA CTC TGT
K
1759

ATT CCA CTC CCC TTT GCA GAG ATG GGC CGC TTC ATT TTG TAA GAC TTA TTA CAT
                           <----------
ACA TAC ACA                 R20 (at 71771bp)
```

(Exon # 21)

```
GTG CTA GAT ACT TTC ACA CAG GTT CTT TTT TCA CTC TTC CAT CCC AAC CAC ATA
AAT AAG TAT TGT CTC TAC TTT ATG AAT GAT AAA ACT (71883)/(77427) CAG TAG
TGT TCT GGA CAT TGG ACT GCT TGT CCC TGG GAA GTA GCA GCAGAA ATC ATC AGG
TGG TGA ACA GAA GAA AAA GAA AAG CTC TTC CTT TTT GAA AGT CTG TTT TTT GAA
TAA AAG CCA ATA TTC TTT TAT AAC TAG ATT TTC CTT CTC TCC ATT CCC CTG TCC
                  ---------->
                  F21 (at 77554bp)

CTC TCTCTT CCT CTC TTC TTC CAG[ATC TTC AGG GGG CTA GAA ATC TGT TGC TAT
                                I   F   R   G   L   E   I   C   C   Y
                              1760 1761 1762 1763 1764 1765 1766 1767 1768 1769
GGG CCC TTC ACC AAC ATG CCC ACA G]GT AAG AGC CTG GGA GAA CCC CAG AGT TCC
 G   P   F   T   N   M   P   T
1770 1771 1772 1773 1774 1775 1776 1777
AGC ACC AGC CTT TGT CTT ACA TAG TGG AGT ATT ATA AGC AAG ATC CCA CGA TGG
                                           <----------
                                           R21 (at 77754bp)
GGG TTC CTC AGA TTG CTG AAA TGT TCT AGA GGC TAT TCT ATT TCT CTA CCA CTC
TCC AAA CAA AAC AGC ACC TAA ATG TTA TCC TAT GGC AAA AAA AAA CTA TAC CTT
GTC (77868)/(79380) CAT ACC CCT ACT ATT TTA AGA CCA TTG TCC TTT GGA GCA
GAG AGA CAG ACT CTC CCA TTG AGA GGT CTT GCT ATA AGC CTTCAT CCG GAG AGT
```

(Exon # 22)

```
GTA GGG TAGAGG GCC TGG GTT AAG TAT GCA GAT TAC TGC AGT GAT TTT ACA TCT
          ---------->
          F22
AAA TGT CCA TTT TAG [AT CAA CTG GAA TGG ATG GTA CAG CTG TGT GGT GCT TCT
                    D   Q   L   E   W   M   V   Q   L   C   G   A   S
                  1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790
GTG GTG AAG GAG CTT TCA TCA TTC ACC CTT GGC ACA]G TAA GTATTG GGT GCC CTG
 V   V   K   E   L   S   S   F   T   L   G   T
1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801
TCA GAG AGG GAG GAC ACA ATA TTC TCT CCT GTG AGC AAG ACT GGC ACC TGT CAG
                                           <----------
                                           R22
TCC CTA TGG ATG CCC CTA CTG TAG CCT CAG AAG TCT TCT CTG CCC ACA TAC CTG
TGC CAA AAG ACT CCA TCT GTA AGG GAT GGG TAA GGA TTT
```

(Exon # 23)

(80703) ATG CCT ATAATC CCG GGG GTG AAG TTG AGC CCA GGA GTT TGA GAC CAG
CCT GGG CAA CAT GGCAAA ACC CTG TCT CTA CCA AAA ATA CAA AAA AAT TAG CCA
GGG GTG GTG GTA CGT GTC TGT AGT TCC AGC TAC TTA GGA GGC TGA GAT GGA AGG
ATT GCT TGA GCC CAGGAG GCA GAG GTG GCA GTG AGC TGA GAT CAC ACC ACT GCA
CTC CAG CCT GGG TGA CAG AGC AAG ACC CTG TCT CAA AAA CAA ACA AAA AAA **ATG**
ATG AAG TG**A CAG TTC CAG TAG TCC TAC** TTT GAC ACT TTG AAT GCT CTT TCC TTC

F23 (at 80973bp)

CTG GGG ATC CAG [GGT GTC CAC CCA ATT GTG GTT GTG CAG CCA GAT GCC TGG ACA
G V H P I V V V Q P D A W T
1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815

GAG GAC AAT GGC TTC CAT G]GT AAG GTG CCT G**CA TGT ACC TGT GCT ATA TGG** GGT

R23 (at81203bp)

E D N G F H
1816 1817 1818 1819 1820 1821
CCT TTT GCA TGG GTT TGG TTT ATC ACT CAT TAC CTG GTG CTT GAG TAG CAC AGT

(Exon # 24)

TCT TGG CAC ATT TTA AAT ATT TGT TGA ATG AAT GGC TAA AAT GTC TTT TTG ATG
TTT TTA TTG TTA TTT GTT TTA TAT TGT AAA AGT AAT ACA TGA ACT GTT TCC ATG
GGG TGG GAG TAA GAT ATG AAT GTT CAT CAC AAA AAC ATA AAT (81333)/(82782)
CAA GAC TCC GTC TCA AAA AAA AAA AAA AAA AAA TTA GCT TCT ACC TCA TTA ATC
CTA AGA ACT CAT ACA ACC AGG ACC **CTG GAG TCG ATT GAT TAG A**GC CTA GTC CAG

F24a (at 82860bp)

GAG AAT GAA TTG ACA CTA ATC TCT GCT TGT GTT CTC TGT CTC CAG[CA ATT GGG
A I G
1822 1823 1824

CAG ATG TGT GAG GCA CCT GTG GTG ACC CGA GAG TGG GTG TTG GAC AGT GTA GCA
Q M C E A P V V T R E W V L D S V A
1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842
CTC TAC CAG TGC CAG GAG CTG GAC ACC TAC CTG ATA CCC CAG ATC CCC CAC AGC
L Y Q C Q E L D T Y L I P Q I P H S
1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860

Stop codon
(
CAC TAC TGA] CTG CAG CCA GCC ACA GGT ACA GAG CCA CAG GAC C
H Y .
1861 1862 1863

CCA AGA ATG AGC TTA CAA AGT **GGC CTT TCC AGG CCC T**GG GAG CTC CTC TCA CTC

R24a **F24b**

TTC AGT CCT TCT ACT GTC CTG GCT ACT AAA TAT TTT ATG TAC ATC AGC CTG AAA
AGG ACT TCT GGC TAT GCA AGG GTC CCT TAA AGA TTT TCT GCT TGA AGT CTC CCT
TGG AAA TCT GCC ATG AGC *ACA AAA TTA TGG TAA TTT TTC* ACC TGA GAA GAT TTT

F24c **R24b**

AAA ACC ATT TAA ACG CCA CCA ATT GAG CAA GAT GCT GAT TCA TTA TTT ATC AGC

R24b

CCT ATT CTT TCT ATT CAG GCT GTT GTT GGC TTA GGG CTG GAA GCA CAG AGT GGC
TTG GCC TC*A AGA GAA TAG CTG GTT TCC* CTA AGT TTA CTT CTC TAA AAC CCT GTG

F24d **R24c**

TTC ACA AAG GCA GAG AGT CAG ACC CTT CAA TGG AAG GAG AGT GCT TGG GAT CGA

TTA TGT GAC TTA AAG TCA GAA TAG TCC TTG GGC AGT TCT CAA ATG TTG GAG TGG
AAC ATT GGG GAG GAA ATT CTG AGG CAG GTA TTA GAA ATG AAA AGG AAA CTT GAA
ACC TGG GCA TGG TGG CTC ACG CCT GTA ATC CCA GCA CTT TGG GAG GCC AAG GTG

R24d

GGC AGA TCA CTG GAG GTC AGG AGT TCG AAA CCA GCC TGG CCA ACA TGG TGA AAC
CCC ATC TCT ACT AAA AAT ACA GAA ATT AGC CGG TCA TGG TGG TGG ACA CCT GTA
ATC CCA GCT ACT CAG GTG GCT AAG GCA GGA GAA TCA CTT CAG CCC GGG AGG TGG
AGG TTG CAG TGA GCC AAG ATC ATA CCA CGG CAC TCC AGC CTG GGT GAC AGT GAG
ACT GTG GCT CAA AAA AAA AAA AAA AAA AGG *AAA ATG AAA CTA GGA AAG GTT* TCT

F24e

TAA AGT CTG AGA TAT ATT TGC TAG ATT TCT AAA GAA TGT GTT CTA AAA CAG CAG
AAG ATT TTC AAG AAC CGG TTT CCA AAG ACA GTC TTC TAA TTC CTC ATT AGT AAT
AAG TAA AAT GTT TAT TGT TGT AGC TCT GGT ATA TAA TCC ATT CCT CTT AAA ATA
TAA GAC CTC TGG CA*T GAA TAT TTC ATA TCT ATA* AAA TGA CAG ATC CCA CCA GGA

F24f R24e

AGG AAG CTG TTG CTT TCT TTG AGG TGA TTT TTT TCC TTT GCT CCC TGT TGC TGA
AAC CAT ACA GCT TCA TAA ATA ATT TTG CTT GCT GAA GGA AGA AAA AGT GTT TTT
CAT AAA CCC ATT ATC CAG GAC TG[**T TTA TAG CTG TTG GAA GGA C**TA GGT CTT CCC

F24g R24f

(The new polymorphism (C/T) detected by melt-MADGE)

TAG CCC CCC CAG TGT GCA AGG GCA GTG AAG ACT TGA TTG TAC AAA ATA CGT TTT

A-tail

GTA AAT GTT GTG CTG TTA ACA CTG C*AA ATA A*AC TTG GTA GCA AAC ACT TCC ACC

ATG AAT GAC TGT TCT TGA GAC TTA GGC CAG **CCG ACT TTC TCA GAG CCT TTT]** CAC

R24g

TGT GCT TCA GTC TCC CAC TCT GTA AAA TGG GGG TAA TGA TAG TAT CTA CCT CCT
AGG ATT TAT TGA GGC AGC TTA AAT ACC TTT TGT ATT TCC TGT TGC TGC CAA AAC
AAA TTG TTG CAA GGT CAG AAG TCT GAG GTG GCT CAA CTG TTT CTT TGT TTC AGG
TTT CAT GAG GCC AAA ATA AAG GTG TTC GCA GGG CGT GTT CCC TTC TAG AGG CTC

CHAPTER SIX

References

1. Albert B, et al. (1994a) Macromolecules: structure, shape and information. Molecular biology of the cell.

2. Albert B, et al. (1994b) the cell-division cycle. Molecular Biology of the cell.

3. Antoniou A et al (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies Am.J.Hum.Genet. 72 (5):1117-1130.

4. Backe J et al (1999) Frequency of BRCA1 mutation 5382insC in German breast cancer patients Gynecol.Oncol. 72 (3):402-406.

5. Begg CB (2002) On the use of familial aggregation in population-based case probands for calculating penetrance J.Natl.Cancer Inst. 94 (16):1221-1226.

6. Bounpheng M et al (2003) Rapid, inexpensive scanning for all possible BRCA1 and BRCA2 gene sequence variants in a single assay: implications for genetic testing J.Med.Genet. 40 (4):e33.

7. Brookes AJ (1999) The essence of SNPs. Gene 234 (2):177-186.

8. Brzovic PS et al (2001a) BRCA1 RING domain cancer-predisposing mutations. Structural consequences and effects on protein-protein interactions J.Biol.Chem. 276 (44):41399-41406.

9. Brzovic PS et al (2001b) Structure of a BRCA1-BARD1 heterodimeric RING-RING complex Nat.Struct.Biol. 8 (10):833-837.

10. Burke W and Austin MA (2002) Genetic risk in context: calculating the penetrance of BRCA1 and BRCA2 mutations J.Natl.Cancer Inst. 94 (16):1185-1187.

11. Cargill M et al (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat.Genet 22 (3):231-238.

12. Chakravarti A (1999) Population genetics--making sense out of sequence. Nat.Genet 21 (1 Suppl):56-60.

13. Chapman MS and Verma IM (1996) Transcriptional activation by BRCA1 Nature 382 (6593):678-679.

14. Claes K et al (2003) Differentiating pathogenic mutations from polymorphic alterations in the splice sites of BRCA1 and BRCA2Genes Chromosomes.Cancer 37 (3):314-320.

15. Claus EB et al (1998) Effect of BRCA1 and BRCA2 on the association between breast cancer risk and family history J.Natl.Cancer Inst. 90 (23):1824-1829.

16. Collins FS, Brooks LD, and Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation Genome Res. 8 (12):1229-1231.

17. Conne B, Stutz A, and Vassalli JD (2000) The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? Nat.Med. 6 (6):637-641.

18. Cotton RGH. 1997. Mutation Detection. Oxford: Oxford University press.

19. Day INM et al (2003) Electrophoresis in Microplate formats. In S.Ye, I.N.M Day (eds) Microarrays & Microplates; Applications in Biomedical Sciences. BIOS Scientific Puplisher Limited.

20. Day IN et al (2001) Epidemiology and the genetic basis of disease. Int.J.Epidemiol. 30 (4):661-667.

21. Day IN et al (1995a) Electrophoresis for genotyping: temporal thermal gradient gel electrophoresis for profiling of oligonucleotide dissociation Nucleic Acids Res. 23 (13):2404-2412.

22. Day IN et al (1999a) Microplate array diagonal gel electrophoresis (MADGE), CpG-PCR and temporal thermal ramp-MADGE (Melt-MADGE) for single nucleotide analyses in populations. Genet Anal. 14 (5-6):197-204.

23. Day IN et al (1999b) Microplate array diagonal gel electrophoresis for mutation research in DNA banks. Electrophoresis 20 (6):1250-1257.

24. Day IN et al (1998) Microplate-array diagonal-gel electrophoresis (MADGE) and melt-MADGE: tools for molecular-genetic epidemiology. Trends Biotechnol. 16 (7):287-290.

25. Day IN et al (1995b) Dried template DNA, dried PCR oligonucleotides and mailing in 96-well plates: LDL receptor gene mutation screening. Biotechniques 18 (6):981-984.

26. Durocher F et al (1996) Comparison of BRCA1 polymorphisms, rare sequence variants and/or missense mutations in unaffected and breast/ovarian cancer populations. Hum.Mol.Genet. 5 (6):835-842.

27. Easton DF, Ford D, and Bishop DT (1995) Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium Am.J.Hum.Genet. 56 (1):265-271.

28. Eccles DM et al (1998) BRCA1 mutations in southern England. Br.J.Cancer 77 (12):2199-2203.

29. El Harith e et al (2002) BRCA1 and BRCA2 mutations in breast cancer patients from Saudi Arabia. Saudi.Med.J. 23 (6):700-704.

30. Ellis CA and Clark G (2000) The importance of being K-Ras. Cell Signal. 12 (7):425-434.

31. Excoffier L and Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population Mol.Biol.Evol. 12 (5):921-927.

32. Fallin D and Schork NJ (2000) Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. Am.J.Hum.Genet. 67 (4):947-959.

33. Gambaro G, Anglani F, and D'Angelo A (2000) Association studies of genetic polymorphisms and complex disease. Lancet 355 (9200):308-311.

34. Gayther SA et al (1995) Germline mutations of the BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype-phenotype correlation. Nat.Genet 11 (4):428-433.

35. Ghaderi A et al (2001) Germline BRCA1 mutations in Iranian women with breast cancer. Cancer Lett. 165 (1):87-94.

36. Gilks WR, Richardson S, Spiegelhalter DJ. 1996. Markov chain Monte Carlo in practice. London: Chapman & Hall.

37. Goel V (2001) Appraising organised screening programmes for testing for genetic susceptibility to cancer BMJ 322 (7295):1174-1178.

38. Guo SW and Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48 (2):361-372.

39. Hamann U (2000) Hereditary breast cancer: high risk genes, genetic testing and clinical implications. Clin.Lab 46 (9-10):447-461.

40. Hoffman JD et al (1998) Implications of a novel cryptic splice site in the BRCA1 gene Am.J.Med.Genet. 80 (2):140-144.

41. Hogervorst FB et al (1995) Rapid detection of BRCA1 mutations by the protein truncation test. Nat.Genet. 10 (2):208-212.

42 Hogervorst FB et al (2003) Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. Cancer Res. 63 (7):1449-1453.

43. Huyton T et al (2000) The BRCA1 C-terminal domain: structure and function Mutat.Res. 460 (3-4):319-332.

44. Jensen DE and Rauscher FJ, III (1999) BAP1, a candidate tumor suppressor protein that interacts with BRCA1 Ann.N.Y.Acad.Sci. 886:191-194.

45. Ke SH and Wartell RM (1993) Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA; determination by temperature-gradient gel electrophoresis. Nucleic Acids Res. 21 (22):5137-5143.

46. Ke SH and Wartell RM (1995) Influence of neighboring base pairs on the stability of single base bulges and base pairs in a DNA fragmentBiochemistry 34 (14):4593-4600.

47. Knudson AG, Jr. (1971) Mutation and cancer: statistical study of retinoblastoma. Proc.Natl.Acad.Sci.U.S.A 68 (4):820-823.

48. Kozlowski P and Krzyzosiak WJ (2001) Combined SSCP/duplex analysis by capillary electrophoresis for more efficient mutation detection. Nucleic Acids Res. 29 (14):E71.

49. Krawczak M et al (2000) Human gene mutation database-a biomedical information and research resource. Hum.Mutat. 15 (1):45-51.

50. Krawczak M, Reiss J, and Cooper DN (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum.Genet. 90 (1-2):41-54.

51. Krontiris TG et al (1993) An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. N.Engl.J.Med. 329 (8):517-523.

52. Kwiatkowski D (2000) Genetic dissection of the molecular pathogenesis of severe infection. Intensive Care Med. 26 Suppl 1:S89-S97.

53. Kwok C et al (1996) Mutation analysis of the 2 kb 5' to SRY in XY females and XY intersex subjects. J.Med.Genet. 33 (6):465-468.

54. Lander ES et al (2001) Initial sequencing and analysis of the human genomeNature 409 (6822):860-921.

55. Lerman LS (1987) Detecting sequence changes in a gene. Somat.Cell Mol.Genet 13 (4):419-422.

56. Lerman LS and Silverstein K (1987) Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. Methods Enzymol. 155:482-501.

57. Levitt NC and Hickson ID (2002) Caretaker tumour suppressor genes that defend genome integrity Trends Mol.Med. 8 (4):179-186.

58. Li J et al (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 275 (5308):1943-1947.

59. Li Y et al (2002) Clustering of variations and haplotype analysis in the highly variable region of exon 11 of BRCA1 in Chinese women with sporadic breast cancer. Hum.Mutat. 20 (5):404-405.

60. Lorick KL et al (1999) RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc.Natl.Acad.Sci.U.S.A 96 (20):11364-11369.

61. Lyamichev VI, Panyutin IG, and Lyubchenko Y (1982) Gel electrophoresis of partially denatured DNA. Retardation effect: its analysis and application. Nucleic Acids Res. 10 (15):4813-4826.

62. Maniatis T and Tasic B (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418 (6894):236-243.

63. Markoff A et al (1998) Comparison of conformation-sensitive gel electrophoresis and single- strand conformation polymorphism analysis for detection of mutations in the BRCA1 gene using optimized conformation analysis protocols. Eur.J.Hum.Genet 6 (2):145-150.

64. Mazoyer S et al (1998) A BRCA1 nonsense mutation causes exon skipping. Am.J.Hum.Genet 62 (3):713-715.

65. Miki Y et al (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266 (5182):66-71.

66. Montagna M et al (1999) Identification of a 3 kb Alu-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families. Oncogene 18 (28):4160-4165.

67 Monteiro AN, August A, and Hanafusa H (1996) Evidence for a transcriptional activation function of BRCA1 C-terminal region. Proc.Natl.Acad.Sci.U.S.A 93 (24):13595-13599.

68. Moslehi R et al (2000) An unaffected individual from a breast/ovarian cancer family with germline mutations in both BRCA1 and BRCA2. Clin.Genet 57 (1):70-73.

69. Myers RM et al (1985a) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. Nucleic Acids Res. 13 (9):3131-3145.

70. Myers RM et al (1985b) Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. Nucleic Acids Res. 13 (9):3111-3129.

71. Myers RM, Maniatis T, and Lerman LS (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. Methods Enzymol. 155:501-527.

72. Nagamine CM, Chan K, and Lau YF (1989) A PCR artifact: generation of heteroduplexes. Am.J.Hum.Genet. 45 (2):337-339.

73. Nataraj AJ et al (1999) Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. Electrophoresis 20 (6):1177-1185.

74. Neuhausen SL (1999) Ethnic differences in cancer risk resulting from genetic variation Cancer 86 (11 Suppl):2575-2582.

75 Newton CR et al (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. 17 (7):2503-2516.

76. Nickerson DA et al (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat.Genet. 19 (3):233-240.

77. Nollau P and Wagener C (1997) Methods for detection of point mutations: performance and quality assessment. The IFCC Scientific Division, Committee on Molecular Biology Techniques. J.Int.Fed.Clin.Chem 9 (4):162-170.

78. Offit K (1998) the common hereditary cancer. clincal cancer genetics.Wiley-Liss.

79. Orita M et al (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc.Natl.Acad.Sci.U.S.A 86 (8):2766-2770.

80. Peelen T et al (1997) A high proportion of novel mutations in BRCA1 with strong founder effects among Dutch and Belgian hereditary breast and ovarian cancer families. Am.J.Hum.Genet 60 (5):1041-1049.

81 Perrin-Vidoz L et al (2002) The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. Hum.Mol.Genet. 11 (23):2805-2814

82. Puget N et al (1997) A 1-kb Alu-mediated germ-line deletion removing BRCA1 exon 17. Cancer Res 57 (5):828-831.

83. Rebeck GW et al (1999) Lack of independent associations of apolipoprotein E promoter and intron 1 polymorphisms with Alzheimer's disease. Neurosci.Lett. 272 (3):155-158.

84. Roest PA et al (1993) Protein truncation test (PTT) to rapidly screen the DMD gene for translation terminating mutations. Neuromuscul.Disord. 3 (5-6):391-394.

85. Sakorafas GH and Tsiotou AG (2000) Genetic predisposition to breast cancer: a surgical perspective. Br.J.Surg. 87 (2):149-162.

86. Satagopan JM et al (2001) The lifetime risks of breast cancer in Ashkenazi Jewish carriers of BRCA1 and BRCA2 mutations Cancer Epidemiol.Biomarkers Prev. 10 (5):467-473.

87. Schneider S, Roessli D, Excoffier L. 2000. Arlequin, version 2000.Genetics and Biometry Laboratory, University of Geneva, Switzerland.

88. Schork NJ (1997) Genetics of complex disease: approaches, problems, and solutions Am.J.Respir.Crit Care Med. 156 (4 Pt 2):S103-S109.

89. Schork NJ, Fallin D, and Lanchbury JS (2000) Single nucleotide polymorphisms and the future of genetic epidemiology. Clin.Genet 58 (4):250-264.

90. Schoumacher F et al (2001) BRCA1/2 mutations in Swiss patients with familial or early-onset breast and ovarian cancer. Swiss.Med.Wkly. 131 (15-16):223-226.

91. Schwabe JW and Klug A (1994) Zinc mining for protein domains. Nat.Struct.Biol. 1 (6):345-349.

92. Seitz S et al (1997) Strong indication for a breast cancer susceptibility gene on chromosome 8p12-p22: linkage analysis in German breast cancer families. Oncogene 14 (6):741-743.

93. Sevilla C et al (2002) Testing for BRCA1 mutations: a cost-effectiveness analysis. Eur.J.Hum.Genet. 10 (10):599-606.

94. Shapiro MB and Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15 (17):7155-7174.

95. Sheffield VC et al (1989) Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. Proc.Natl.Acad.Sci.U.S.A 86 (1):232-236.

96 Slater HR et al (2003) Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). J.Med.Genet. 40 (12):907-912.

97. Smith SA et al (2001) BRCA1 germline mutations and polymorphisms in a clinic-based series of ovarian cancer cases: a Gynecologic Oncology Group study. Gynecol.Oncol. 83 (3):586-592.

98. Smith TM et al (1996) Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. Genome Res 6 (11):1029-1049.

99. Somasundaram K (2003) Breast cancer gene 1 (BRCA1): Role in cell cycle regulation and DNA repair-perhaps through transcription. J.Cell Biochem. 88 (6):1084-1091.

100. Somasundaram K et al (1997) Arrest of the cell cycle by the tumour-suppressor BRCA1 requires the CDK-inhibitor p21WAF1/CiP1. Nature 389 (6647):187-190.

101. Spanakis E., Al-dahmesh M.A., and Day INM (2001) Scanning for point variation in large populations by melt-MADGE. Technical Tips Online (1):117.

102. Spanakis E, Day IN (1997) The molecular basis of genetic variation: mutation detection methodologies and limitations. In Day IN, Humphries SE (eds) Genetics of common diseases. Bios Scientific Publishers.

103. Stephens M, Smith NJ, and Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am.J.Hum.Genet. 68 (4):978-989.

104. Struewing JP et al (1997) The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N.Engl.J.Med. 336 (20):1401-1408.

105. Szabo CI and King MC (1997) Population genetics of BRCA1 and BRCA2. Am.J.Hum.Genet 60 (5):1013-1020.

106. Thorlacius S et al (1998) Population-based study of risk of breast cancer in carriers of BRCA2 mutation. Lancet 352 (9137):1337-1339.

107. Thornton K et al (1999) Purification, characterization, and crystallization of the distal BRCT domain of the human XRCC1 DNA repair protein. Protein Expr.Purif. 16 (2):236-242.

108. Tishkoff SA et al (2000) The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus. Am.J.Hum.Genet. 67 (2):518-522.

109. Trinh TQ and Sinden RR (1991) Preferential DNA secondary structure mutagenesis in the lagging strand of replication in E. coli. Nature 352 (6335):544-547.

110. Vallon-Christersson J et al (2001) Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer familiesHum.Mol.Genet. 10 (4):353-360.

111. Van Orsouw NJ et al (1999) A highly accurate, low cost test for BRCA1 mutations. J.Med.Genet 36 (10):747-753.

112. Vijg J and van Orsouw NJ (1999) Two-dimensional gene scanning: exploring human genetic variability. Electrophoresis 20 (6):1239-1249.

113. Vineis P, Schulte P, and McMichael AJ (2001) Misconceptions about the use of genetic tests in populations. Lancet 357 (9257):709-712.

114. Wagner T et al (1999) Denaturing high-performance liquid chromatography detects reliably BRCA1 and BRCA2 mutations. Genomics 62 (3):369-376.

115. Wang L and Freedman SD (2002) Laboratory tests for the diagnosis of cystic fibrosis m.J.Clin.Pathol. 117 Suppl:S109-S115.

116. Wang WW et al (2001) A single nucleotide polymorphism in the 5' unranslated region of RAD51 and risk of cancer among BRCA1/2 mutation carriers. Cancer Epidemiol.Biomarkers Prev. 10 (9):955-960.

117. Welcsh PL, Owens KN, and King MC (2000) Insights into the functions of BRCA1 and BRCA2. Trends Genet 16 (2):69-74.

118. Williams RS, Green R, and Glover JN (2001) Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. Nat.Struct.Biol. 8 (10):838-842.

119. Wilson JMG. and Jungner G (1968) Principles and practice of screening for disease. Public Health Paper Number 34.Geneva: WHO.

120. Wooster R et al (1995) Identification of the breast cancer susceptibility gene BRCA2. Nature 378 (6559):789-792.

121. Xu CF and Solomon E (1996) Mutations of the BRCA1 gene in human cancer. Semin.Cancer Biol. 7 (1):33-40.

122 Ye S et al (2001) An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res. 29 (17):E88

123. Zhong Q et al (1999) Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. Science 285 (5428):747-750.