UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

SCHOOL OF MEDICINE

Division of Cancer Sciences

The Role of Genetic Factors in Breast Cancer Aetiology

By

Victoria Naomi Hammond

Thesis for degree of Doctor of Philosophy

June 2010

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

SCHOOL OF MEDICINE

Doctor of Philosophy

THE ROLE OF GENETIC FACTORS IN BREAST CANCER AETIOLOGY

By Victoria Naomi Hammond

Breast cancer is the most common cancer in women and is also the leading cause of cancer mortality in women. There are several known risk factors for breast cancer including genetic factors which account for at least 25% of the incidence of breast cancer, although only a small proportion of this is a result of mutations in known high penetrance susceptibility genes. The majority of genetic risk is now thought to be due to common genetic variants, for example single nucleotide polymorphisms (SNPs). We investigated whether SNPs in candidate genes, with a biological reason for being of interest to study in relation to breast cancer, were correlated with the development of tumours with a certain phenotype, such as grade, lymph node involvement, oestrogen receptor status and the presence of distant metastases.

We genotyped 206 SNPs across 30 candidate genes in 1001 patients. Association was performed using Cochran-Armitage trend test and 2-by-3 tables of disease by genotype**.**

We replicated observations from previous studies such as the association of SNPs in FGFR2, TNRC9 and ATM with oestrogen receptor status and identified novel associations of SNPs in the oestrogen receptor gene and matrix metalloproteinase-9 gene (MMP-9) with grade and presence of distant metastasis respectively.

The function of two promoter SNPs in MMP-9 were further investigated using luciferase reporter gene assays. The C allele of rs3918242 had a 1.5 fold increase in MMP-9 expression in MDA-MB-231 cells and the A allele of rs3918241 showed a slight increase in MMP-9 expression in MCF-7 and NIH-3T3 cell lines although not significant.

The novel results identified need to be replicated for validation but this study provides evidence that common genetic variants play a role in predisposing to certain tumour types.

List of Contents

List of Tables

List of Figures

Author's Declaration

I, the undersigned, Victoria Naomi Hammond, declare that the thesis entitled 'The Role of Genetic Factors in Breast Cancer Aetiology' and the work presented herein is entirely my own and reference to or quotation from the work of any other person has been correctly acknowledged within the work.

Part of this work is published as:

William Tapper, Victoria Hammond, Sue Gerty, Sarah Ennis, Peter Simmonds, Andrew Collins, the Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) Steering Group and Diana Eccles. The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. *Breast Cancer Research* 2008, **10**:R108

Signed……

Dated……..

Acknowledgements

I would like to thank my principal supervisor Professor Diana Eccles for her continued support and guidance throughout my research.

Thanks should also be attributed to Professor Andy Collins, my second supervisor, and Dr Will Tapper from the Department of Human Genetics, University of Southampton, for their assistance with bioinformatics.

I would also like to thank Dr Jeremy Blaydes from the Cancer Sciences Division, University of Southampton, for his guidance with the functional studies.

Finally, thank you to Dr Jon Strefford for his continued support and enthusiasm and help with proof reading my thesis.

This work was funded by Breast Cancer Research Trust.

Definitions

Aetiology	The cause of a disease
Allele	An allele is an alternative form of a gene (one member of a pair) that is located at a specific position on a specific chromosome
Autosomal	Pertaining to a chromosome that is not a sex chromosome
Autosomal dominant	Autosomal dominant is one of several ways that a trait or disorder can be passed down through families. If a disease is autosomal dominant, it means you only need to get the abnormal gene from one parent in order for you to inherit the disease.
Basement membrane	Membrane separating the organ parenchyma from the underlying stroma
Carcinogenesis	The process by which normal cells are transformed into cancer cells
Endothelial cells	Cells that form the inner lining of a blood vessel
Enzyme	large protein molecules capable of affecting the speed of a reaction without being altered thenselves
Genetic marker	A gene or DNA sequence with a known location on a chromosome
Genome	The entirety of an organism's hereditary information
Genotype	The genetic constitution of a cell, an organism, or an individual (the specific allele makeup of the individual)
Haplotype	Combination of alleles at multiple loci that are transmitted together on the same chromosome
Heterogeneity	In genetics, heterogeneity refers to multiple origins causing the same disorder in different individuals
Heterozygous	Having two different alleles for a single trait
Homozygous	Having identical alleles for a single trait
In silico	An expression used to mean "performed on computer or via computer simulation"
Incidence	The number of new cases occurring, expressed as an absolute number of cases per year or as a rate per 100,000 persons per year

Linkage	The tendency of certain loci or alleles to be inherited together
Metastasis	Spread of a disease from one organ or part to another non-adjacent organ or part
Oncogene	A gene that, when mutated or expressed at high levels, helps turn a normal cell into a tumour cell
Penetrance	The proportion of individuals carrying a particular variation of a gene (allele or genotype) that also express an associated trait (phenotype)
Phenotype	Any observable characteristic or trait of an organism: such as its morphology, development, biochemical or physiological properties, or behavior
Polymorphism	Multiple alleles of a gene within a population, usually expressing different phenotypes
Prevalence	The number of persons alive at a particular point in time with the disease of interest
Proliferation	An increase in the number of cells as a result of cell growth and cell division.
Proto-oncogene	A normal gene that can become an oncogene due to mutations or increased expression
Single nucleotide polymorphism	DNA sequence variation occurring when a single nucleotide — A, T, C, or G — in the genome (or other shared sequence) differs between members of a species
Somatic	Refers to cells of the body, rather than gametes
Sporadic	When a genetic disease occurs without any family history or genetic defects in the parents, the disease is called a sporadic genetic disease
Stroma	Structural tissue of organs, namely the connective tissues. In cancer, stroma refers to surrounding connective tissue and associated cells that supports it - connective, functionally supportive framework of a biological cell, tissue or organ.
Transfection	The process of deliberately introducing nucleic acids into cells

Transformation	The genetic alteration of a cell resulting from the uptake, genomic incorporation, and expression of environmental genetic material (DNA), occurring ansformation occurring most commonly in bacteria
Wild type	The phenotype of the typical form of a species as it occurs in nature

Abbreviations

ADH1B1	Alcohol dehydrogenase 1B (class 1, beta polypeptide)
AR	Androgen receptor
ATM	Ataxia telangiectasia mutated gene
BRCA1/2	Breast cancer susceptibility gene 1/2
CASP8	Caspase 8
CGAS	Candidate gene association study
CYP1B1	Cytochrome P450, subfamily 1, polypeptide 1
DAPK1	Death associated protein kinase
DDFS	Distant disease-free survival
dH2O	Distilled water
ddH2O	Double distilled water
ER	Oestrogen receptor
ERBB2	(Her-2/neu) Epidermal growth factor receptor B2
ESR1	Oestrogen receptor 1 (encoding oestrogen receptor α)
FGFR2	Fibroblast growth factor receptor 2
FN1	Fibronectin 1
GATA3	GATA binding protein 3
GWAS	Genome wide association study
IGF-1	Insulin-like growth factor 1
IL-6	Interleukin-6
ITGB4	Integrin beta 4 subunit
LB broth	Luria-Bertani broth
LN	Lymph node
LSP1	Lymphocyte specific protein
Maf	Minor allele frequency

MAP3K1	Mitogen-activated protein kinase kinase kinase 1
MDM2	Mouse double minute 2 homolog
MMP	Matrix metalloproteinase
NMSC	Non-melanoma skin cancer
PGR	Progesterone receptor
POSH	Prospective study of Outcomes in Sporadic versus Hereditary breast cancer
RHOC	Ras homolog gene family, member C
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
TGF-β	Transforming growth factor β
TNF-α	Tumour necrosis factor α
TOX3	TOX high mobility group box family member 3
TP53	Tumour protein 53 (p53) gene
TWIST1	Twist transcription factor
VEGF	Vascular endothelial growth factor

Chapter 1 Introduction

1.1 Research Framework

This study investigates whether genetic variations in certain genes are associated with tumours with a specific phenotype (grade and oestrogen receptor status) or propensity to metastasise (lymphovascular invasion, presence of lymph node involvement and distant metastasis).

Introduction: Breast cancer is the most common cancer in women. There are many risk factors for breast cancer, including genetic factors which account for 25-30% of the incidence. Only 15-30% of the heritable component of breast cancer is due to known highly penetrant genes such as BRCA1 and BRCA2. A significant proportion of genetic factors remain undetermined. Recent Genome Wide Association Studies (GWAS) have confirmed the existence of several common genetic variants affecting breast cancer risk. It has also been shown that some low penetrance variants may increase the risk of a specific tumour phenotype. Genetic variants may have an impact on breast cancer prognosis, currently routinely predicted on the basis of tumour characteristics and stage, either by influencing the development of specific breast cancer subtype or by influencing the host tumour defence mechanisms. We investigated whether there was evidence that single nucleotide polymorphisms (SNPs), the most common type of genetic variant, in candidate genes might influence breast cancer phenotype and prognosis.

Method: We genotyped blood DNA from 1001 young onset symptomatic breast cancer cases from the POSH study. 206 tagging SNPs were typed reporting on 30 candidate genes. We used Cochrane Armitage Trend test to compare phenotypic extremes (for example ER+ve vs ER–ve and Grade 1 vs Grade 3) and Kaplan Meier survival analysis to compare relapse free survival according to genotypic subgroups. The functional consequences of the potentially disease influencing SNPs located in the promoter regions of the MMP-9 gene were further investigated using Luciferase reporter assays.

Results: Replicating observations from recently published studies, SNPs in FGFR2 (p=0.000003) and TOX3 (p=0.0014) were associated with ER+ve breast cancer, whereas ATM SNPs were associated with ER–ve disease (p= 0.0000985). SNPs associated with known prognostic tumour features included two adjacent SNPs in the oestrogen receptor alpha gene (tumour grade, p=0.0026 and p=0.0040) and three SNPs in the MMP-9 gene (distant metastasis, p= 0.00092,

p=0.0034, p=0.0037 (p=0.000066, p=0.0.00044, p=0.00038 in ER+ patients)). The wild type (C) allele of rs3918242 was found to have a 1.5 fold increase in MMP-9 expression over the minor allele in highly metastatic MDA-MB-231 cells, and the minor allele (A) of rs3918241 showed a slight increase in MMP-9 expression in oestrogen receptor positive MCF-7 cells and NIH-3T3 fibroblasts.

Conclusion: These novel findings will need to be replicated in further studies but suggest inherited genetic variants may be either modulating the host response to malignant cell growth or predisposing to certain types of somatic mutation that drive malignant cell proliferation.

1.2 Background

1.2.1 Introduction to Cancer

Cancer is a term that refers to a collection of diseases that have the common feature of uncontrolled cell growth and invasion. Cancer is one of the leading health problems worldwide[1], in the year 2000 there were an estimated 22 million people living with cancer and 10 million new cases were diagnosed in that year[2]. In the UK, one in three people will develop cancer during their lives and one in four people will die from some form of cancer[3]. It is predominantly a disease of older people with approximately 25% of cases occurring in people over 60 years of age[4].

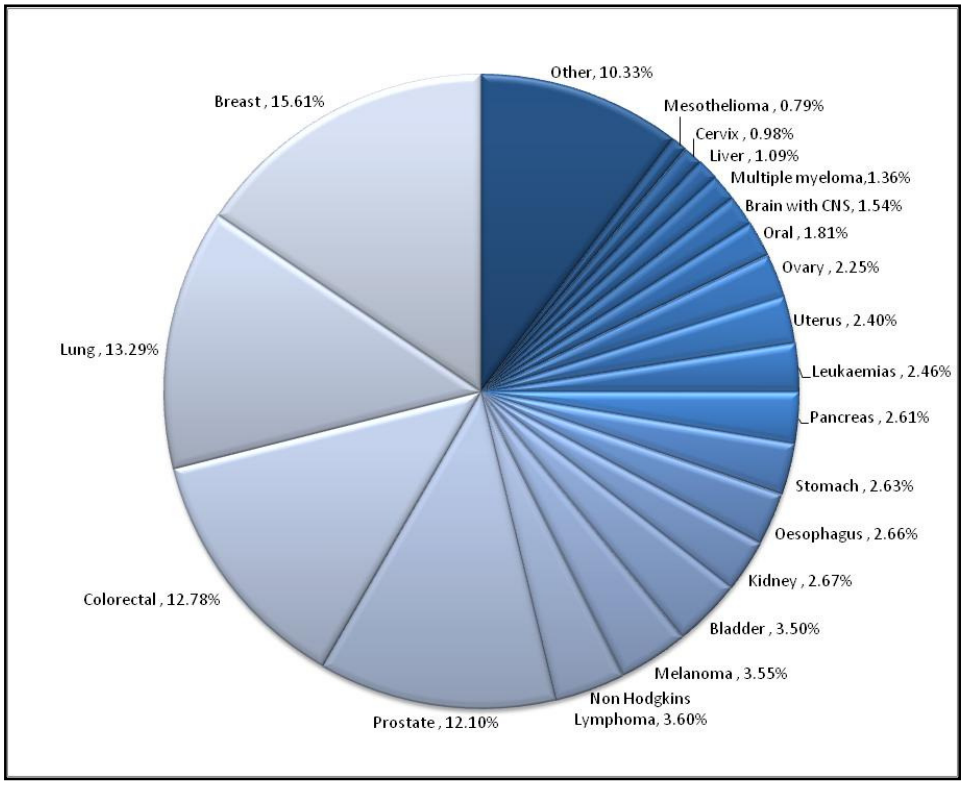


Figure 1: The 20 most commonly diagnosed cancers (excluding non-melanoma skin cancers, NMSC) UK, 2006. Data from CRUK[1]

In western countries, breast and prostate cancers are the most common, whereas cancers of the stomach and cervix are far more prevalent in developing countries. Lung cancers are also more common in the Western countries[2]. In the UK four types of cancer; breast, lung, colorectal and prostate, account for over half of all cancer cases. All cancer excludes

non-melanoma skin cancers because of the difficulties of measurement and consequent lack of data (Figure 1).

There are many possible causes of cancer, for example exposure to carcinogens, either in the external environment or through lifestyle choices such as smoking, is a major influence. The impact of lifestyle and environmental factors are shown by the prevalence of certain cancers in different countries and the results of migration studies showing that breast cancer rates change when women move to a new country, providing evidence for the importance of lifestyle and environment in breast cancer risk. [2;5;6].

It was first suggested in 1914 by the German scientist Theodor Boveri, prompted by previous observations of aberrant mitoses by David von Hansemann[3], that there is a genetic component to cancer in his work entitled 'Zur Frage der Entstehung Maligner Tumoren' (The origin of malignant tumours) which was recently translated and republished [7]. Boveri noted that there were chromosomal abnormalities in cancer cells and he proposed that the chromosomes are carriers of genetic information and that predisposition to cancer was inherited by inheriting a copy of a chromosome unable to sufficiently suppress tumour growth.

More recent studies have confirmed Boveri's predictions. In 1971 Knudson performed statistical analysis on cases of retinoblastoma, a tumour of the retina, and he noted that sporadic cases affected older patients and were invariably unifocal, whereas the inherited form of the disease affected patients at younger ages than average and in the majority of cases they develop more than one tumour. Based on these observations Knudson proposed a 2-step model of cancer[8]. He assumed that most cancers arose from a single cell and at least two mutational events are required for carcinogenesis (affecting both copies of a gene) and suggested that cancers could be classified into two categories, sporadic and inherited. In sporadic cases two somatic mutations, occurring by chance, were required, however in inherited cases, one germline copy of a damaged gene is inherited so only one somatic mutation was required to the good copy of the gene which

would rapidly lead to cancer explaining the younger age of onset in patients with the inherited form of the disease.

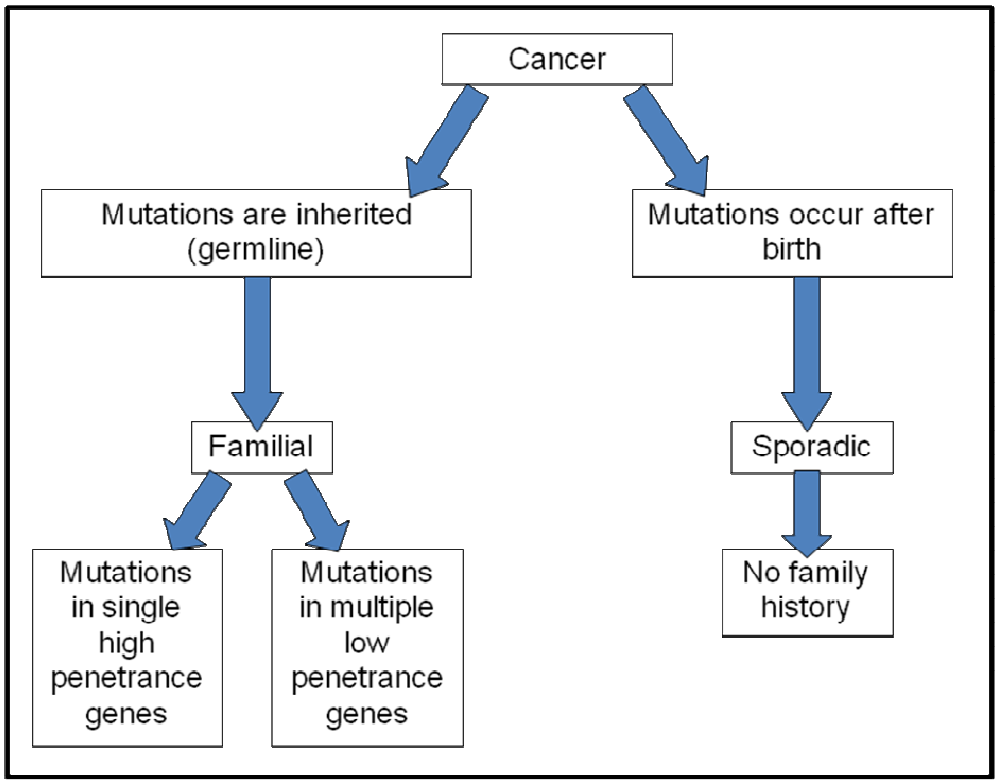


Figure 2: Cancers are classified as familial or sporadic. Sporadic cancers are characterised by a lack of family history and arise from somatic mutations occurring after birth, whereas inherited germline mutations contribute to familial cancer. Adapted from Slijepcevic 2007[9].

It is now accepted that the process of carcinogenesis is a multi-step process resulting from the accumulation of errors in regulatory pathways in the cell affecting proliferation, survival, apoptosis and cell motility. It is thought that at least three to six mutations in key genes are needed for a cell to overcome all the normal cellular restraints and become cancerous[10]. However, in reality, most adult cancers are very complex and will also contain many acquired, but not critical, mutations; these are sometimes referred to as driver (critical) and passenger (background mutation 'noise') mutations. Inherited mutations may be a single mutation in a high penetrance gene or multiple mutations in low penetrance genes (Section 1.2.2.6). Somatic mutations that occur may, for example, enable the invasion of neighbouring tissue, evasion of immune system detection,

recruitment of a new blood supply, dissemination and targeting of new sites, and the penetration and reinvasion through new blood and tissue layers. Over time successful metastasis occurs.

1.2.2 Introduction to Genetics

The process of tumourigenesis is multifaceted and the role of genetics in the process is complex. This section explains the basic terminology that is of relevance in this study.

1.2.2.1 DNA

The biological information an organism needs to reproduce itself is contained in its DNA.

DNA is a polymer containing chains of nucleotide monomers. Each nucleotide contains a sugar,2'-deoxyribose, one to three phosphate groups and a base. There are four types of base, the purines adenine (A) and guanine (G) and the pyrimidines thymine (T) and cytosine (C). Nucleotides occur as individual molecules or polymerised as nucleic acids DNA or RNA. The sequence of bases encodes the genetic information.

Two polynucleotide strands wrap around each other forming a DNA molecule known as a double helix, with the sugar phosphate part of the molecule forming a backbone and the bases facing inwards. Hydrogen bonds form between the bases on the two strands, stabilising the double helix. Due to the space that is available between the two strands in the helix, the pairing of bases is restricted such that purines always interact with pyrimidines, therefore A pairs only with T and G only with C, so that both strands of the helix are related to each other with one sequence determining and predicting the sequence of the other. This is known as complementary base pairing and ensures the genetic information is preserved during replication. The two polynucleotide strands run in opposite directions. The double helix is right-handed, with a turn every 10 bases. The major groove of the helix interacts with proteins.

The human genome (the total human DNA sequence) contains over 3 billion DNA base pairs.

1.2.2.2 Genes

The DNA in an organism is organised into a large number of genes. Genes are discrete segments of DNA and each gene codes for the amino acid sequence of a polypeptide. Genes can vary in size from less than 100 base pairs up to several million base pairs.

Genes are very dispersed, being separated by non-coding DNA sequences which can be very long so that the coding gene sequences account for a very small percentage of the genome. The human genome contains an estimated 20,000–30,000 protein-coding genes which account for only 1.5% of the genome.

Only one of the two strands of DNA carries biological information. This strand is called the template strand (also known as the antisense or non-coding strand) and is used to direct synthesis of a complementary RNA molecule (pre-mRNA) in a process known as transcription. The other strand is known as the non-template strand (sense/coding). Both strands can act as the template strand and individual genes may be encoded on different strands.

Genes are composed of both coding (exons) and non coding (introns) regions of DNA. The number of introns in a gene can vary greatly, from 0 to 50 or more. The length of both the introns and exons vary greatly, but the introns are usually far longer, accounting for the majority of the gene sequence. For the biological information contained in a gene to be used to synthesise a protein, the introns must be removed from the pre-mRNA molecules synthesised from the template strand of DNA. This process is known as splicing and results in the formation of a mature mRNA molecule which is then exported from the nucleus to cytoplasmic organelles known as ribosomes where it acts as a template for protein synthesis (translation). The amino acid (building blocks of protein) sequence of the protein determines its three-dimensional structure which in turn dictates its function.

Expression of the biological information contained in the genes is highly regulated and the human genome has many different regulatory sequences which are crucial to controlling gene expression. Not all of the genes that are present in a cell are active and different

genes are expressed in different cell types. This differential expression determines the characteristics of a cell and the role it plays within an organism. A segment of DNA upstream of the coding region of the gene, known as the promoter, regulates synthesis of mRNA from the gene and therefore expression of the gene. The promoter region may extend up to about 1kbp (1000 base pairs) upstream of the gene. Conserved regions in the promoter are recognised and bound to by transcription factors and RNA polymerase which initiate the synthesis of a gene.

There are also other regulatory sequences that influence the transcription of a gene and may occur at sites much further away than the promoter, for example enhancers, which turn on and speed up the process of transcription, silencers, which slow or stop transcription and insulators, which protect genes from the effects of silencers and enhancers.

The majority of the genome is made up of extragenic DNA sequences, this is all of the DNA sequence apart from the genes and gene-related sequences (introns, exons, pseudogenes and gene fragments). Extragenic DNA sequences can be unique or exist only as a small number of copies which are known as interspersed genome-wide repeats, for example short interspersed nuclear elements (SINEs) and long interspersed nuclear elements, (LINEs), but some may contain long stretches of repeat elements known as tandemly repeating DNA and includes satellite DNA, minisatellites (variable number tandem repeats) and microsatellites (simple tandem repeats). Some of these sequences are involved in the regulation of gene expression, some act as spacers and other regions have functions as yet undiscovered.

1.2.2.3 Chromosomes

The human genome is located in the nucleus of every cell in the body (except for red blood cells). It is packaged into organized thread-like structures called chromosomes. Each of the chromosomes consists of DNA that is tightly coiled around proteins, known as histones, that maintain the structure. The DNA and histones are then wound round and round itself in supercoils and this dense DNA is known as chromatin. Each chromosome has a constriction point called the centromere which separates it into two arms; the shorter arm, above the centromere, is the p arm and the longer arm, below the centromere, is the q arm. The location of the centromere on each chromosome gives the chromosome its characteristic shape and can be used to describe the location of specific genes.

1.2.2.4 Genotypes and Phenotypes

A genotype is the genetic makeup of a person, whereas phenotype refers to the physical manifestation of an inherited trait or disease. In cancer, both genotype and phenotype keep changing over time, contributing to the heterogeneity of the disease.

1.2.2.5 Alleles

The human genome is stored on 23 chromosomes, 22 autosomal chromosomes and one that is sex-determining. Each individual has two copies of each chromosome (46 chromosomes in total, 23 chromosomal pairs), one is maternally inherited and one paternally inherited, giving each individual two copies of each gene. There can be more than one form of a gene (allele) at a given locus and in humans two alleles for each gene will be inherited, one from each parent. Paired alleles (one on each of two paired chromosomes) that are the same are called homozygous and those that are different are called heterozygous. Complex traits such as height and longevity are usually caused by the interactions of numerous pairs of alleles, while simple traits such as eye colour may be caused by just one pair.

1.2.2.6 Penetrance

Penetrance is a measure of the probability of a gene or genetic trait (phenotype) being expressed. It is based on the proportion of individuals with a mutation causing a particular disorder who exhibit clinical symptoms of that disorder.

A condition is said to have complete penetrance if clinical symptoms are present in all the population who have the disease causing mutations (usually inherited in an autosomal dominant fashion).

A condition is said to have reduced penetrance, or incomplete penetrance, if the clinical symptoms are not always present in all the population who have the disease causing mutations.

Reduced penetrance is common in many familial cancers such as hereditary nonpolyposis colorectal cancer (HNPCC) and breast cancer. HNPCC results from germline mutations in mismatch repair genes, however 20% of the carriers of these mutations will not develop the disease[11]. Mutations in the high penetrance breast cancer susceptibility genes BRCA1 and BRCA2 confer a lifetime risk of approximately 80% of developing breast cancer[12;13], which is five to ten times the risk in the control female population[14]. It is not possible to predict which people with the mutations will get the disease or when it may develop. Reduced penetrance is likely to result from a combination of genetic, environmental and lifestyle factors, many of which are still unknown.

Penetrance is an age related trait, as traits not normally expressed at birth occur with an increased frequency as a carrier ages, for example mutations in mismatch repair genes associated with hereditary nonpolyposis colorectal cancer are incompletely penetrant and therefore not all people carrying these mutations will develop the disease, however the risk of disease increases with age[15].

Penetrance can be affected by carcinogens, hormones, mutations in DNA damage response genes and modifier genes, which may affect the expression of some alleles that may increase or decrease penetrance of a germline mutation, such as an altered susceptibility allele.

1.2.2.7 SNPs

The DNA sequence in each individual is 99.9% the same. The remaining 0.01% of the human DNA sequence differs between individuals. The greatest type of genetic variation is a single nucleotide polymorphism (SNP, pronounced 'snip'). SNPs are certain sites within a human genome at which some individuals will have one nucleotide present while other individuals will have a different one. These single base substitutions occur, on average, once in every 1200 bases. SNPs are point mutations that become established in a population over time. In order to be called a SNP, the single base substitution must be present in over 1% of a large population. It is estimated that there are at least 10 million

common SNPs in the human genome. Each base is known as an allele and collections of alleles make up a persons genotype.

SNPs can be used as markers to help locate genes in DNA sequences. For example, if a large group of patients affected with a disease, such as breast cancer, is compared with the same number of healthy individuals and if a SNP is identified where one particular allele dominates in the disease group at that specific locus then that allele at that location is associated with the disease. This association does not necessarily imply causation as large regions of genetic information are inherited together in sections known as haplotype blocks. This means that all individuals with the same SNP at a particular locus will also have identical variants at all the other SNPs in the same haplotype block. Therefore, when trying to identify genes involved with a certain disease, only one SNP, a tagging SNP, needs to be used in each haplotype block. It is estimated that 300,000 to 600,000 tagging SNPs will contain most of the information on genetic variation in the human genome. Information from the Human Genome Project[16] can be used to identify the location of the SNP in the genome, which may be within or near to a specific gene.

1.2.2.8 Linkage Disequilibrium

In population genetics linkage equilibrium occurs where the genotype present at one locus is independent of the genotype at a second locus. If, however, two alleles are not independent of one another and occur more or less frequently in a population than would be expected from a random formation of haplotypes from alleles based on their frequencies they are said to be in linkage disequilibrium. Non-random associations between polymorphisms at different loci are measured by the degree of linkage disequilibrium (LD), that is, the degree of deviation from the expectation of non-association along a genome.

1.2.3 Genetic Association Studies

Researchers have spent decades trying to identify genes involved in breast cancer using different types of association studies. The first breast cancer susceptibility gene BRCA1 was mapped and cloned in the early 90s in a family-based linkage study and several further high penetrance susceptibility genes have since been identified in this way. In the past few years, several new loci associated with various degrees of breast cancer risk have been identified using candidate gene association studies (CGAS) and genome wide association studies (GWAS).

Genetic association studies look at a group of people with a particular disease (cases) and a group without the disease (controls) to see if a certain genotype (single-locus alleles or multilocus haplotype) is more frequent in the cases than the controls. If the allele plays a causative role in the disease or is correlated with a causal allele it will have a higher frequency in the case population than the control population.

Genetic association studies may take one of two approaches, either family-based or population based.

1.2.3.1 Family-based Association Studies – Linkage Studies

Linkage is the tendancy for genes and other genetic markers to be inherited together due to their proximity on the same chromosome and the genetic markers can be be of use in studying diseases running in families. Identifying genetic markers that run in a family, in the same way as the disease, can locate an area of the genome in which the causative gene is assumed to be. It can be difficult to genotype several generations of a family, so a common approach is to genotype affected siblings.

Linkage studies identify large chromosomal regions (millions of base pairs long) and so the resolution is low and there could be many genes in the area that need to be tested. The strongest linkage signals come from recessive and highly penetrant and therefore generally rare diseases.

The high penetrance breast cancer susceptibility genes BRCA1[17;18] and BRCA2[19] were identified using family-based linkage studies.

1.2.3.2 Population-based Association Studies - GWAS and CGAS

The classical linkage studies identify high penetrance gene mutations linked to disease in affected families, however, mutations in low penetrance genes cannot be identified in this way as linkage studies lack power to detect the alleles responsible for low to moderate risk of developing breast cancer.

These alleles are now being detected by population based gene association studies. These studies use thousands of known SNPs present in the human genome, originally focusing on SNPs in candidate genes. The disadvantage of such candidate gene association studies (CGAS) is the need to guess the gene in advance and type the causative SNP (or a tightly linked SNP in strong linkage disequilibrium), however these studies can be used to look for association of alleles with a disease in nominated candidate genes from linkage studies. This has a higher resolution than linkage studies.

As it becomes increasingly more evident that many diseases result from polygenic effects of multiple low penetrance susceptibility alleles, there is a need for larger and larger association studies in order to have sufficient power to detect these loci. Many technological advances such as large scale genomic projects including the sequencing of the human genome (The Human Genome Project[16]) and the HapMap project[20] as well as new genotyping technology has allowed association to move from candidate gene studies to unbiased whole genome searches (Genome Wide Association Studies, GWAS) identifying moderate risk alleles without prior knowledge of position or function[21] making linkage scans unnecessary.

CGAS and GWAS clearly have the power to identify common variants that are associated with low susceptibility loci for a disease although those susceptibility genes conferring very low risks may still be missed[22]. Also, a great degree of caution is needed to interpret the results of such studies correctly, considering factors such as sample size, genotyping

quality controls and successful replication of results[23]. Case control studies are always subject to problems such as confounding due to population stratification which is not an issue in family based studies.

The lack of success so far with association studies may be due to small sample sizes not having sufficient power to detect associations, or it may be that the studies have not looked at the right markers and there are important genes that are yet to be examined[22].

The low penetrance susceptibility alleles are common in the general population and each confer only a very small overall risk (generally 1.3 fold or less[24]) precluding the use of individual susceptibility alleles in genetic counselling[14]. However, the combined effects may be sufficiently large to be useful for risk prediction, as well as targeted screening and prevention, particularly as more loci are identified.[24]

1.2.4 Breast Cancer Genetics

Breast cancer is the most common cancer in the UK even though it is very rare in men (Figure 3). In 2006, 45,822 people in the UK were diagnosed with breast cancer, 314 of which were men. Breast cancer rates are increasing and it has become one of the biggest causes of death from cancer in women. In 2007, 12,082 women died from breast cancer. It has been calculated that there is a lifetime risk of one in nine for women in the UK developing breast cancer [1].

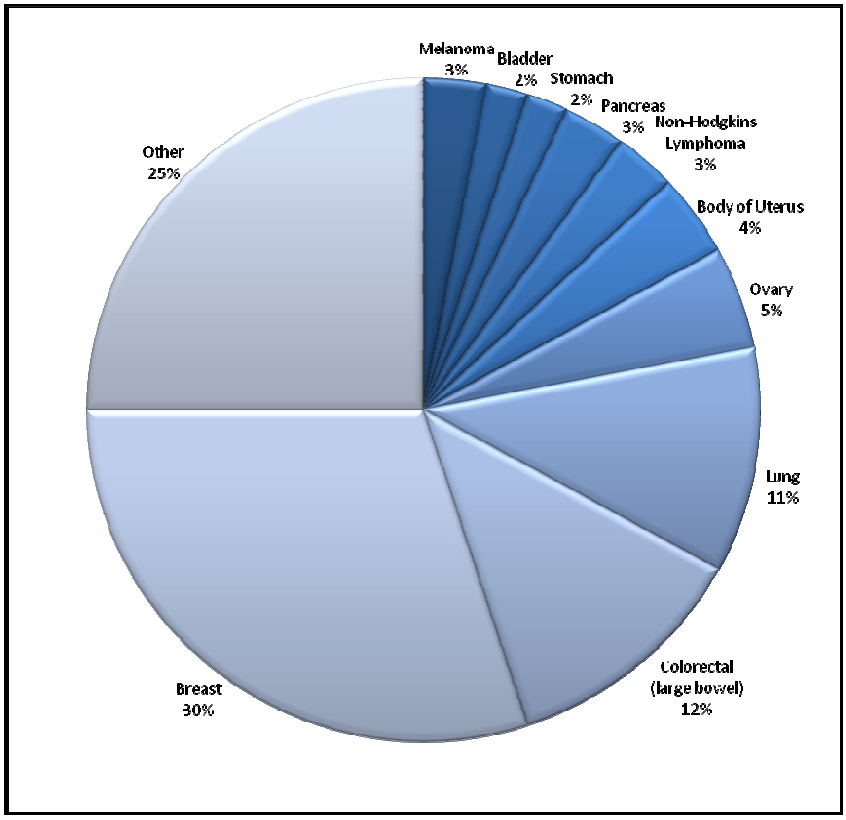


Figure 3: The most common cancers diagnosed in women in the UK in 2001 out of a total of 136,153 cases, excluding NMSC. Adapted from ABPI[25]

Breast cancer is the most common cancer in women worldwide with approximately one million new cases in the world each year[2;5], accounting for 22% of all cancers[6]. The rates of breast cancer vary in different countries, being less common among women in developing countries, with low incidence rates in most African and Asian populations. More than half of the breast cancer cases are in industrialised countries with the highest

incidence rates occurring in Europe, North America, Australia and New Zealand[2;5](Figure 4).

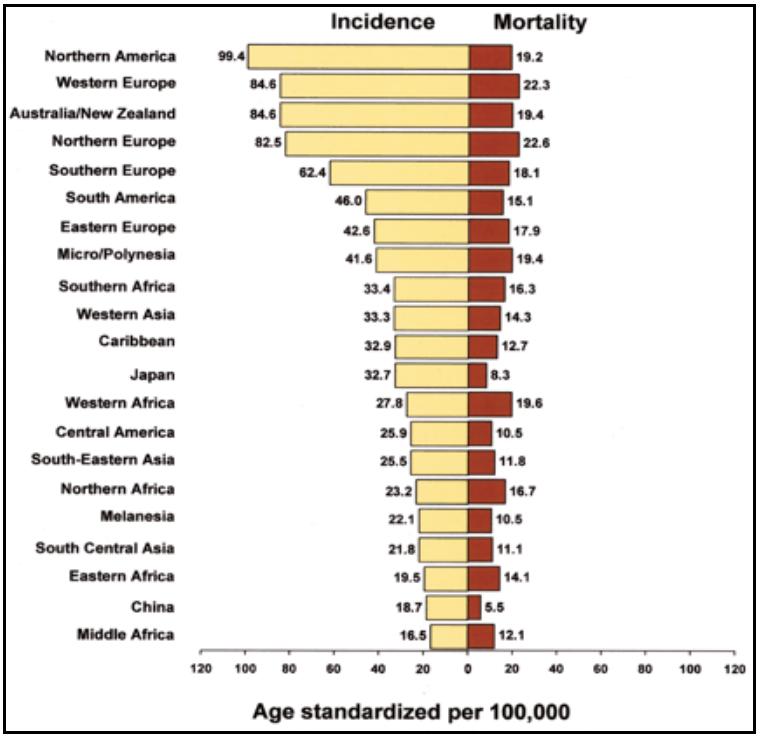


Figure 4: Age standardised incidence and mortality rates for breast cancer. Data shown per 100,000[5]

There is a striking variation in the risk of different cancers by geographic area. The high incidence of breast cancer in developed countries may be largely due to the screening programs that can detect early invasive cancers that may otherwise have been diagnosed later if at all. However some of this variation may be explained by exposure to risk factors that may be related to lifestyle, such as smoking, or environmental risk factors, such as exposure to radiation. This is also demonstrated by migration studies with the risk for women who migrate from low to high risk countries typically increasing, for example; Japanese migrants to the USA experience rapidly increasing breast cancer rates,[26;27] Italian and Polish women had a higher risk of breast cancer after they migrated to Australia, where incidence rates are higher, and breast cancer incidence in Polish migrant communities in America tripled within one generation, reaching the high rates of U.S.-born women[6]. Many factors have been associated with these changes in breast cancer

risk including having fewer children later in life, dietary changes, increasing rates of obesity and exposures to carcinogenic agents resulting from adoption of a Westernised lifestyle[6].

It is thought that breast cancer is caused by an interaction between genes, the environment and a person's lifestyle but very little is known about the exact causes of breast cancer. Several risk factors have been identified including age, reproductive history, personal or family history of breast cancer, diet, lifestyle, exercise, exposure to ionising radiation, mammographic density and height.[28]

Most breast cancers are sporadic (approximately 65-75%)[29]. These cancers affect mainly older people and they are characterized by a lack of family history (no two first degree relatives in a family are affected). A family history can increase a woman's chance of being affected by breast cancer by two to three fold[14]. Inherited cancers affect younger patients and the bilateral form of the disease (affecting both breasts) is more common than in sporadic cancers[9].

A Scandinavian register study has shown that daughters who get breast cancer if their mothers had breast cancer and died have a worse prognosis compared with those whose mother had breast cancer and survived[30] providing evidence that breast cancer prognosis is inherited.

Many years of intensive studying have focussed on understanding susceptibility to breast cancer. In the 1990s the high penetrance breast cancer susceptibility genes BRCA1[17;18] and BRCA2[19] were identified using family-based linkage studies. These genes are critical for genome integrity having an important role in DNA double strand break repair by homologous recombination[31;32]. Mutations in the BRCA genes account for approximately 30-40% of all hereditary breast cancers[33] (Figure 5) and this accounts for only 2-3% of all breast cancers [12;22]. Women with an inactivating mutation in one of these genes have a risk of approximately 60%-88% to develop breast cancer[12;13], which is five to ten times that of the control female population[14].

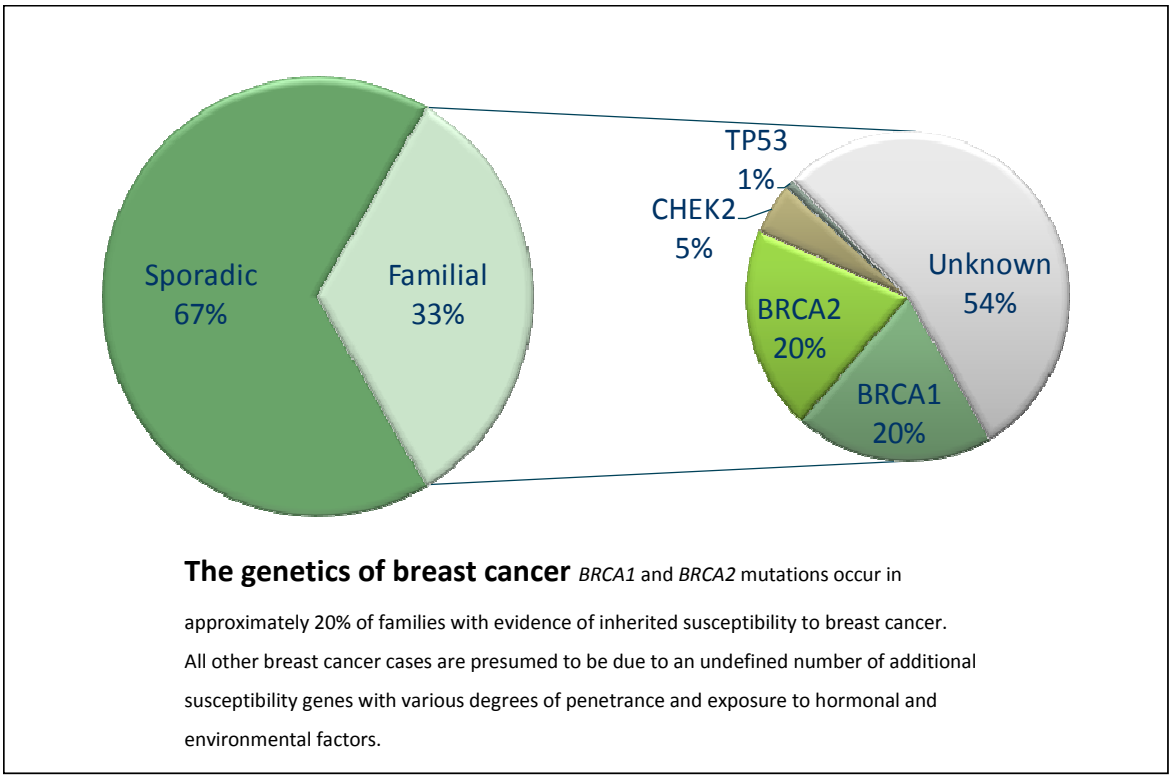


Figure 5: The genetics of breast cancer, adapted from[12;29]. Approximately one third of all breast cancer cases are familial, of these, over half of the cases result from as yet unknown genetic causes.

Further genetic linkage studies were instrumental in identifying additional breast cancer susceptibility genes with high (TP53) and intermediate penetrance (PTEN, CHEK2, ATM, RAD50, BRIP1, PALB2 and NSB1) for inherited breast cancer[34]. However, despite intense efforts, linkage studies have failed to identify additional breast cancer susceptibility genes for familial breast cancer and it is likely there will be no other single gene responsible for a large proportion of the familial aggregation not arising due to mutations in BRCA1 and BRCA2[35;36] as the searches that mapped these genes would have had sufficient power to map a third gene, had it been responsible for a significant proportion of the remaining families[22]. Mutations in the known high and intermediate penetrance genes account for around 50% of the familial breast cancer cases[23] and so it is assumed the remainder of the genetic risk results from mutations in low penetrance susceptibility genes. Linkage studies lack the power to detect low risk alleles so the search for further susceptibility genes has moved from the classical linkage studies to population-based genetic association studies made possible by completion of the 'HapMap'[20], the Human Genome Project[16] and

significant advances in techniques such as DNA sequencing and SNP screening technologies.

The earlier population-based studies focused on the candidate gene approach looking for associations of SNPs in a limited number of genes with breast cancer, however this approach has not been very successful, yielding mixed results[23]. Large candidate gene association studies have been carried out by SEARCH and BCAC (Breast Cancer Association Consortium). The study by SEARCH reported that common variants in the BRCA1, BRCA2, ATM, CHEK2 and TP53 genes are not likely to increase risk of breast cancer[37] and BCAC reported that common coding variants in caspase 8 and TGFβ1 were associated with a significant risk of invasive breast cancer[38]. Many studies however, including an earlier study by BCAC identifying an association of SNPs in caspase 8, IGFBP3, PGR, SOD2 and TGFβ1 with breast cancer, have only reached borderline statistical significance[39].

More recently several genome-wide association studies have been conducted utilising unbiased scans of the genome to identify SNPs that may be associated with breast cancer risk and, over a decade since the discovery of the high penetrance BRCA genes, several low penetrance variants associated with breast cancer have been published.[21;40;41]

A recent genome wide association study by a large research consortium lead by a Cambridge group identified five new loci showing significant association with breast cancer, four of which contain plausible causative genes[21] (FGFR2, TOX3, MAP3K1, LSP1 and 8q) and SNPs in FGFR2[40] and TOX3[41] have both been associated with oestrogen receptor positive breast cancer in independent GWASs.

The new low penetrance susceptibility genes identified probably explain less than 5% of the excess familial risk of breast cancer[14] and so many more susceptibility factors are expected to be identified each conferring only a small increase in breast cancer (approximately 1.2-1.5 fold increase[23]). The low predicitive value and high frequency of the variants in the population precludes the use of these individual risk factors in genetic counselling.

Table 1- Familial breast cancer genes identified to date[21;34]

Gene	*Penetrance*	*Function*	*Reference*
ATM	Intermediate	DNA damage response	*Walsh et al 2007*
BRCA1	High	DNA damage response	*Walsh et al 2007*
BRCA2	High	DNA damage response	*Walsh et al 2007*
BRIP1	Intermediate	DNA damage response	*Walsh et al 2007*
CHEK2	Intermediate	DNA damage response	*Walsh et al 2007*
FGFR2	Low	Tumour Suppressor	*Easton et al 2007*
Gene Desert	Low	Unknown	*Easton et al 2007*
LSP1	Low	Invasion/metastasis	*Easton et al 2007*
MAP3K1	Low	Cell death (apoptosis)	*Easton et al 2007*
NBS1	Intermediate	DNA damage response	*Walsh et al 2007*
P53	High	Tumour suppressor	*Walsh et al 2007*
PALB2	Intermediate	DNA damage response	*Walsh et al 2007*
PTEN	Low	Tumour suppressor	*Walsh et al 2007*
RAD50	Low	DNA damage response	*Walsh et al 2007*
TOX3	*Low*	*Transcription factor*	*Easton et al 2007*

A study was conducted to analyze the occurrence of breast cancer in relatives of affected individuals to find a genetic model that may account for the familial aggregation not due to mutations in the BRCA genes. The study found that the model that best fit the data was, indeed, the polygenic model, with the susceptibility to breast cancer being conferred by a large number of alleles. The risk that is associated with any one allele is small but the effects are multiplicative so a woman with several susceptibility alleles will be at high risk[35]. Breast tumours are very heterogeneous which suggests that they arise from multiple genetic events[42] likely to be caused by low penetrance genes, the number and identity of these genes is unknown [12;35;36;42].

Mechanisms by which the SNPs at these certain loci cause the susceptibility are still unknown[43], but some SNPs have been shown to affect the intrinsic properties and function of the proteins to a variable degree[44-49].

The polygenic models suggest that multiple commonly occurring low to modestly penetrant SNPs of cancer related genes might have a greater effect on a disease when considered in combination and it has been found that the risk of breast cancer is approximately six times as great among women carrying 14 risk alleles as among those carrying no risk alleles at these loci[50], however this model does not specifically consider their possible interactions[35;51].

Recently, attempts have been made to identify breast cancer risk that is conferred by SNP-SNP interactions[44]. The strategy developed has the potential to identify complex biological links among breast cancer genes and processes. The genetic effect of combinations of functionally relevant SNPs may additively or synergistically contribute to increased breast cancer risk[52].

Many studies have also identified associations of mutations with certain breast cancer phenotypes. There is evidence that mutations in the BRCA genes can cause tumours with a specific molecular phenotype and these phenotypes are accompanied by specific genetic profiles.[53] Mutations in TP53 are associated with a complex phenotype of different tumours called Li-Fraumeni syndrome[54]. Nordgard et al[43] investigated the expression levels of previously identified susceptibility genes (FGFR2, TOX3, LSP1 and MAP3K1)[21] in tumours and found that they are highly significantly differentially expressed between the five established breast cancer subtypes, suggesting expression of these genes in the tumour influence breast cancer subtype. Therefore stratification of patients on the basis of their molecular subtypes may give much more power to classic case control studies and genes of no or borderline significance may become identifiable and possibly even appear to be high-penetrance genes for certain subtypes.

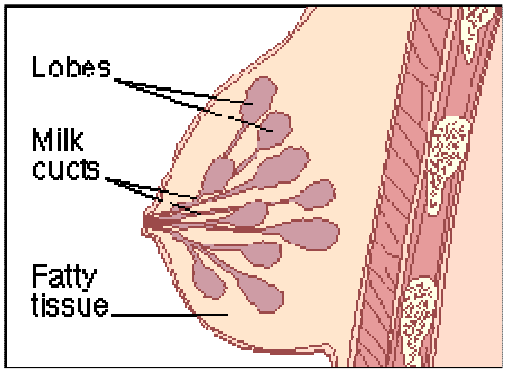
1.2.5 Breast Cancer Biology

Breast cancer comprises a heterogeneous group of tumours with variable prognosis and aggressiveness, which is related to the capacity of malignant cells to invade neighbouring tissues and to metastasize to distant sites[55;56].

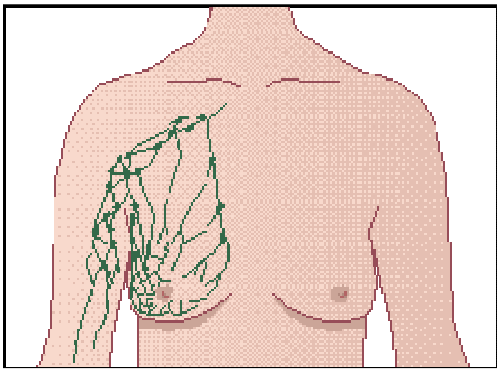
1.2.5.1 Biology

The breasts are made up of fat, connective tissue and glandular tissue. Each breast is divided into sections called lobes. There are 15-20 lobes in each breast. The lobes contain several smaller lobules, which contains groups of tiny milk-producing glands. The lobules end in milk-producing bulbs called terminal ductal lobular units. Milk flows from the lobules, through thin tubes called ducts, to the nipple. The spaces between the lobules and the ducts are filled with fat.

Structure of the breast



Lymph vessels and lymph nodes



Breasts also contain lymph vessels. The lymph vessels carry lymph, a clear fluid that travels through the arteries and circulates through tissues to cleanse them. The lymph vessels lead to lymph nodes, small, pea-sized organs that can be found in the axilla (underarm), near the collarbone and near the breastbone as well as many other sites in the body. Lymph nodes are barriers to the spread of infection, acting as filters along the lymphatic system to trap or remove foreign particles such as bacteria, viruses, cancer cells and other unwanted substances to make sure that they are removed from the body.

Breast tumours are derived from epithelial cells that line the terminal duct lobular unit of the breast.

The most important factor determining the prognosis of a breast cancer is whether it is non-invasive (in situ) or invasive (infiltrating).

Non-invasive cancers confine themselves to the ducts (ductal carcinoma in situ, DCIS) or lobules (lobular carinoma in situ, LCIS) and have not spread to the surrounding tissues in the breast or other parts of the body. They can, however, develop into or increase your risk of developing a more serious, invasive cancer[57].

If cancer cells disseminate outside the basement membrane of the milk duct or milk-making glands in the lobules and grows into surrounding adjacent normal tissue the cancer is known as invasive. Invasive cancers can spread to other parts of the body and form metastases (secondary tumours) at distant sites including regional lymph nodes or visceral organs such as the lungs, liver, brain, and bone (Figure 6). When the lymph nodes are involved in the cancer, they are known as lymph node positive. When lymph nodes are free of cancer, they are known as lymph node negative. In most cases, the more extensive the lymph node involvement, the more aggressive the cancer will be. The route of spread of cancer may include blood stream (hematogenous), lymphatic vessels (lymphovascular invasion) or third space extension (i.e. ascitic fluid dissemination as seen in ovarian cancer). [58] The bone, bone marrow[59] and brain[60] are the most common sites of metastasis in breast cancer. Despite significant improvements in diagnostic and therapeutic modalities for the treatment of cancer, metastasis remains the overwhelming cause of death for cancer patients, responsible for 90% of all cancer deaths[58;61] .

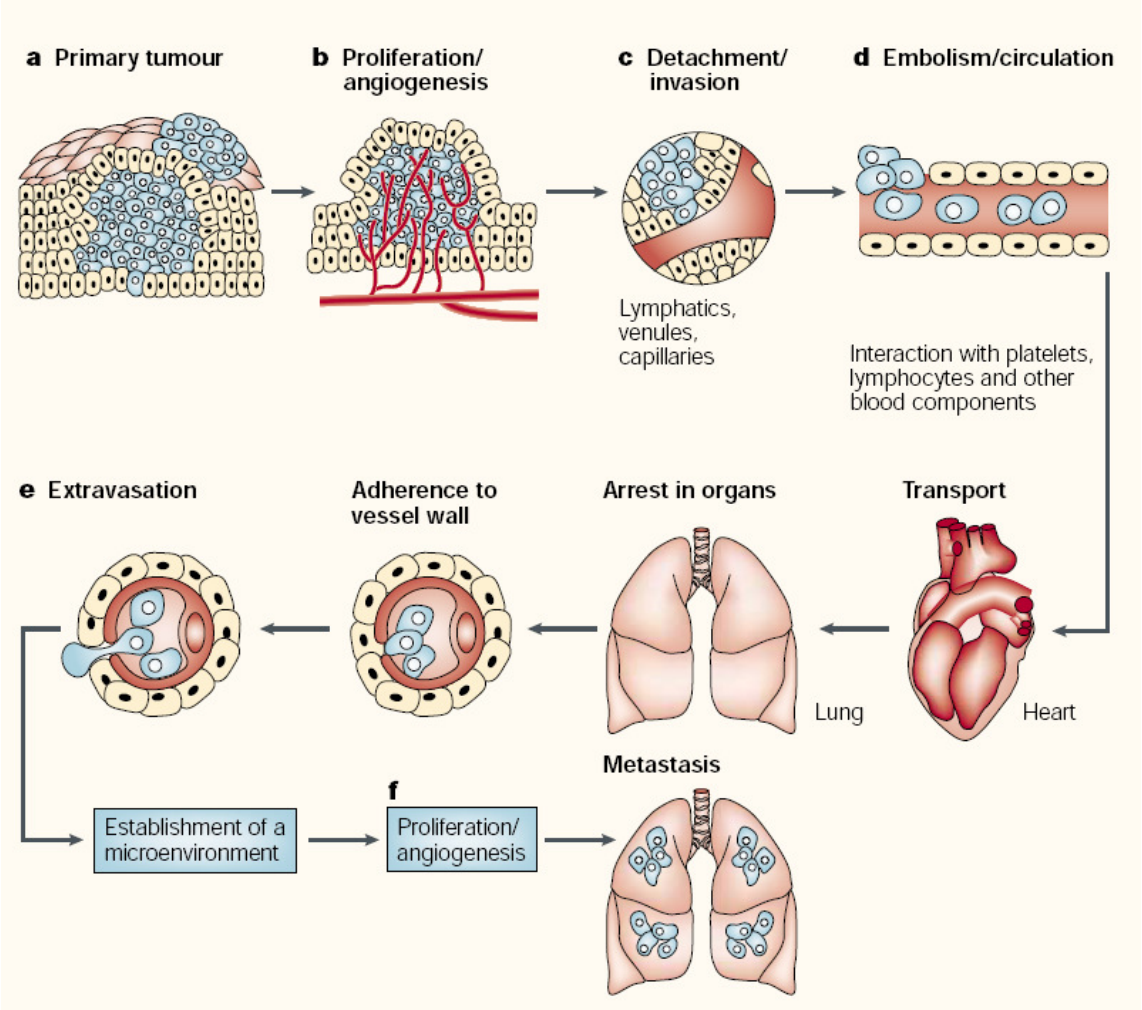


Figure 6: The main steps in the formation of a metastasis. a | Cellular transformation and tumour growth. Growth of neoplastic cells must be progressive, with nutrients for the expanding tumour mass initially supplied by simple diffusion. b | Extensive vascularization must occur if a tumour mass is to exceed 1–2 mm in diameter. The synthesis and secretion of angiogenic factors establish a capillary network from the surrounding host tissue. c | Local invasion of the host stroma by some tumour cells occurs by several parallel mechanisms. Thin-walled venules, such as lymphatic channels, offer very little resistance to penetration by tumour cells and provide the most common route for tumour-cell entry into the circulation. d | Detachment and embolization of single tumour cells or aggregates occurs next, most circulating tumour cells being rapidly destroyed. After the tumour cells have survived the circulation, they become trapped in the capillary beds of distant organs by adhering either to capillary endothelial cells or to subendothelial basement membrane that might be exposed. e | Extravasation occurs next — probably by mechanisms similar to those that operate during invasion. f | Proliferation within the organ parenchyma completes the metastatic process. To continue growing, the micrometastasis must develop a vascular network and evade destruction by host defences. The cells can then invade blood vessels, enter the circulation and produce additional metastases.[62]

The size of a tumour, whether the lymph nodes are involved and whether or not the cancer has spread to other parts of the body are assessed to give an indication of the tumour stage (tumour node metastasis, TNM staging). The appearance of the cancer cells under the microscope is also assessed to give an indication of the grade of the cancer and indicates how quicly the cancer may develop. There are three grades, grade 1 being low grade, grade 2 moderate or intermediate and grade 3 high grade. The appearance of the cells are scored (1-3) on the frequency of mitosis (rate of cell division), tubule formation (percentage of the cancer composed of tubular structures) and nuclear pleomorphism (change in cell size and uniformity) and these three scores added together to allow classification of the grade. A score of 3-5 indicates grade 1, 6-7 grade 2 and 8-9 grade 3. This is known as the Bloom and Richardson grade, after the originators of the scoring system.

As well as the above clinical features oestrogen receptor status is also used as a prognostic marker in breast cancer. The absence of oestrogen receptors in breast tumours is associated with an early recurrence independent of other known prognostic factors, such as axillary lymph node status and tumour size[63], and breast cancer patients with tumours that are ER-positive and/or PR-positive have lower risks of mortality after their diagnosis compared to women with ER-negative and/or PR-negative disease.[64]

Breast tumours are diverse in their history and treatment response. A number of microarray studies have shown that the majority of human breast tumours can be classified into clinically relevant subtypes based on differences in their gene expression patterns.[65-69] Gene expression profiling allows stratification of breast tumours into these groups through the use of expression levels of thousands of genes rather than looking at one single prognostic factor. The ability to stratify breast tumours into different subgroups illustrates how the phenotypic diversity of breast tumours is accompanied by an underlying diversity in gene expression patterns.[66-68;70-74]

There are four main subtypes that have been consistently identified:

1. Basal like subtype
2. ERBB2+ subtype (HER-2 over-expressing)
3. Luminal/ER+ subtype (may be divided into 2 or 3 subclasses with distinct expression profiles and clinical outcomes [66])
4. Normal breast-like subtype

These groups have different clinical outcomes, the basal subgroup having the worst prognosis, HER-2 over-expressing having a poor prognosis, luminal B an intermediate prognosis and the normal breast-like and luminal A subgroups having a good prognosis. The basal and HER-2 over-expressing groups are ER-negative, whereas the Luminal subgroup is ER-positive reflecting the significant role of ER status in prognosis.

Currently tumour grade, stage and hormone receptor status are assessed to predict disease-free and overall survival and determine a patients treatment. Recent work has demonstrated that gene expression profiles may also be used in prognostication. A 'poor prognosis' profile has been developed that is indicative of a short time before developing metastases in lymph node negative patients.[68] Being able to identify the type of tumour a person is likely to develop may help select those patients that may benefit from adjuvant therapy and advise patients on prophylactic treatment.

1.2.5.2 Tumour and Stroma

The stroma (connective, functionally supportive framework) surrounding a tumour consists of many cell types and differs depending on the origin of the tumour. Recent studies have shown that the cells in the stroma influence the process of tumourigenesis to varying degrees depending on the type of tumour (Section 1.2.5.5).[75]

As well as the neoplastic cells, the cellular microenvironment of a malignant tumour also contains a variety of non-neoplastic cellular elements such as adipocytes, fibroblasts, nerve cells, vascular cells (for example, lymph-endothelial, vascular-endothelial cells and pericytes) and large numbers of leukocytes involved in inflammation and immunity (for example lymphocytes (T-cells (helper, suppressor and cytotoxic), B cells, natural killer (NK) cells), dendritic cells, granulocytes, monocytes, and macrophages) surrounded by an extracellular matrix (Figure 7).[75-83] These non-neoplastic cellular elements, along with diffusible growth factors and cytokines, chemokines and other secreted molecules, make up the tumour stroma. [58;84]

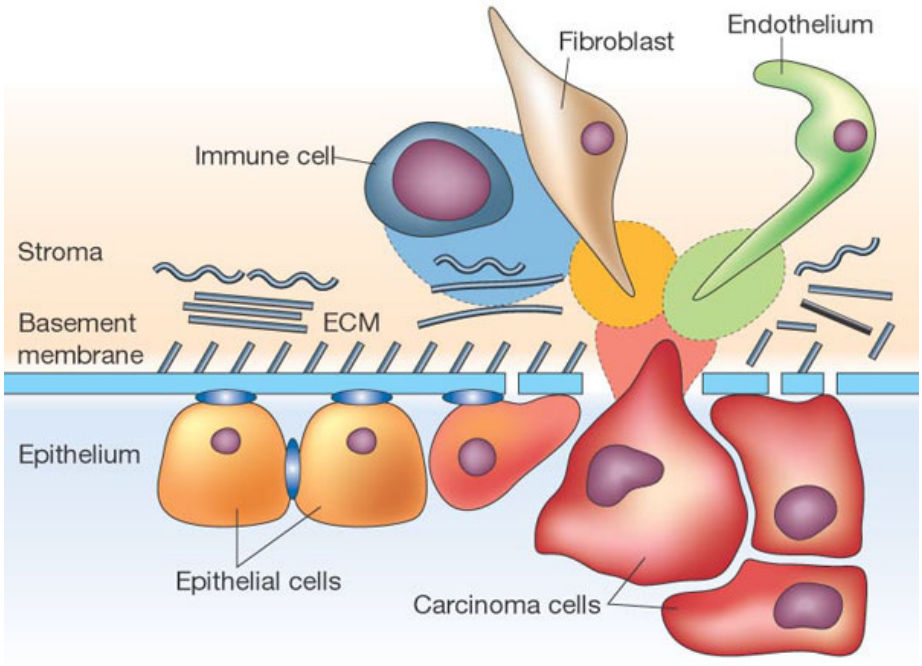


Figure 7: Diagram of the tumour stroma from Liotta and Kohn, 2001[77]. The tumour stroma is composed of many non-neoplasic cellular elements including immune cells and fibroblasts that may contribute to the metastastic potential of a tumour.

The presence of immune cells in human tumours was first described in 1863 by Rudolf Virchow, who observed the presence of immune cells in neoplastic tissues and hypothesized that cancer originates at sites of previous chronic inflammation. Over the past decade our understanding of the inflammatory microenvironment of malignant tissues has supported Virchow's hypothesis and it is now recognized that cancers frequently arise in areas of chronic inflammation for example; colon carcinoma, which is associated with inflammatory bowel disease, stomach cancer following *Helicobacter pylori* infection and hepatocellular carcinomas after hepatitis C infection[85].

Tumour stroma formation and wound healing have been shown to share many important properties[86] and tumours have been described as wounds that do not heal as wound healing is usually self limiting, but tumour cells and/or tumour associated leukocytes and platelets secrete several factors such as inflammatory cytokines, growth factors and chemokines that lead to a persistent immune response, for example by stimulating proliferation of epithelia and generating reactive oxygen species that can cause DNA damage[87] which may directly contribute to malignant progression[81] as well as influencing cancer initiation[84].

As well as tumour-associated inflammatory cells arising from chronic inflammation, the malignant tumour cells can recruit vasculature and stroma through the production and secretion of stimulatory growth factors and cytokines[77;88] and the 'activated' stroma of the tumour microenvironment, consisting of several components including growth factors, other secreted molecules, fibroblasts and immune cells, highly influences the behavior of tumour cells[77] critical to the overall survival of the cancer and, in many cases, is thought to enhance the tumour's malignant capabilities.[89;90] An activated 'metastatic tumour stroma' is sufficient to convert non-metastatic cells to metastatic.[58]

It has been reported that inflammatory cells can account for as much as 50% of the total tumour mass[91] and up to 80% of the total cell mass in breast carcinoma.[83;92]

The metastatic potential of the tumour stroma is in part determined by macrophage and fibroblast secretion of matrix metalloproteinases, which initiate a cascade of events that contribute to activation of the stroma and chemokines (chemotactic proteins) and their associated receptors, which direct and modulate the cellular component of the stroma[58;93], therefore fibroblasts and macrophages play a significant role in breast cancer progression.

1.2.5.3 Fibroblasts

It has been demonstrated that several types of malignant cells, including breast and colon, actively recruit fibroblasts into tumours and, in fact, fibroblasts often represent the majority of the stromal cells within various types of human carcinomas, such as invasive breast cancer.[94] Recent studies have revealed that fibroblasts have a more profound influence on the development and progression of carcinomas than was previously appreciated,[84] however the specific contributions of these cells to tumour growth are poorly understood.[95]

It was initially believed that fibroblasts assumed a fairly passive role in breast cancer, responding to signals released from tumour cells, but recent evidence shows fibroblasts to play a far more active role in tumourigenesis[84] and are associated with cancer cells at all stages of cancer progression, being responsible for the synthesis, deposition and remodelling of much of the ECM in the tumour stroma, increasing the extent of extracellular matrix degradation (a prerequisite for invasion) and are a source of growth factors that influence the growth of carcinoma cells. [84;96;97] It has long been known that these factors can be secreted by neoplastic cells, however it has become clear only in recent years that a major fraction is also produced by stromal fibroblasts. The fibroblasts found in the tumour stroma are termed carcinoma-associated fibroblasts (CAFs) and have a different phenotype to normal fibroblasts. There is substantial evidence that stromal fibroblasts from breast cancer are phenotypically distinct from those found in normal breast[98;99], with fibroblasts extracted from a number of invasive human breast carcinomas being more competent at promoting the growth of mammary carcinomas cells and tumour angiogenesis than comparable cells derived from outside of these tumour masses.[95]

Whereas normal fibroblasts at different anatomical sites have been shown to exhibit a considerable phenotypic and functional diversity[97], fibroblastic cells of the tumour stroma are less well characterized.[99]

Experiments have also shown orthotopic fibroblasts conferred morphogenic and mitogenic induction of breast cancer cells, while ectopic fibroblasts did not always affect the growth of breast cancer cells.[100] This has been supported by more recent work, with co-inoculation of breast cancer cells and breast fibroblasts having increased tumourigenicity and tumour size compared with inoculation of breast cancer cells alone[101]. The potent influence of organ-specific fibroblasts on epithelial morphogenesis and tumourigenesis may arise from the release of one or more soluble factors from orthotopic fibroblasts that exert more stimulatory activity upon cancer cells than factors from ectopic fibroblasts, playing a significant role in cell-cell regulation of cancer progression.

The interaction of fibroblasts and tumour cells may also be crucial in promoting the tumourigenic process, for example the interaction of lung cancer cells and stroma fibroblasts promotes the expression of the angiogenic factor IL-8 both in cancer cells and in the stroma fibroblasts.[102]

1.2.5.4 Macrophages

Macrophages are released from the bone marrow as immature monocytes which circulate in the blood stream before migrating into tissues where they undergo differentiation into resident macrophages. The phenotype of these fully differentiated, resident macrophages can vary markedly within tissues, from that of microglial cells in the brain, Kupffer cells in the liver and Langerhans cells in the skin.[92;103]

Despite these organ specific differences, resident macrophages share a multitude of common functions. They are vital for host defence, playing a fundamental role in immune and inflammatory responses, guarding against microbial infections and helping to repair sites of injury, as well as regulating normal cell turnover and tissue remodelling.[103;104]

Macrophages are pivotal members of the inflammatory infiltrate in the tumour stroma and represent a significant component of the leukocytic infiltrate closely associated with tumour cells seen in most, if not all, malignant tumours[79;81;83;104-107] including breast cancers.[92;108-110] It has been reported that macrophages can comprise 50% of the total mass of a tumour[108] and in invasive breast carcinomas, the neoplastic cell population is often outnumbered by stromal cells such as macrophages, which can constitute a large portion (up to 80%)[92] of the cell mass.[75;111] These macrophages are known as tumour-associated macrophages (TAMs).

There is still some controversy over whether the TAMs are derived from peripheral blood monocytes recruited into the tumour from the local circulation, rather than resident macrophages present in the healthy tissue before tumour development.[79;112] Most TAMs are derived from peripheral blood monocytes recruited into the tumour mass from the circulation throughout the life span of the tumour[113], recruited by tumour derived chemokines (chemotactic cytokines) such as monocyte chemotactic protein-1 (MCP-1), colony stimulating factor-1 (CSF-1) and vascular endothelial growth factor (VEGF), [81;83;103;105] however there is some evidence of local proliferation of macrophages.[92;114]

It was historically believed that macrophages found in tumours were part of an intrinsic defense mechanism, mounting an effective host anti-tumour response by producing anti-tumour cytokines, stimulating the anti-tumour activity of T cells and destroying tumour cells[115;116] but it now appears that most components of the immune system are endowed with potential dual functions and macrophages are actually more likely to be recruited by the tumour cells to promote tumour phenotypes, such as angiogenesis, growth and invasion.[81;83;87;111;117-119]

TAMs exhibit a distinct phenotype which differs from that seen in macrophages in non-malignant tissues. This may be a consequence of their migration into tumour sites which exposes them to different microenvironmental signals that may regulate their localization and function[83] and ''educate'' them to perform functions required by tumour cells in those areas.[103]

The phenotype of TAMs has been described as relatively immature. TAMs show a greatly reduced capacity to lyse tumour cells, present tumour-associated antigens to T cells and express immunostimulatory cytokines that stimulate the proliferation and antitumour functions of T cells and natural killer (NK) cells in vitro than macrophages derived from healthy or inflamed tissues. This may be due in part to their exposure to tumour-derived molecules such as IL-4, IL-10 which may induce TAMs to develop into polarized type II (alternatively activated) or M2 macrophages, associated with poor antigen-presenting capability and production of factors that suppress T-cell proliferation and activity, and are generally better adapted to scavenging debris, promoting angiogenesis and repairing and remodelling wounded/damaged tissues than the classically activated type I or M1 macrophages that are efficient immune effectors.[103;112]

Upon activation by tumour cells, TAMs release a wide array of growth factors, proteolytic enzymes (predominantly matrix metalloproteinases[58;111;120;121]) and cytokines that modify the extracellular matrix (ECM) and play an important role in angiogenesis and vascularisation allowing proliferation, survival, cell invasion into surrounding normal tissues and metastasis to local and distant sites.[75;79;81;83;103-105;111;119;122-125] Therefore it is

no surprise that the majority of studies published have identified a strong correlation between extensive TAM infiltration and a poor prognosis in a various forms of cancer[58;75;92;104;126;126;127] including carcinomas of the breast[105;109;128;129], ovaries, cervix and bladder, however there is conflicting evidence for their role in prostate, lung and brain tumours[92] inferring a clear role for macrophages in driving tumour progression. In some cancers TAM numbers have been shown to be an independent prognostic factor.[92;103]

In breast cancer a high level of TAM infiltration has been associated with increased angiogenesis and increased involvement of local lymph nodes [103;105;127;130], relapse free and overall survival[105;109], high tumour grade, large tumour size and tumour necrosis[131], high tumour grade and poor prognosis[109], promoting the progression of preinvasive mammary tumours to malignant lesions and increasing formation of lung metastases[110] and also with a poor prognosis in invasive ductal tumours[129].

Several of the actions of TAMs (promotion of angiogenesis, induction of tumour growth and enhancement of tumour cell migration and invasion) all contribute to the invasiveness and matrix-degrading activity of cancer cells[104] and this is by far the most common phenotype associated with TAM infiltration with many studies identifying a correlation between a high number of macrophages and the metastatic potential of a tumour.[83;105;110;132] TAMs seem to play roles in both the release of metastatic cells from the primary tumour as well as the establishment of secondary tumours at distant sites[103].

Studies have shown that the increase in metastastatic potential resulting from high numbers of TAMs is MMP dependent with tumour-macrophage interactions inducing macrophages to release MMPs, supporting tumour cell invasion and leading to an increased invasive capacity of the tumour cell.[79] Expression of a panel of MMPs, MMP-7, -9, -11, -13, -14, and their inhibitors, the tissue inhibitors of matrix metalloproteinase (TIMPs), TIMP-1 and -2, have been shown to characterise a phenotype of infiltrating macrophages associated with a poorer prognosis in breast cancer[91]. In particular, MMP-9 production by TAMs has been implicated with an increased malignant potential of

tumours in several studies[81;104], for example promoting formation of lung[133;134], ovarian[135] and brain metastases[60].[58]

1.2.5.5 Role of the Microenvironment

In the past decades the major focus of cancer research has been the transformed tumour cell itself[136;137], while the role of the cellular microenvironment in tumourigenesis has not been widely explored, however it is now beginning to be appreciated that interactions between the tumour and the host environment contribute to a state of malignancy. [138]

The English surgeon Stephen Paget (1855–1926) is credited with being the first to suggest the key role played by the microenvironment in formation of metastasis proposing the 'seed and soil' theory (1889), the concept of which has been supported and confirmed by numerous publications[61;139;140].

Paget analysed autopsy records of 735 women with breast cancer and identified a non-random pattern of metastasis to visceral organs and bones, suggesting that the process was not due to chance but rather that certain tumour cells had a specific affinity for the environment of certain organs. Paget therefore proposed that the process of metastasis did not occur by chance, but rather that certain tumour cells with the ability to metastasise (the 'seed') had a special affinity for the growth-enhancing environment present within specific organs (the 'soil'). Paget concluded that the sites where metastases occur are defined not only by the tumour cell but also the microenvironment of the secondary metastatic site with metastases developing only when the seed and soil were compatible, that is, where the microenvironment of the tumour provided a 'fertile soil' for cells enabled with the ability to grow under the specific conditions provided by the soil, suggesting that the site of metastasis depended on the affinity of the tumour for the microenvironment.[141]

Stephen Paget's theory fits with the current knowledge of tumourigenesis and the process of metastasis. Tumours are biologically heterogeneous and contain subpopulations of cells with different angiogenic, invasive and metastatic properties. Cells possessing the metastatic phenotype may represent a very small fraction of cells within a heterogeneous primary tumour.[139] The outcome of metastasis is dependent on multiple interactions of

metastatic cells with homeostatic mechanisms and the cell's ability to resist stresses that would otherwise impede metastasis. [58;142]

Tumour cells leave the primary tumour through regulated lysis of surrounding stroma, pass through the tumour basement membrane and then through or between endothelial cells (invasion) in order to enter the circulation. While in the circulation, the tumour cells must resist the process of anoikis (programmed cell death associated with loss of cellular contact), evade immune recognition and cope with the sheer physical stress of the circulatory system. Ultimately the metastasizing cells arrest in a distant capillary bed and extravasate (leave the circulation) into the organ parenchyma.[142] At the distant site the cell must survive the stresses of a new and likely hostile microenvironment, proliferate, induce angiogenesis and/or co-opt existing blood vessels and then successfully grow into a metastatic lesion. [58;143]

Survival of metastatic cells at the secondary site, where the new microenvironment (the "soil") encountered by the metastatic cell is considered to be foreign or inhospitable, may be a consequence of intrinsic features of the metastatic cell such as the ability to effectively engage in a molecular cross-talk with its surroundings, modulating the environment of the secondary site. Aberrant expression of proteins by the tumour cell is thought to provide enhanced proliferative and survival capabilities[144-148] in the tumour cell's new microenvironment,[58] for example, integrins, key mediators of cell–cell interactions, often have deregulated expression in tumour cells and may play a crucial role in this process[149;150].

Successful metastatic cells that arrive at a secondary site in the body interact with cells in the microenvironment to modulate the secondary site, producing an environment that is conducive to survival of the metastatic cell. Recent evidence has suggested that development of such an environment may happen even before the arrival of metastatic cells themselves via priming by bone marrow derived cells[58;151] or may even be related to the patient's germline genetics and therefore precede primary tumour formation.[152] This

links back to Paget's hypothesis, suggesting that specific genetic determinants of the host (i.e. soil) contribute to the success of the metastatic process (i.e. seed).

In summary, the work by several groups suggests that the risk for metastatic progression is in part defined by the genetics of the patient, genetic changes that develop early in the process of tumour development and the subsequent emergence of cells within the tumour that possess the cellular tools needed for metastasis.[58]

1.3 Hypothesis Development

This study is designed to investigate whether common genetic variants may have a functional significance in a large group of breast cancer affected women, i.e. predetermine breast cancer phenotype and/or prognosis, as well as statistical significance, i.e. determining risk. This will help to understand susceptibility in young onset breast cancer cases and identify better ways to pick out high risk gene carriers.

Identifying genetic variants that predict the type of breast cancer a woman may develop will be of significant benefit in advising risk management and directing adjuvant therapies, for example if a patient at risk of breast cancer was likely to get a particularly aggressive form of the disease, they may wish to have prophylactic treatment, such as mastectomy, whereas if they were likely to develop a less aggressive, oestrogen receptor positive cancer with low metastatic potential the patient may decide to have no prophylactic treatment and, if the cancer occurs, treat with Tamoxifen. It will also be of use in determining who will respond to certain therapies and who will not.

Several lines of evidence including the ethnic differences in tumour types,[153;154] specific subtypes of cancer seen with germline BRCA1 mutations[53] and the low penetrance variants like FGFR2[43], and the Scandinavian register study showing a worse prognosis for daughters who get breast cancer if their mothers had breast cancer and died compared with those whose mother had breast cancer and survived[30] suggest it is likely that genetic polymorphisms play a role in determining some of the biological features of the tumour, including invasiveness, metastatic potential and prognosis, as well as dictating the environment in which the tumour is growing which will also affect its growth and invasiveness. Therefore this study was set up to investigate the association of common genetic variants in a range of candidate genes with the following prognostic phenotypes; grade, ER status, LN involvement, lymphovascular invasion and distant metastasis.

Chapter 2 Methods for Association Studies

2.1 Overview

To identify SNPs that might be associated with particular disease phenotypes, 30 candidate genes were selected that had a biological reason for being of interest to study in relation to breast cancer. After the 30 genes had been selected 260 SNPs were identified in these genes, some were previously reported, others were identified from genome databases or selected purely for their location within the gene to ensure coverage of the whole gene. 206 of these SNPs were genotyped successfully in our cohort of 1001 patients selected from the POSH study. Patients were selected who had no significant family history of breast or ovarian cancer and where plenty of DNA was available.

DNA for genotyping was extracted from whole blood samples and the DNA genotyped by Sequenom (San Diego, CA, USA) using the iPLEX service based on mass spectrometry.

The genotype data was subjected to a screening procedure to remove poor quality data from the analysis and just over one third of all SNPs and 58 individuals were removed. To avoid confounding due to population stratification, multidimensional scaling analysis was was carried out on the data to ensure only Caucasians were included in the final analysis. A further 44 samples, 35 African and 9 Asian, were excluded to leave only individuals of western European ancestry.

The cases were then stratified into subgroups based on ER status (positive and negative), LN involvement (positive and negative), lymphovascular invasion (present or absent), distant metastasis (present and absent) and grade (grade one and grade three). Grade is a subjective classification, and some grade 2 tumours might have a gene expression signature that is similar to tumours of grade 1 or 3[155] and a systematic expert review may lead some grade 2 cases to be reclassified into either grade 1 or grade 3 whereas grade 3 and grade 1 are less often likely to move to grade 2 so, for clarity, we compared only the phenotypic extremes of grade 1 and grade 3 in this study.

Associations of the SNPs passing QC with disease phenotypes were tested using the Cochran-Armitage trend test and 2-by-3 tables of disease by genotype using Plink[156], a whole genome analysis toolset.

Associations of SNPs with prognosis were also investigated. The mean follow up time for the patients in our study who passed quality control was 2.4 years so we used distant disease free survival times (taken as the time between initial diagnosis and the first distant metastasis) as a substitute for overall survival since the majority of patients developing distant metastases will ultimately die from the disease[157].

2.2 Resources

To perform the association studies, we used data from two openly available resources on the internet.

2.2.1 UCSC

The University of California Santa Cruz Genome Browser Database (UCSC)[158-160] is developed and maintained by the Genome Bioinformatics Group within the Center for Biomolecular Science and Engineering at the University of California Santa Cruz. It is a genome browser website containing the reference sequences for a large collection of genomes. The UCSC genome browser integrates the work of many scientists worldwide, including data from the MWG (Mammalian Gene Collection), RefSeq and Ensembl, in an interactive, graphical display that presents both experimentally validated and computer-predicted genes as well as evidence to help scientists recognize key features of genes and predict their function. As of September 2009, genomic sequence and a basic set of annotation 'tracks' (known genes, predicted genes, ESTs, mRNAs, CpG islands, assembly gaps and coverage, chromosomal bands, mouse homologies) were provided for 47 organisms.

The UCSC site also provides links to many off-site repositories such as PubMed, GenBank, Entrez and OMIM.

The Genome Browser, other tools, downloadable data files and links to documentation and other information can be found at http://genome.ucsc.edu/.

2.2.2 HapMap

The International HapMap project[20;161-163] is a large collaboration among scientists and funding agencies from Japan, the United Kingdom, Canada, China, Nigeria, and the United States. It began in 2002 to identify and catalogue common genetic variants occurring in humans.

It gives information about these variations, including their location in the genome and distribution among populations from different countries.

The Information collected by the HapMap project can be used by researchers to identify links between genes and common diseases as well as individual responses to medications and environmental factors.

Data from four different populations with different ancestry are being collected:

- Yoruba in Imbadan, Nigeria (YRI) – African
- Japanese in Tokyo, Japan (JPT) - Asian
- Han Chinese in Beijing, China (CHB) – Asian
- Utah residents with ancestry from northern and western Europe (data provided by the Centre d'Etude du Polymorphisme Humain (CEPH)) (CEU) – Caucasian

The DNA samples for the HapMap have come from 270 people, 30 sets of samples (two parents and an adult child, each set called a trio) from the Yoruba people of Ibadan, 30 trios from U.S. residents with northern and western European ancestry, 45 unrelated individuals from Tokyo, Japan, and 45 unrelated individuals from Beijing, China.

2.2.3 Plink

The statistical analysis was carried out using the analysis software PLINK[156]. PLINK is a free, open-source, whole genome association analysis toolset developed by Shaun Purcell at the Center for Human Genetic Research (CHGR), Massachusetts General Hospital (MGH) and the Broad Institute of Harvard & MIT, with the support of others. It is designed to perform a range of basic, large-scale analyses in a computationally efficient manner. The focus of PLINK is purely on analysis of genotype/phenotype data. The statistical analysis was performed using SPSS version 17.

2.3 Gene Selection

Many genes have been implicated as potential candidates for breast cancer susceptibility. We selected 30 genes that were either potential candidate genes or contained SNPs that had previously been reported to have an association with breast cancer susceptibility. One SNP selected was a previously reported SNP found in a gene desert[21].

These 30 genes can be broadly divided into different categories based on their function, however some may fall into more than one category, for example BRCA1 is involved in DNA replication and repair but, due to its impact on the cell cycle, is also known as a tumour suppressor.

The following section details the categories of genes we selected and evidence implicating the particular genes in breast cancer.

2.3.1 Transcription Factors

Transcription factors are proteins that bind to specific DNA sequences and regulate the transcription of DNA to mRNA. They act by increasing (activators) or decreasing (repressors) the binding of RNA Polymerase to specific genes. Transcription factors can act by themselves or as part of a complex with other proteins.

Genes encoding three members of the nuclear receptor superfamily, the androgen receptor, progesterone receptor and oestrogen receptor alpha were all chosen along with three additional transcription factor-encoding genes GATA3, TOX3 and TWIST.

2.3.1.1 Androgen Receptor (AR)

The androgen receptor, alternatively known as the dihydrotestosterone receptor, is a steroid hormone receptor. It is activated by the binding of the androgenic hormones testosterone or dihydrotestosterone. It is a member of the nuclear receptor superfamily and functions as a ligand activated transcription factor regulating gene expression.

Nantermet et al[164] found that AR modulates the expression of genes involved in proliferation and differentiation, repressing expression of several key cell cycle inhibitors while modulating members of the Wnt and Notch signalling pathways, multiple growth factors, peptide hormone signalling systems and genes involved in MAP kinase and calcium signalling.

There is evidence that androgens directly stimulate growth of human breast cancer cell lines[165]. ARs are expressed in approximately 70%-90% of invasive breast cancers, comparable with or higher than the frequency for ER (70%-80%) or PGR (50%-70%)[166-168]. ARs are the sex hormone receptors found most frequently in both primary and secondary breast cancer and in metastatic tissues are present with twice the frequency of PGR, with one in four of these tumours expressing only AR[169]. There is evidence indicating steroids can upregulate MMPs, contributing to invasiveness of the tumour via degradation of the basement membrane and extracellular matrix[170]. A few studies have investigated the effect of AR expression on breast cancer prognosis but most studies have found no significant associations[166;171;172], however a specific value of AR expression has been shown to be a prognostic indicator in breast cancer although the functional role of AR in these neoplasms is still unclear.[173]

Many polymorphisms have been identified in the AR in prostate cancer[174] but have not been extensively investigated in relation to breast cancer, however long repeat lengths of the CAG repeat length polymorphism in exon 1 of the AR gene have been associated with a decreased ability to activate androgen responsive genes and a decreased age at diagnosis in BRCA1 mutation carriers, suggesting pathways involving androgen signalling may affect the risk of BRCA1-associated breast cancer.[175]

2.3.1.2 Progesterone Receptor (PGR)

The progesterone receptor is a member of the steroid receptor superfamily and exists as two functionally distinct isoforms, PRA and PRB. Hopp et al[176] indicated that PRA rich breast tumours have a heightened aggressiveness. Leygue et al[177] identified several

variants in the PGR in breast cancer, mainly exon deletions but the significance of these variants is unknown. Yeates et al[178] identified a truncated PGR protein, found in breast cancers, that is ligand binding and seems to be derived from PRA. De Vivo et al[179] identified variants in the PR gene that may predispose to endometrial cancer. One promoter region polymorphism, +331G>A, created a unique transcription start site, increasing transcription of the PGR gene and favouring production of the PRB isoform.

2.3.1.3 Oestrogen Receptor Alpha (ESR1)

The gene ESR1 encodes the oestrogen receptor ERα which belongs to a family of transcription factors known as the nuclear receptor superfamily. These receptors are responsible for mediating the effects of steroids on development, reproduction, proliferation, cellular homeostasis and gene expression. In particular, the oestrogen receptor has a role in the growth, differentiation and function of the normal breast, therefore makes a good candidate gene for breast cancer susceptibility. Many early studies reported that certain ESR1 variants were more common in breast cancer affected patients than in controls[180-182]. Previously associated variants include 478G>T (Gly160Cys)[181], 908A>G (Lys303Arg)[183] and 975C>G (common synonymous SNP in exon 4, rs18011132)[182;184], however Schubert et al[185] investigated these and seven other ESR1 polymorphisms and found that inherited genetic variation is not a mechanism by which the oestrogen receptor is commonly involved in breast cancer development[185]. The best characterized SNPs in the ERS1 gene are the *Pvu*II and *Xba*I restriction site polymorphisms. The polymorphisms, c454-397T→C (rs2234693) and c454-351A→G (rs9340799), are 397 and 351 bp upstream of exon 2 and have been described by the name of detecting restriction enzyme, *Pvu*II or *Xba*I respectively[186;187]. These SNPs are in strong linkage disequilibrium. Boyapati et al[188] reported no overall association of these SNPs with breast cancer survival but the CC genotype of the PvuII polymorphism had a worse prognosis in ER-ve breast cancer patients. Rapuri et al[189] found that women with either the CC or GG genotype had lower bone remodelling, lower rates of bone loss and had more beneficial effects from hormone therapy in postmenopausal women and Van Duijnhoven et al[190] found that these polymorphisms were associated with an increase in mammographic

density, a strong risk factor for breast cancer, and that these polymorphisms are associated with an increased susceptibility to the effects of hormone therapy on mammographic density[191].

2.3.1.4 GATA Binding Protein 3 (GATA3)

GATA3 encodes a transcription factor, belonging to a family of at least six members, however its function remains largely unknown. The GATA proteins play critical roles in development, including cell fate specification, regulation of differentiation and control of cell proliferation and movement. Hoch et al[192] first described the correlation in expression between GATA3 and oestrogen receptor status, and proposed that GATA3 is likely to regulate a set of genes involved in the hormone responsive phenotype. GATA3 negative tumours have an eight fold higher risk of being unresponsive to hormone treatment when compared to GATA3 positive tumours[193]. GATA3 is highly expressed by luminal epithelial cells in the breast[194] and has been found to have low expression in invasive carcinomas with a poor clinical outcome (higher histological grade, positive lymph node status, larger tumour size, ER and PR negative and ERBB2 positive as well as a shorter disease free and overall survival)[194]. High levels of GATA3 have been found in the luminal A subtype of breast cancer that is associated with oestrogen receptor expression and a favourable prognosis[67;68;192;195;196]. Usary et al[196] suggested that loss of GATA3 contributes to tumourigenesis in oestrogen receptor positive breast cancers.

2.3.1.5 TOX3

TOX3 (TNRC9) is a member of the high mobility group family of non-histone chromatin proteins, a large and diverse family of HMG-box proteins that function as architectural factors in the modification of chromatin structure by bending and unwinding DNA.

Increased expression of TOX3 has already been associated with bone metastasis of breast cancer[197]. Smid et al[198] found that TOX3 was one of 69 genes that was differentially expressed in patients who experienced breast cancer relapse to bone versus those who experienced relapse elsewhere in the body. Several SNPs in TOX3 have been investigated

in relation to breast cancer. Three TOX3 SNPs (rs12443621, rs8051542 and rs3803662) that are in strong linkage disequilibrium were identified as having a significant association with breast cancer risk.[21] The association between rs3803662 and breast cancer risk was not confirmed by Huijts et al[199], however they found that the minor allele of rs3803662 was associated with a younger age of onset but this difference was not significant. Huijts et al also found that the SNP rs12443621 was significantly more often negative for the progesterone receptor and the minor allele of SNP rs12443621 was significantly associated with a lower body mass index.

2.3.1.6 TWIST1

The TWIST gene encodes a basic helix-loop-helix transcription factor. It has been shown to play a role in halting differentiation, inhibiting apoptosis and interfering with the p53 tumour-suppressor pathway.[200] Mutations in this gene have been found in patients with Saethre-Chotzen syndrome, an autosomal dominantly inherited craniosynostosis (premature closure of one or more cranial sutures). Sahlin et al[201] studied Saethre-Chotzen affected women and found a novel association with this syndrome and breast cancer, indicating TWIST 1 may be a novel breast cancer susceptibility gene. Since then Martin et al[202] found that TWIST is inappropriately expressed in breast cancer and may play a part in the progression of human breast tumours. It has been shown to play an essential role in metastasis[203] and high expression levels are associated with invasive lobular carcinoma, a highly infiltrating tumour type.[203] Mironchik et al[204] demonstrated that TWIST overexpression in breast cancer cells can induce angiogenesis, correlates with chromosomal instability and promotes epithelial-mesenchymal-like transition that is pivotal for the transformation into an aggressive breast cancer phenotype.

2.3.2 Growth Factors

Growth factors are naturally occurring proteins or steroid hormones that are capable of stimulating a variety of cellular processes including cellular growth, proliferation and cellular differentiation. Growth factors act as signalling molecules between cells, binding to cell surface receptors. Many growth factors are quite versatile, stimulating cellular

division in numerous different cell types; while others are specific to a particular cell-type, for example, bone morphogenic proteins stimulate bone cell differentiation while fibroblast growth factors and vascular endothelial growth factors stimulate blood vessel differentiation (angiogenesis). Therefore mutations in growth factors or their receptors may contribute to excessive growth or proliferation of cells and may influence the growth rate of some cancers.

2.3.2.1 Mitogen Activated Protein Kinase Kinase Kinase 1/MEK kinase 1 (MAP3K1)

MAP3K1 is a serine threonine kinase that is activated by many apoptotic stimuli and induces apoptosis.[205-207] Deng et al[208] found that MAP3K1 transmits signals leading to the transcriptional activation of genes that are involved in cell migration. A SNP in MAP3K1 (rs889312) has been identified as having a significant association with familial breast cancer risk in a genome wide association study[21] and Huijts et al[199] found that patients carrying one or two minor alleles of the rs889312 SNP were less likely to have lymph node positive breast cancer than those with no copies of the minor allele.

2.3.2.2 Fibroblast Growth Factor Receptor 2 (FGFR2)

FGFR2, a member of the fibroblast growth factor family, is a receptor with tyrosine kinase activity. Activation of the receptor can induce cell migration, proliferation and differentiation, and can also stimulate angiogenesis. FGFR2 can acquire transforming potential through gene amplification, protein over-expression and/or mutations, leading to a subsequent activation and aberrant downstream signalling. It has previously been shown to be important in mammary gland development and neoplasia[209] and its amplification or over expression has been identified in breast cancer[210]. Two recent genome wide studies have identified SNPs in FGFR2 that have an association with breast cancer. Hunter et al[40] identified four SNPs (likely in high LD due to their similar degree of association) in intron 2 of FGFR2 that were significantly associated with breast cancer (rs11200014, rs2981579, 1219648, rs2420946) Easton et al also identified rs2981582 as a susceptibility allele in breast cancer.[21] Huijts et al[199] investigated the association of the

seven significant polymorphisms identified by Easton et al[21] with breast cancer characteristics in a Dutch cohort and found that FGFR2 was associated with a positive family history of breast and/or ovarian cancer. F*GFR2* rs1219648 and rs2981582 genotypes were significantly associated with breast cancer in European-American women with oestrogen receptor-positive (ER+), progesterone receptor-positive (PR+) and HER2/Neu-negative (HER2–) tumours[211].

2.3.2.3 ERBB2

The ERBB2 gene, also known as HER-2 or neu, encodes a member of the epidermal growth factor receptor family of tyrosine kinases. ERBB2 has no ligand binding domain of its own but forms heterodimers with other members of the EGF receptor family and enhances the kinase mediated activation of downstream signalling pathways, for example those involving mitogen activated protein kinase (MAPK) and phosphatidylinositol-3 kinase (PI3K). Its amplification and/or overexpression has been reported in numerous cancers including breast and ovarian tumours[212;213]. In a population based case control study Xie et al[214] found that the valine allele of the val655ile polymorphism (rs1136201) was associated with an increased risk of breast cancer, particularly among younger women. The frequency of this allele was found to vary between different ethnic groups[215].

2.3.2.4 Vascular Endothelial Growth Factor (VEGF)

VEGF is an endothelial cell-specific mitogen that induces angiogenesis. The formation of new blood vessels (neovascularisation) is needed by solid tumours in order to supply its metabolic demands and to provide potential routes for tumour dissemination and metastasis[216;217] and therefore plays a crucial role in the progression of cancer. There is variability in the angiogenic profiles of tumours and it has been suggested that some of this variability could result from germline mutations in genes controlling angiogenesis[218] and since pre-invasive and invasive tumours vary significantly in their angiogenic requirements, some women may be genetically predisposed to an invasive breast cancer phenotype[219]. Individuals with more significant angiogenic profiles have been shown to have a greater risk of breast cancer[220] and breast cancers of those with a more significant

angiogenic profile are more likely to be histologically aggressive[221;222] and have a greater likelihood of metastatic disease.[223]

Several studies have investigated the association of VEGF polymorphisms with breast cancer risk but have had inconsistent results. SNPs that have been investigated include 936C/T (rs3025039), -1154G/A, -2578C/A, -634G/C (rs699947), -2489C>T, -1498C>T, -7C>T, C-460T (promoter, rs83306), G+405C (5'UTR,rs2010963) and C+936T (3'UTR) C-406T and G+450C, 1612G>A[219;224-230].

2.3.2.5 Insulin-Like Growth Factor 1 (IGF-1)

IGF-1 has been implicated in breast cancer because of its role in stimulating mitogenesis and promoting differentiation as well as mammary gland cell proliferation and survival[231]. Oestrogen plays a role in the aetiology of breast cancer and both regulates the IGF family and is itself influenced by the IGF family. Results of studies investigating the association of IGF-1 with breast cancer have been inconsistent.[232-235] Several studies have investigated the effects of SNPs in IGF-1 on breast cancer risk. The SNPs rs1520220 and rs2946834 have been associated with mammographic density, one of the strongest risk factors for breast cancer[236]. Canzian et al[237] identified a small but significant association between polymorphisms at the 5' end of the IGF-1 (rs21262679, rs35765 and rs35767) gene and breast cancer risk, especially in younger women, and a UK case-control study of 4647 cases and 4564 controls investigating the effects of nine SNPs in the IGF-1 gene and circulating IGF-1 levels and breast cancer risk found the tag SNP rs1520220 had the most significant association with IGF-1 levels, however it was suggested that rs6220 in the 3'UTR region of the gene was most likely the causative SNP[238].

2.3.3 Cytokines

Cytokines are a diverse class of powerful signalling proteins that are used extensively in cellular communication, immune function and embryogenesis. They are released from inflammatory tissue, connective tissue and immune system cells and act by autocrine (on the cell that released them) and paracrine (on another cell) mechanisms. More than 50

cytokines have been identified and the cytokine superfamily includes interleukins, chemokines, colony-stimulating factors, growth factors, interferons and the transforming growth factor and tumour necrosis factor families. The cytokine network is complex and not yet fully understood, however it is clear they are important pathophysiological mediators and are implicated in the pathogenesis of numerous disease states.

The terms growth factor and cytokine are sometimes used interchangeably. Historically, cytokines were associated with haematopoietic (blood forming) cells and immune system cells (e.g., lymphocytes and tissue cells from spleen, thymus and lymph nodes), however it is now known that some of the same signalling proteins that the haematopoietic and immune systems use are also used by many other cells and tissues during development and in the mature organism.

The cytokines involved in immune and inflammatory responses (interferon, interleukins, and TNF-alpha) are produced mainly from lymphocytes (lymphokines) and macrophages (monokines) but also from other leukocytes, endothelial cells and fibroblasts. Cytokines act on high affinity receptors on the target cell. Many of the lymphokines are also known as **i**nterleukins (ILs)**,** since they are not only secreted by leukocytes but also able to affect the cellular responses of leukocytes. Specifically, interleukins are growth factors targeted to cells of haematopoietic origin.

2.3.3.1 Interleukin-6 (IL-6)

IL-6 is a pleiotropic proinflammatory cytokine involved in various physiological and pathophysiological processes in the body, including inflammation, bone metabolism and carcinogenesis. IL-6 has been shown to have antiadhesive effects[239] and may increase the activity of 17-β- hydroxysteroid dehydrogenase which converts oestrone to oestradiol, which may in turn increase the concentration of oestradiol around breast tumours.[240] Several studies have associated the presence of IL-6 with breast cancer. Elevated serum levels have been found to be independent prognostic factors in breast cancer with increased IL-6 levels being associated with advanced stage disease and a worse

prognosis[241-244] and it is one of several genes overexpressed in the 'poor prognosis signature' of inflammatory breast cancer.[245] In contrast IL-6 inhibits growth of breast cancer cell lines[246] and high levels of expression of IL-6 mRNA and protein have been associated with a better prognosis and less malignant breast cancer phenotype.[247-249] Polymorphisms in the IL-6 promoter region have been studied extensively with inconsistent results. The SNPs -572G>C and -174G>C have both been associated with breast cancer[250] and -174G>C (rs1800795), shown to influence protein expression rates in vivo by decreasing the rate of transcription[251;252], has been associated with an increased risk of breast cancer[253] and a more aggressive phenotype.[254]. In contrast, the -174G>C SNP has been linked with an earlier stage of disease and significantly better survival in ovarian cancer[255], an improved outcome in high risk breast cancer[256] and has been shown to have no effect on breast cancer risk or severity[257]. Patients who are exposed to over ten hours of passive smoke a week and have the IL-6 SNP rs2069832 have a four fold increase in risk of breast cancer.[258]

2.3.3.2 Transforming Growth Factor Beta (TGFB)

TGFB is a multifunctional cytokine, regulating proliferation and differentiation in many different cell types, acting as a negative autocrine growth factor. Dysregulation of TGFB activation and signalling may lead to apoptosis. Upregulation of TGFB has been found in tumour cells.[259] Grainger et al[260] identified two polymorphisms in the promoter region, -800G>A and -509C>T. The -509C>T polymorphism (rs1800469) was found to be significantly associated with plasma concentrations of TGFB1, suggesting that predisposition to various diseases, including some forms of cancer, may be correlated with the presence of particular alleles at the TGFB1 locus. TGFB has been studied in relation to breast cancer. The activated TGFBR1 (TGF beta type 1 receptor) has been implicated in the formation of lung metastases, while impairing ERBB2 induced tumour growth.[261] Dunning et al[262] found that the promoter SNP C-509T and the T+29C signal peptide SNP (encoding Leu10Pro), in strong linkage disequilibrium, were significantly associated with increased incidence of invasive breast cancer. The C-509T SNP is not in a known consensus sequence for a promoter regulatory element so is unlikely to affect TGFB1

expression whereas the Leu10Pro signal peptide substitution may potentially effect TGFB1 secretion so it is likely it is this SNP having the effect, rather than C-509T and the L10P SNP (rs1982073) was also associated with risk of invasive breast cancer based on data contributed to the Breast Cancer Association Consortium, however this significant association was confined to patients with progesterone receptor negative tumours.[38]

2.3.3.3 Tumour Necrosis Factor Alpha (TNFA)

The TNFA gene encodes a proinflammatory cytokine belonging to the TNF superfamily and is a major mediator of inflammation involved in regulating many biological processes including cell differentiation, proliferation and apoptosis. Dysregulation and overproduction of TNFA have been implicated in a variety of human diseases including cancer. When chronically produced, TNFA may act as an endogenous tumour promoter, contributing to the tissue remodelling and stromal development necessary for tumour growth and spread.[81] Overexpression of TNF confers invasive properties on some tumour cell lines.[263] In breast cancer, infiltrating macrophages are a major source of TNF, which may regulate an angiogenic enzyme in the tumour epithelium.[264]

The TNFA gene is a candidate predisposing gene, however its location within the MHC, a highly polymorphic region encoding numerous genes involved in immunological responses, makes its association difficult to study.[265]

2.3.4 Tumour Suppressors

A tumour suppressor gene, or anti-oncogene, is a gene whose product reduces the probability that a cell in a multicellular organism will turn into a tumour cell. If a tumour suppressor gene becomes mutated causing a loss or reduction in its function it can contribute to the development of cancer. This usually occurs in combination with other genetic changes. The proteins coded for by the tumour suppressor genes may act through a variety of mechanisms. They may have a repressive effect on genes essential for the continuing of the cell cycle, couple the cell cycle to DNA damage in the cell or promote apoptosis if DNA damage cannot be repaired. Some cell adhesion proteins may prevent

tumour cells from dispersing by promoting contact inhibition, consequently loss of contact inhibition will promote metastasis of tumour cells.

2.3.4.1 Tumour Suppressor Protein p53 (TP53)

TP53 is the gene encoding the tumour suppressor p53. p53 mediates cell growth arrest, senescence and apoptosis in response to DNA damage, chemotherapeutic drugs, UV light, hypoxia and aberrant growth (signals from oncoproteins) by blocking progression through the cell cycle. Therefore polymorphisms in this gene are good candidates for genetic susceptibility to breast cancer. The role of TP53 mutations in carcinogenesis is shown by the high incidence of early onset cancers, including breast cancer, in patients with Li-Fraumeni Syndrome, a rare familial disease resulting from germline mutations in the TP53 gene. TP53 is mutated or deleted in many types of tumour[266;267] and is the most commonly mutated tumour suppressor gene in human cancers[268] suggesting it has a crucial role in the carcinogenesis of many tumours. Mutated p53 is found in approximately 20% of all breast tumours[269] and inactivating mutations in the TP53 gene cause a loss in negative regulation on growth and proliferation, giving rise to an increase in cell proliferation. One SNP, R72P (rs1042522) in exon 4 of the gene appears to have a significant association with the risk of breast cancer[270] and has frequently been studied and has been suggested as a prognostic marker in breast cancer[271;272] and associated with low grade histology.[273] However a recent pooled analysis by the Breast Cancer Association Consortium (BCAC) found strong evidence against there being an association between the R72P SNP and breast cancer.[39]

2.3.4.2 Mouse Double Minute-2 (MDM2)

The human homologue of the mouse double minute 2 gene is a nuclear phosphoprotein that acts as a negative regulator of p53, by binding p53 and inhibiting p53 dependant transcription, and can enhance its degradation. It has p53-independent effects on DNA double strand break repair.[274-276] MDM2 has been shown to be upregulated in approximately 40% of breast cancers. [277] One well studied SNP in MDM2 is the SNP IVS1+309 (rs2279744). This SNP is in the intronic promoter of MDM2 that is utilized by

both the p53 and ras pathways to activate MDM2 transcription. This SNP was first analysed by Bond et al[278;279]who found this SNP was associated with the early onset of different cancers. The minor allele of this SNP was shown to have an increased affinity for Sp1, a transcriptional activator, causing an increase in MDM2 transcription and therefore higher levels of the MDM2 protein and increased inhibition of p53 followed by a subsequent acceleration in tumour formation. Boersma et al found that there was a strong association between this SNP and the p53 status of a tumour which appears to modify the association between p53 status and breast cancer survival.[280] In contrast to this, Milikan et al[281] conducted a population based case-control study of African-Americans and Whites in North Carolina and found that there was no association with this SNP and breast cancer risk, as did Ma et al[282] who investigated this SNP in a Chinese population and Wilkening et al found that this SNP alone did not have any effect on the risk or the age of onset of inherited breast cancer.[283]

2.3.5 DNA Replication and Repair

DNA damage can occur through both normal metabolic activities and environmental factors such as UV light and radiation resulting in as many as 1 million individual molecular lesions per cell per day. The damage to the DNA molecule may alter or abolish the cell's ability to transcribe the gene that the affected DNA encodes or induce potentially harmful mutations in the cell's genome, which may affect the survival of its daughter cells after it undergoes mitosis. DNA damage may also occur through mistakes in replication. Most DNA replicates with fairly high fidelity, however mistakes do occur, with polymerase enzymes sometimes inserting the wrong nucleotide or too many or too few nucleotides into a sequence during replication. Most of these errors are fixed through various DNA repair processes with repair enzymes recognizing structural imperfections between improperly paired nucleotides and excise the incorrect bases and put the correct bases back in their place. Some replication errors make it past these mechanisms and become permanent mutations. Therefore the DNA repair process is crucial to maintaining the integrity of the genome and when normal repair processes fail and apoptosis does not

occur, irreparable DNA damage may occur. If the genes encoding the DNA repair enzymes themselves become mutated, mistakes begin accumulating at a much higher rate and can lead to cancer.

Inherited mutations that affect DNA repair genes are strongly associated with high cancer risks in humans, for example hereditary nonpolyposis colorectal cancer (HNPCC) is strongly associated with specific mutations in the DNA mismatch repair pathway and mutations in BRCA1 and BRCA2, both associated with a large number of DNA repair pathways, confer an increased risk of breast cancer.

BRCA1 regulates the cell cycle by keeping cells from growing and dividing too rapidly or in an uncontrolled way. The protein made by the BRCA1 gene is directly involved in the repair of damaged DNA through interaction with the protein produced by the RAD51 gene in the nucleus of many cell types to mend breaks in DNA.

2.3.5.1 Ataxia Telangiectasia Mutated Gene (ATM)

ATM plays a crucial role in maintaining the integrity of the genome. It is activated in response to DNA double strand breaks caused by ionising radiation. Once activated it triggers the phosphorylation of various proteins promoting cell cycle arrest and activating repair of the damaged DNA. It is mutated in a rare, recessive disorder, Ataxia Telangiectasia, which is characterised by progressive neurodegeneration, cell cycle check point defects, radiosensitivity and an increased predisposition to cancer.[284-288] There is an increased risk of breast cancer in relatives of patients with Ataxia Telangiectasia (AT)[284;289] and mutations in the ATM gene may cause an increased risk of developing tumours with a poor prognosis.[290] Several SNPs in ATM, including rs1800054 (S49C) and rs1800056 (F858L), have been shown to be associated with breast cancer susceptibility[291], however results are contradictory.[292-294] This may be because the many common variants that are found in the ATM gene have only a small increase in risk of breast cancer which is not detectable in most case control studies due to small sample size.[295]

2.3.5.2 BRCA1

The BRCA1 gene encodes a nuclear phosphoprotein that plays a role in maintaining genomic stability.[296] It acts as a transcription factor and plays a role in DNA repair of double strand breaks and recombination.[297] Mutations in the BRCA gene are responsible for approximately 40% of inherited breast cancers and more than 80% of breast and ovarian cancers. It has been indicated that BRCA1 may normally serve as a negative regulator of mammary epithelial cell growth and that this function is compromised in breast cancer either by direct mutation or by alterations in gene expression.[298] Several SNPs have been identified in BRCA1 that alter susceptibility to both breast and ovarian cancer.[299-301]

2.3.6 Metabolising Genes

Metabolism is a term used to describe all the chemical processes occurring within cells. Some of these reactions involve the synthesis of larger molecules from smaller ones (anabolism) and some involve the breakdown of larger molecules into simpler ones (catabolism) and are aided by a complicated network of enzyme- catalyzed reactions that occurs in cells.

Several breast cancer risk factors are thought to act by influencing lifetime exposure to steroid hormones, for example age at menarche, age at menopause, postmenopausal obesity and postmenopausal hormone use which all influence the dose and duration of oestrogen and progesterone exposure. Endogenous steroid hormones exert growth-promoting effects and induce breast cell proliferation by binding to intracellular receptors and regulating gene transcription and are therefore important in the development and progression of breast cancer. Oestrogens are metabolized by a number of oxidative and conjugate reactions that lead to their deactivation and subsequent elimination and inherited alterations in the activity of any of these enzymes hold the potential to define differences in breast cancer risk associated with oestrogen carcinogenesis.

Alcohol intake is also associated with increased breast cancer risk. Ethanol is mainly oxidised in the liver by alcohol dehydrogenases to acetaldehyde, a weak mutagen, and is further detoxified to acetate by aldehyde dehydrogenases. The metabolism of ethanol also releases reactive oxygen species that may induce oxidative damage. Alcohol interferes with oestrogen pathways in multiple ways, influencing hormone levels and effects on the oestrogen receptors. Functional variants in genes involved in alcohol metabolism result in differences between individuals in exposure to carcinogenic acetaldehyde, suggesting a possible interaction of genetic susceptibility and alcohol exposure in cancer.[302;303]

2.3.6.1 Cytochrome P450 1B1 (CYP1B1)

CYP1B1, a monoxygenase, is a major enzyme catalysing the hydroxylation of oestrogens into the genotoxic catechol oestrogen 4-hydroxy-oestradiol, which has been shown to induce uterine adenocarcinoma[304] and can generate potentially mutagenic free radicals that induce single strand breaks in DNA.[305] CYP1B1 is expressed in various tissues in the body, but expression is particularly high in the breast, supporting a role in hormone mediated breast cancer[306] and is genetically polymorphic, which may account for some of the genetic susceptibility to breast cancer risk.[307] For example the Val432Leu (rs1056836) polymorphism, which has been associated with oestrogen mediated carcinogenicity, having a decreased catalytic efficiency for 4-hydroxylation of oestradiol as well as alterations in enzymic activity towards various other mammary carcinogens.[308-310] This SNP has also been associated with oestrogen receptor positive breast cancer in Caucasian patients[311], breast cancer susceptibility in Turkish women, especially those with a BMI (body mass index) greater than 24kg/m^2 [312] and a higher risk of breast cancer in smokers with the Valine allele.[313;314] In contrast, there have been several studies that have reported no significant associations of CYP1B1 polymorphisms and breast cancer[306;311;315] including a large Swedish case control study.[316]

2.3.6.2 Alcohol Dehydrogenase-1B (ADH1B)

Epidemiological studies have identified alcohol consumption as a risk factor for breast cancer[317-320]. The first step in the detoxification of ethanol is oxidation and this can contribute to oxidative cellular and DNA damage, which can contribute to carcinogenesis[321] and its metabolism results in the production of an acetaldehyde, a reactive electrophile that can act as a weak mutagen and carcinogen[322]. Alcohol is metabolised by several mechanisms, one of which involves ADH, alcohol dehydrogenase, which is expressed in the mammary epithelium and its expression is significantly reduced in invasive breast cancer.[323] Genetic susceptibilities in genes affecting ethanol metabolism can affect breast cancer risk and the polymorphism G47A may modify the risk associated with high levels of alcohol consumption.[324]

2.3.7 Apoptosis Related Genes

Apoptosis, or programmed cell death, is the process by which cells of multicellular organisms use specialized cellular machinery to destroy itself and is used to control cell number and eliminate cells that threaten the organisms survival. Apoptosis involves a series of biochemical events that lead to a variety of morphological changes, including changes to the cell membrane such as loss of membrane asymmetry and attachment, cell shrinkage, nuclear fragmentation, chromatin condensation and chromosomal DNA fragmentation. An insufficient amount of apoptosis results in uncontrolled cell proliferation, such as cancer.

2.3.7.1 Death Associated Protein Kinase 1 (DAPK1)

DAPK1 is a calcium/calmodulin dependant enzyme that phosphorylates serine/threonine residues on proteins (serine/threonine kinase).[325] It is involved in an intrinsic p53 dependant apoptotic checkpoint which is designed to eliminate premalignant cells from cancer development[326], reacting to various stimuli including INFγ, TNFα, Fas, C6-ceramide, oncogene expression (such as c-Myc and p53) and detachment from the extracellular matrix.[325;327-330] Due to its proapoptotic actions it is known as a metastasis suppressor and its expression frequently lost in human tumours.[326] Loss of DAPK expression has been

associated with highly aggressive, metastatic behaviour in tumours.[325;328] Transcriptional silencing by hypermethylation of CpG islands in the promoter is a common event in the multistep process of tumour metastasis[331] and has been associated with an invasive phenotype in breast cancer.[332;333] DAPK1 has been shown to play a role in uncoupling of stress fibres and focal adhesions, leading to a perturbation of the balance between contractile and adhesion forces and subsequent cell detachment, which might contribute to its pro-apoptotic activity[334]. The single nucleotide change c.1-6531A>G enhances the binding affinity of transcriptional factor HOXB7 and results in altered protein binding, causing an increased affinity for HOXB7 that downregulates DAPK1 transcription[335].

2.3.7.2 Caspase 8 (CASP8)

Caspases play an important role in apoptosis, which results from a cascade of protease reactions carried out by the caspases (aspartate specific cysteine proteases). Caspase 8 is activated in response to DNA damage or external death signals. A lack of functional caspase 8 results in resistance to TNFα induced apoptosis.[336] The rare H allele of D302H (rs1045485), an aspartic acid to histidine substitution, was associated with a reduced risk of breast.[337;338] BCAC (Breast Cancer Association Consortium), a large consortium established to conduct combined case control analysis with augmented statistical power to attempt to confirm putative genetic associations with breast cancer, typed nine SNPs, all with a previous association to breast cancer, and the strongest association found was with the D302H allele in the CASP8 gene with the D302H polymorphism having a protective effect on breast cancer risk[38].

2.3.8 Invasion and Metastasis Associated Genes

Tumour invasion and development of metastasis are the primary determinants of patient outcome and, consequently, molecules involved in these processes are obvious candidates to be identified as new prognostic markers in breast cancer. Metastasis requires several sequential steps, such as changes in cell-ECM interaction, disconnection of intercellular adhesions, separation of single cells from the solid tumour tissue, degradation of ECM, locomotion of tumour cells into the ECM, invasion of lymph and

blood vessels, evasion of the immune system, adhesion to endothelial cells, extravasation from lymph and blood vessels, proliferation of cells and induction of angiogenesis[339] involving many different genes.

2.3.8.1 Fibronectin 1 (FN1)

The FN1 gene encodes fibronectin, an extracellular matrix glycoprotein of high molecular weight (430kDa), that plays a role in cell adhesion and morphology including processes such as embryogenesis, wound healing, blood coagulation, host defence and metastasis. It binds to cell membranes via a β1 integrin receptor, preventing apoptosis. This interaction is also essential for cell proliferation in normal cells, however it is not necessary for proliferation of cancer cells. It does, however, prevent apoptosis in cancer cells[340]. Previous studies have shown an association between FN1 expression and tumour progression and a shorter overall survival.[341;342] It has been associated with an unfavourable prognosis in ovarian cancer[343] and has been reported to be upregulated by oestrogens in breast cancer cells[344] and downregulated in HER-2 transfected cells[345].

2.3.8.2 Integrin Binding Protein 4 (ITGB4)

Integrins are transmembrane glycoprotein receptors that mediate cell-matrix or cell-cell adhesions and transduce signals regulating gene expression and cell growth.[346] They are heterodimers, comprising a non-covalently associated α and β subunit, and have different ligand binding specificities depending on their combination of α and β subunits. ITGB4 encodes the integrin β4 subunit and has specificity for the laminins. The ITGB4 subunit commonly associates with the alpha 6 subunit and is likely to play a pivotal role in the biology of invasive carcinoma, as it was demonstrated in a breast cancer cell line that the α6β4 integrin promotes carcinoma invasion through a preferential and localised targeting of phosphoinositide-3 OH kinase (PI3K) activity and it was shown that α6β4 regulated PI3K activity is required for the formation of lamellae (dynamic sites of motility) in cancer cells.[347] Dajee et al[348] demonstrated that laminin-5 and ITGB4 are both required for human squamous cell carcinoma tumourigenesis. Murgia et al[349] found that the interactions mediated by the beta-4 cytoplasmic domain are vital for adhesion of epithelia

to the basement membrane and for proper cell cycle control of proliferation and Guo et al[350] investigated the effect of deletion of the cytoplasmic tail of the beta-4 subunit, finding that it promoted tumour progression by amplifying ERBB2 signalling.

2.3.8.3 LSP1

The LSP1 gene encodes a protein involved in leukocyte recruitment. The SNP rs3817198 in intron 10 of the LSP1 gene was recently identified by Easton et al[21] in a genome wide association study as being associated with breast cancer susceptibility and Huijts et al[199] investigated the association of SNPs with breast cancer phenotype and found that the rs3817198 SNP was associated with lymph node positive breast cancer, however it had no effect on overall or disease free survival.

2.3.8.4 Matrix Metalloproteinases (MMPs)

The MMPs are a family of structurally and funtionally related zinc dependant endoproteinases. Their primary function is degradation of the extracellular matrix (ECM) and they play an important role in tissue remodelling, being associated with various physiological and pathological processes such as morphogenesis, angiogenesis, tissue repair, cirrhosis, arthritis and invasion and metastasis[351;352]. There are currently 26 known members of this family[353]. The MMPs are commonly divided into four groups based on their substrate specificity and domain organisation. The four groups are; interstitial collagenases, gelatinases, stromelysins and the membrane type MMPs. The interstitial collagenases degrade collagen types I, II and III, the major components of bone and cartilage. The gelatinases (also known as the type IV collagenases) degrade gelatin (denatured collagen) and types IV, V, VII, IX and X collagen. Type IV collagen is abundant in the basement membrane giving these MMPs a crucial role in invasion and metastasis[354;355]. Two members of this group have been identified; gelatinase A (MMP-2) and gelatinase B (MMP-9). The stromelysins have a broad substrate specificity and catalyse the degradation of many ECM substrates, including proteoglycans, and non collagenous proteins such as laminin and fibronectin. One of the main members of this group is matrilysin (MMP-7). The membrane type group of MMPs have a transmembrane

domain. There are five members of this group, the best characterised being the membrane type I MMP. There are however some MMPs that do not appear to fit into any of these categories and others that may apply to more than one group[356]. MMPs may play a role in breast cancer initiation[121;357;358], as well as angiogenesis, invasion and metastasis[355;357;359;360]. Inhibition of MMPs has been shown to decrease tumour growth.[361;362]

Many studies have implicated an involvement of MMP-2, MMP-7 and MMP-9 in breast cancer. MMP-2 is up-regulated in many tumour types and is associated with prognosis in most of them[363]. MMP-2 expression and activity levels have been associated with breast tumour progression [364], as well as with breast cancer phenotype[365] including large tumour diameter[366], high grade[367] lymph node involvement[364;368-373], bone and visceral metastases[373] and high expression of oestrogen receptors in malignant tumours[369] as well as a shortened overall survival independent of major prognostic indicators[366;374;375]. Expression of both MMP-2 and MMP-9 is abundant in various malignant tumours[376] and has been linked to metastasis of breast cancer to the brain[60] and correlated with poor disease free survival in lymph node negative patients[377]. Expression and activity levels of MMP-9, a previously reported candidate metastasis associated gene[378], are elevated in malignant breast tumours[369], its expression has been associated with the infiltrating lobular breast cancer phenotype[364] and is a useful marker in both the prognosis and of the follow up of breast cancer patients, with an increased level being associated with worse prognosis disease relapse.[379] MMP-7 has been associated with the progression of many different types of cancer[380-383] including breast cancer[384], its expression has been shown to contribute to early stage mammary tumourigenesis.[358]

Several SNPs in the MMP genes have been investigated in relation to cancer. A novel functional genetic variant in the MMP-2 promoter (rs17859818) was identified in 2001[385]. This SNP abolishes the Sp1 binding site causing lower transcriptional activity. It has since been associated with risk of lung cancer[386], development of gastric cardia adenocarcinoma[387], risk of occurrence and metastasis of oesophageal cancer[388], colorectal

cancer development and invasion[389], development and aggressiveness of head and neck squamous cell carcinoma[390] and breast cancer susceptibility and progression[391;392]. This SNP has been found to have different effects on survival depending on the patient's oestrogen receptor (ER) status, with the ER negative patients having a poor survival and the ER positive patients having good survival rates.[365] Another SNP in the MMP-2 gene, g-1575A, was identified that affects the oestrogen receptor α binding site. This leads to reduced transcriptional activity as the oestrogen receptor is essential for enhanced activity.[393] A polymorphism in the MMP-9 promoter (rs3918242) is related to the severity of coronary athersclerosis[394] and has since been linked to good prognostic features in breast cancer[365].

2.3.8.5 RHOC

The RHOC gene encodes a member of the Rho family of small GTPases. RhoC (Ras homolog gene family, member C) is a small (~21 kDa) signalling G protein (more specifically a GTPase) and is a member of the Rac subfamily of the Rho family of GTPases. It functions as a molecular switch in signal transduction cascades. Rho proteins promote reorganization of the actin cytoskeleton and regulate cell shape, attachment and motility. Overexpression of this gene is associated with tumour cell proliferation and metastasis[395].

2.3.9 Gene Desert

An LD block of approximately 110kb, containing no known genes, on chromosome 8q24 has been shown to contain a SNP associated with breast cancer susceptibility.[21]

2.4 SNP Selection

260 SNPs (200 plus an additional 30% as requested by Sequenom in case of in silico design failure) were selected from the 30 candidate genes.

The SNPs were selected for one of four reasons:

1. Previously reported (identified from extensive literature searching)
2. Non synonymous – identified using the UCSC database (Section 2.2.1)
3. Promoter region – SNPs in the region 5kb upstream (5') on the gene
4. LD location – The remainder of SNPs were selected based on their LDU location to ensure coverage of the gene with one SNP selected at least every 0.5 LDU ($r^2 \geq 0.9$ (r-squared between adjacent SNPs in candidate genes) using data from the HapMap project (Section 2.2.2) [161])

All SNPs (other than those previously reported) had been previously reported in the HapMap CEU population (population with ancestry from northern and western Europe). Previously reported SNPs, promoter SNPs and those selected by LDU location had a minor allele frequency (maf) >5% and non synonymous SNPs had a maf>1%.

Table 2: 206 Selected SNPs

Gene	*Chromosome*	*SNP*	*Type*	*location (bp)*	*P*	*OR (95% C.I.)*	*Reference*	*Tested for association*
BRCA1	17q21.31	rs8176318	3' UTR	38,450,800				Yes
		rs8176265	Intron	38,467,522				Yes
		rs3737559	Intron	38,487,830				Yes
		rs16940	Exon synonymous	38,498,763				Yes
		rs799905	Intron	38,530,713				Yes
		rs11653069	Promoter	38,536,903				No
TP53	17p13.1	rs12951053	Intron	7,518,132	0.00009	1.29 (1.06-1.58) ER-ve	Garcia-Closas et al. (2007)	No
		rs1625895	Intron	7,518,840	NA	0.49 (0.27-0.90) Smokers	Gaudet et al. (2008)	No
		rs1042522	Exon non-synonymous	7,520,197	NA	1.32 (1.00–1.74)	Gaudet et al. (2008)	Yes
		rs12602273	Intron	7,523,738				No
		rs2287497	Promoter	7,533,505				Yes
ATM	11q22.3	rs228606	Promoter	107,593,057				Yes
		rs4987876	Promoter	107,597,847		1.21 (0.64-2.29)	Baynes et al. (2007)	Yes
		rs1800054	Exon non-synonymous	107,603,786	0.08	1.13 (0.99-1.30)	Cox et al. (2007)	No
					0.004	1.69 (1.19-2.40)	Stredrick et al. (2006)	
		rs599164	Intron	107,625,649	NA	1.05 (0.88-1.24)	Baynes et al. (2007)	No
		rs4986761	Exon non-synonymous	107,629,971	0.03	0.47 (0.23-0.93)	Stredrick et al. (2006)	No
		rs1800056	Exon non-synonymous	107,643,213	0.03	2.03 (1.05-3.90)	Stredrick et al. (2006)	No
		rs3092991	Intron	107,645,726	NA	1.12 (0.71-1.77)	Einarsdottir et al. (2006)	Yes
		rs1800058	Exon non-synonymous	107,665,560				No
		rs1800889	Exon synonymous	107,668,697	0.04	0.75 (0.49-1.13)	Stredrick et al. (2006)	Yes
		rs1801516	Exon non-synonymous	107,680,672	0.04	0.63 (0.27-1.49)	Edvardsen et al. (2007)	Yes

DAPK1	9q21.33	rs975256	Promoter	89,296,902				No
		rs2058882	Intron	89,304,566				No
		rs17399090	Intron	89,324,847				Yes
		rs10746815	Intron	89,327,937				Yes
		rs4878086	Intron	89,329,724				No
		rs1045042	Intron	89,340,827				Yes
		rs1041326	Intron	89,346,666				Yes
		rs7038971	Intron	89,361,235				No
		rs928114	Intron	89,368,626				No
		rs11141901	Intron	89,380,485				Yes
		rs11141904	Intron	89,386,866				No
		rs7043056	Intron	89,392,167				No
		rs13288504	Intron	89,410,000				No
		rs7033344	Intron	89,426,667				Yes
		rs11141918	Intron	89,436,202				Yes
		rs943855	Intron	89,460,147				No
		rs13285155	Intron	89,470,700				Yes
		rs9410612	Intron	89,474,570				Yes
		rs1007394	Intron	89,482,445				Yes
		rs12056997	Intron	89,506,425				Yes
		rs1056719	Intron	89,511,843				Yes
CASP8	2q33.1	rs12693932	Promoter	201,801,640				Yes
		rs2037815	Intron	201,809,960				No
		rs6760993	Intron	201,819,162				Yes
		rs12990906	Intron	201,822,869				Yes
		rs10931936	Intron	201,852,173				No
		rs1045485	Exon non-synonymous	201,857,834	5.7x10-7	0.74 (0.62-0.87)	Cox et al. (2007)	No
					NA	0.58 (0.39-0.88)	MacPherson et al. (2004)	
					NA			
					NA	0.49 (0.17-1.44)	Frank et al. (2005)	
IGF-1	12q23.2	rs2946834	Intron	101,311,944	0.0004		Tamimi et al. (2007)	Yes
		rs5742714	Intron	101,313,982				Yes

IGF-1		rs1520220	Intron	101,320,652	0.03	1.41 (1.11-1.79)	Ali Al-Zahrani et al. (2006)	Yes
		rs2373721	Intron	101,351,175				Yes
		rs5742632	Intron	101,380,604				Yes
		rs12821878	Intron	101,391,797				Yes
		rs5742620	Intron	101,393,730				No
		rs2162679	Intron	101,395,389				Yes
		rs35766	Promoter	101,404,603				Yes
ERBB2	17q12	rs903501	Promoter	35,093,019				Yes
		rs1801200	Exon non-synonymous	35,133,114		0.68 (0.47-0.98)	Cox et al. (2005)	Yes
		rs28933370	Exon non-synonymous	35,134,904				No
		rs1058808	Exon non-synonymous	35,137,563				Yes
TGFβ1	19q13.2	rs8179181	Intron	46,530,046				Yes
		rs11466338	Intron	46,537,641				Yes
		rs4803455	Intron	46,543,349				No
		rs1800469	Intron	46,552,136	NA 0.009	0.31 (0.27-0.35) 1.25 (1.06-1.48)	Grainger et al. (1999) Dunning et al. (2003)	No
		rs1982072	Promoter	46,556,349				Yes
FGFR2	10q26.13	rs1649202	Intron	123,230,615				No
		rs3135811	Intron	123,232,714				Yes
		rs2912796	Intron	123,238,141				Yes
		rs2912762	Intron	123,266,280				Yes
		rs11199993	Intron	123,281,254				Yes
		rs2912787	Intron	123,315,528				Yes
		rs2981428	Intron	123,319,419				Yes
		rs1219648	Intron	123,336,180	1.1x10-10	1.64 (1.42-1.90)	Hunter et al. (2007)	Yes
		rs2981582	Intron	123,342,307	2x10-76 0.00045	1.63 (1.53-1.72) 1.24 (1.11-1.38)	Easton et al. (2007) Antoniou et al. (2008)	Yes
		rs4412700	Promoter	123,352,565				No
		rs4752566	Promoter	123,257,621				Yes

VEGF	6p21.1	rs699946	Promoter	43,840,647				No
		rs699947	Promoter	43,844,367	0.03	1.99 (1.06-3.74)	Schneider et al. (2007)	Yes
		rs833061	Promoter	43,845,464		1.5 (0.9-2.5)	Lu et al. (2005)	No
		rs833069	Intron	43,850,557				Yes
		rs3024997	Intron	43,853,085				No
		rs3025033	Intron	43,859,053				Yes
		rs10434	3'UTR	43,861,190				Yes
MAP3K1	5q11.2	rs889312	Upstream	56,067,641	7x10-20 0.044 0.02	1.27 (1.19-1.36) LN-ve 1.12 (1.02-1.24) BRCA2	Easton et al. (2007) Huijts et al. (2007) Antoniou et al. (2008)	Yes
		rs17661089	Promoter	56,141,753				Yes
		rs1423622	Intron	56,149,199				Yes
		rs16886383	Intron	56,161,453				No
		rs252921	Intron	56,173,544				No
		rs832574	Intron	56,195,639				Yes
		rs702689	Exon non-synonymous	56,213,200				Yes
IL-6	7p15.3	rs2056576	Promoter	22,727,727				No
		rs1800795	Promoter	22,733,170	0.02	2.0 (1.1-3.6)	Hefler et al. (2005)	Yes
		rs2069832	Intron	22,733,958	0.01	4.4 (1.5-12.8)	Slattery et al. (2007)	No
		rs1474348	Intron	22,734,433				Yes
		rs2069845	Intron	22,736,674				Yes
		rs2069860	Exon non-synonymous	22,737,563				No
TNFα	6p21.33	rs2009658	Promoter	31,646,223				Yes
		rs3093662	Intron	31,652,168				Yes
LSP1	11p15.5	rs907613	5'UTR	1,830,868				Yes
		rs2089910	Exon synonymous	1,830,980				No
		rs599774	Intron	1,835,212				Yes
		rs592373	Intron	1,847,566				Yes
		rs661348	Intron	1,861,868				Yes

LSP1		rs3817198	Intron	1,865,582	3x10-9	1.17 (1.08-1.25)	Easton et al. (2007)	Yes
		rs548195	3'UTR	1,869,953				No
TOX3	16q12.1	rs2075236	Intron	51,037,272				No
		rs1420542	Intron	51,039,776				Yes
		rs1420546	Intron	51,046,208				Yes
		rs1362560	Intron	51,058,858				No
		rs8051542	Intron	51,091,668	10-12	1.19 (1.12-1.27)	Easton et al. (2007)	Yes
		rs12443621	Intron	51,105,538	2x10-19	1.23 (1.17-1.30)	Easton et al. (2007)	No
		rs3803662	Exon synonymous	51,143,842	10-36	1.39 (1.26-1.45)	Easton et al. (2007)	Yes
					0.00027	1.28 (1.11-1.46)	Antoniou et al. (2008)	
ADH1B	4q23	rs1042026	3' UTR	100,447,489	0.044	1.04 (0.95-1.14)	Cox et al. (2007)	Yes
		rs1789882	Exon synonymous	100,454,076				Yes
		rs1353621	Intron	100,460,598				No
		rs1789888	Promoter	100,466,593				Yes
CYP1B1	2p22.2	rs162549	3' UTR	38,148,960				No
		rs1800440	Exon non-synonymous	38,151,643	NA	1.3 (0.3-4.8)	Renee bailey et al. (1998)	Yes
						1.61 (0.96-2.70) risk 2.82 (1.37-5.82) ER-ve	Justenhoven et al. (2007)	
		rs1056836	Exon non-synonymous	38,151,707				Yes
		rs2551188	Intron	38,156,298				No
		rs10175368	Promoter	38,161,365				Yes
AR	Xq12	rs17302090	Promoter	66,672,128				Yes
		rs7061037	Intron	66,738,888				Yes
		rs6624304	Intron	66,792,481				Yes
		rs12011518	Intron	66,849,761				No
		rs12014709	Intron	66,855,191				No
		rs9332969	Exon non-synonymous	66,859,466				No
ESR1	6q25.1	rs2077647	Exon synonymous	152,170,770				No
		rs827423	Intron	152,197,890				No

ESR1		rs6557170	Intron	152,244,797				Yes
		rs4870062	Intron	152,279,311				Yes
		rs9371564	Intron	152,303,806				Yes
		rs3020410	Intron	152,308,070				Yes
		rs3020314	Intron	152,312,365				Yes
		rs1884051	Intron	152,324,972				No
		rs2982700	Intron	152,328,216				Yes
		rs3020403	Intron	152,337,408				Yes
		rs2144025	Intron	152,349,399	0.001		Einarsdóttir et al. (2008)	Yes
		rs12212176	Intron	152,351,700				No
		rs6905370	Intron	152,367,890				Yes
		rs2207232	Intron	152,381,981				Yes
		rs2982735	Intron	152,410,380				Yes
		rs9383607	Intron	152,422,011				Yes
		rs9397484	Intron	152,435,930				No
		rs9341052	Intron	152,458,318				Yes
		rs9341066	Intron	152,461,219				Yes
		rs2228480	Coding synonymous	152,461,788				Yes
		rs3798577	3'UTR	152,462,823	0.047		Zhang et al. (2008)	Yes
TWIST	7p21.2	rs2285681	Promoter	19,125,291				Yes
		rs2189000	Promoter	19,127,743				Yes
GATA3	10p14	rs1244181	Promoter	8,131,383				No
		rs1399180	Intron	8,138,725				No
		rs3802604	Intron	8,142,278	NA	0.9 (0.76-1.06) risk	Garcia-Closas et al. (2007)	Yes
					0.02	0.72 (0.54-0.96) ER-ve		
		rs570613	Intron	8,146,508	0.004	0.82 (0.69-0.96) risk	Garcia-Closas et al. (2007)	No
				8,151,415	0.006	0.71 (0.54-0.94) ER-ve		
		rs422628	Intron	8,158,099		0.97 (0.77-1.24)		Yes
		rs263419	Intron	113,048,650				Yes

RHOC	1p13.2	rs2999156	Intron	113,050,314			Yes
		rs12144044	Intron	113,055,790			Yes
		rs7522283	Promoter	100,414,396			Yes
PGR	11q22.1	rs11224563	3'UTR	100,418,375			Yes
		rs606789	Intron	100,438,622			Yes
		rs1042838	Exon non-synonymous	100,442,270	1.06 (0.97-1.15)	Johnatty et al. (2008)	Yes
		rs11224579	Intron	100,473,333			Yes
		rs694070	Intron	100,500,919			No
		rs506487	Intron	100,503,981			Yes
		rs3740753	Exon non-synonymous	100,505,711			No
		rs518162	5'UTR	100,513,712			Yes
		rs4754732	Promoter	71,218,460			Yes
ITGB4	17q25.1	rs820227	Promoter	71,232,322			No
		rs3862481	Intron	71,247,907			No
		rs2290458	Intron	71,265,098			Yes
		rs871443	Exon non-synonymous	215,940,366			No
FN1	2q35	rs1250214	Intron	215,944,957			No
		rs17449032	Exon non-synonymous	215,976,689			No
		rs7572169	Intron	215,988,767			No
		rs10207245	Intron	216,015,348			Yes
		rs1250233	Promoter	67,493,429			No
MDM2	12q15	rs1470383	Intron	67,502,788			Yes
		rs3730556	Intron	67,518,247			No
		rs3730646	Intron	54,062,829			No
MMP-2	16q12.2	rs16955194	Promoter	54,066,374			Yes
		rs1116195	Promoter	54,069,307			Yes
		rs243865	Promoter	54,074,209			No
		rs857403	Intron	54,081,206			No
		rs243849	Exon synonymous	54,085,183			Yes

MMP-2		rs183112	Intron	54,089,186				No
		rs17859981	Intron	54,089,959				No
		rs2287076	Intron	54,094,264				Yes
		rs11541998	Exon synonymous	54,097,115				Yes
		rs7201	3'UTR	101,903,803				Yes
MMP-7	11q22.2	rs10502001	Exon non-synonymous	101,905,546				Yes
		rs11225309	Intron	101,906,843				Yes
		rs11568819	Promoter	101,912,401				Yes
		rs1943779	Promoter	101,914,290				Yes
		rs4754850	Promoter	101,919,669				No
		rs17352599	Promoter	44,062,697				Yes
MMP-9	20q13.12	rs6094237	Promoter	44,065,774				Yes
		rs6065912	Promoter	44,069,142				Yes
		rs3918241	Promoter	44,069,383				Yes
		rs3918242	Promoter	44,073,632		1.88 (0.97-3.63)	Lei et al. (2007)	No
		rs17576	Exon non-synonymous	44,075,813				Yes
		rs2250889	Exon non-synonymous	44,076,518				No
		rs2274756	Exon non-synonymous	44,076,999				No
		rs3918261	Intron	128,424,800				Yes
Desert	8q24.21	rs13281615	Genomic	1.28E+08	5x10-12	1.18 (1.101.25)	Easton et al. (2007)	Yes

2.5 Patient Selection

Patient material from the POSH study was used for our study.

2.5.1 POSH Study (Prospective Outcomes in Sporadic versus Hereditary breast cancer (POSH) study)[396]:

The POSH study is a study of 3000 patients with invasive breast cancer with an age of onset less than 41 years who have been recruited from across the UK. The study has been designed to give accurate information about hereditary factors that determine predisposition, prognosis and treatment responses in breast cancer. Blood samples and pathology blocks are available from all recruits. Family history for each recruit is recorded. The collection of genomic DNA and the meticulous documentation of clinical details about treatment and diagnosis, and a systematic pathological review will give the opportunity to investigate the role of the host genome in determining clinical outcomes, including tumour presenting characteristics and treatment responses. This provides a unique dataset for hypothesis generation, looking for evidence to be replicated by larger studies.

<u>Advantages of using this cohort:</u>

- Cases that have a young age of onset are more likely to be due to an inherent predisposition to the disease.
- Cases selected have no family history therefore are unlikely to have an underlying high risk genetic predisposition that could mask more subtle traits.
- Almost all cases are symptomatic (<2% screen detected) and so represent the true natural history of the tumour. Breast screening may alter the presenting features of disease which causes problems when studying the influence of the genome on the natural history of breast cancer.
- The systematic pathology review, prospective clinical follow up and the remarkably consistent treatment choices allows for associations to be made with various phenotypes as well as risk and will help interpret the outcome data.

- The large sample size gives good statistical power to detect any associations in data generated.

2.6 DNA Extraction

DNA was extracted from whole blood samples as documented below (3.3.1.1).

2.7 Genotyping

Genotyping of our samples was carried out by Sequenom, using tier 2 of their iPLEX Gold service based on single-base primer extension and MALDI-TOF mass spectrometry (Section 9.1.1.1). All of the 260 SNPs selected underwent in silico design and the top 210 were selected (ensuring there was still full coverage of each gene) for genotyping in the 1001 patients, of these 206 were successfully typed.

2.8 Quality Control

Missing genotype rates are an indicator of poor DNA quality. 58 individuals were removed from analysis as they had $\geq$ 10% missing genotypes, leaving 943 individuals. SNPs with over 10% data missing were identified and removed from analysis (60 SNPs). A further 12 SNPs were removed which had a minor allele frequency <5%.

Deviations from Hardy-Weinberg equilibrium in the control samples would typically be used to exclude further SNPs, however in this study, only cases were genotyped and therefore the Hardy Weinberg equilibrium could not be applied as deviations may be a result of an association with breast cancer, rather than reflecting poor genotyping. Therefore we noted any SNPs with significant deviations from Hardy-Weinberg equilibrium ($\chi^2 \geq 10$) and the genotyping calls were verified by examination of the Sequenom cluster plots for these SNPs (Section 9.3). On this basis, one additional SNP was removed. This left 133 SNPs (64%) for analysis which is consistent with the genotyping tier (Section 9.2).

2.8.1 Population Stratification

Population stratification refers to differences in allele frequencies between cases and controls due to systematic differences in ancestry rather than association of genes with disease. It can cause false positive results to occur in association studies as genetic variations could be due to the individuals being of different ethnic origin and not a true association with disease. The majority of individuals in the POSH cohort are of Western European ancestry therefore any individuals that appeared to be of different ancestry were removed. To identify these individuals we performed a multi-dimensional scaling analysis on genome-wide average identities by state on the cohort using PLINK. First the data was reduced to 82 autosomal SNPs with maximum pairwise $r^2 \leq 0.2$ to avoid confounding by including non-independent SNPs in strong linkage disequilibrium with eachother. We ensured that all genotypes were configured to the positive strand of the reference sequence (National Cancer for Biotechnology Information [NCBI] build 36.1) and then merged the data from the POSH cohort with individuals from the HapMap project[161] (60 Western European, 60 Nigerian, 45 Han Chinese and 45 Japanese). It is recommended that over 10000 autosomal SNPs are used for this analysis[156], however we found plotting the first two components (which represent geographic and genetic variation) from the multidimensional scaling analysis of the 82 SNPs produced 3 separate clusters with our data representing the African, Asian and Caucasian populations. We were therefore able to exclude 44 samples from the POSH cohort that were separate from the cluster of Caucasians (35 Africans and 9 Asians) leaving 899 patients (Figure 8). The call rate for the 133 SNPs in the remaining 899 individuals was 99%.

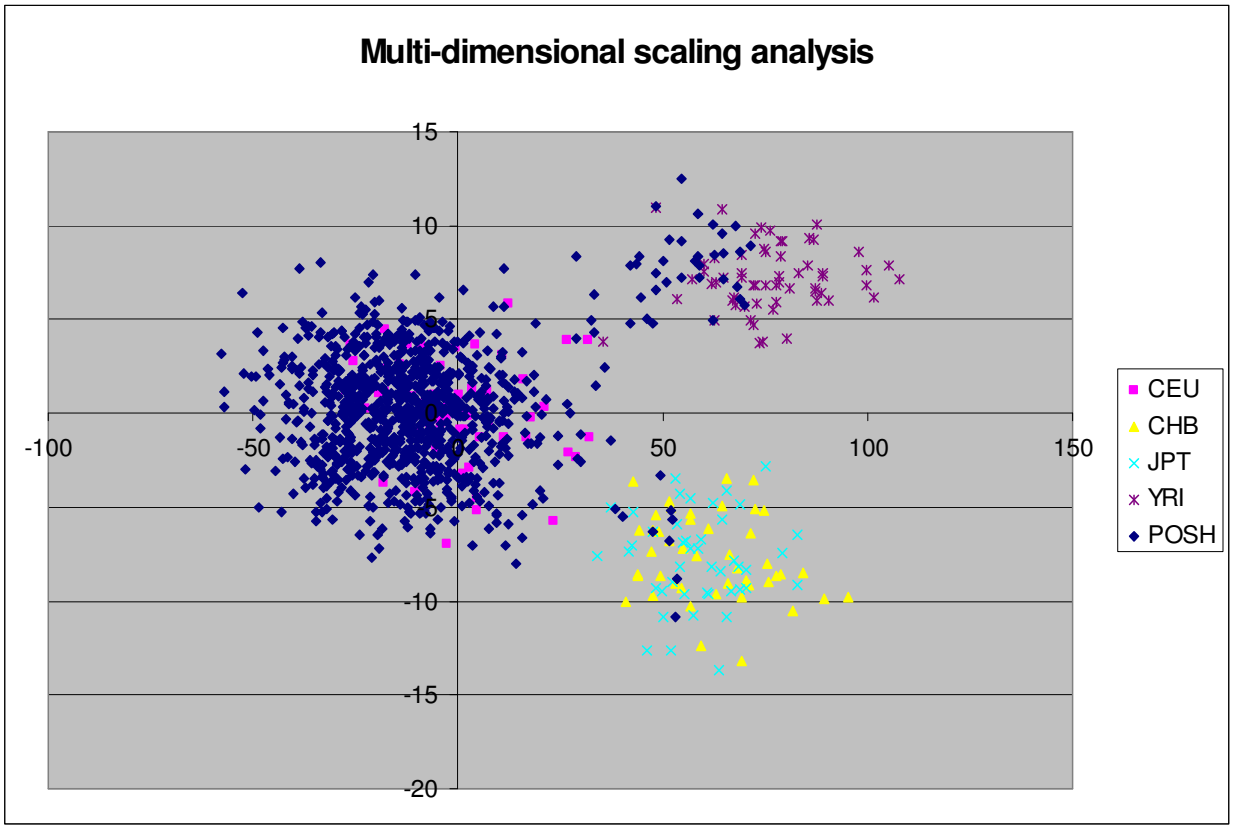


Figure 8: Multi-dimensional scaling analysis- Inference of ancestry by multidimensional scaling-POSH and HapMap samples plotted for the first two principal components obtained by multidimensional scaling of a matrix of pairwise identity by state (IBS) values. The blue diamonds occurring near the YRI and CHB+JPT clusters are the individuals excluded from the POSH cohort. - CHB, Chinese; CUE, Western European; JPT, Japanese; POSH, Prospective study of Outcomes in Sporadic versus Hereditary breast cancer; YRI, Nigerian.

2.8.2 Associations

2.8.2.1 Association with Survival

As the median time to follow up for the individuals passing QC in our study was only 2.4 years and most individuals developing distant metastases will die from the disease, we used distant disease-free survival, defined as the time between diagnosis and the first distant metastasis, in place of survival in our risk associations. Univariate analyses of DDFS (distant disease-free survival) were performed, by plotting Kaplan-Meier curves for each SNP and using the log rank test to compare Kaplan-Meier DDFS curves for the three genotypes of each SNP. Tests were performed using SPSS v17 (SPSS Inc., San Francisco, CA, USA). In seven cases the date of distant metastasis was not available and so the date of death was used. In one of these cases the cause of death was known to be unrelated to breast cancer.

2.8.2.2 Association with Breast Cancer Phenotypes

To identify SNPs associated with breast cancer biology the data was stratified into phenotypic subgroups. The phenotypes investigated were ER status, lymph node involvement, lymphovascular involvement, distant metastasis and grade. All phenotypes were divided on the basis of being positive or negative for that phenotype except for grade where the phenotypic extremes of grade 1 and grade 3 were compared since grade is a subjective phenotype and some grade 2 tumours have gene expression signatures similar to those of grade 1 or grade 3 tumours[155] and may be reclassified as eith grade 1 or 3. Associations were tested for using the Cochran-Armitage trend test and two-by-three tables of disease by genotype (tests performed using PLINK[156]).

2.8.2.3 Significance

To correctly identify accurate levels of significance in association studies it is important to correct results for the number of tests performed. HapMap suggest that the number of independent tests, rather than the total number of tests are corrected for therefore we corrected our data for the 82 independent tests performed, as determined by the multidimensional scaling procedure with maximum pair-wise r^2 values of less than 0.2, giving a threshold of significance of $P< 0.0006$ (0.05/82).

Chapter 3 Methods for Functional Studies

3.1 Overview

A series of experiments were conducted to evaluate the function of SNPs in the promoter region of the MMP-9 gene.

Firstly, our panel of cell lines was screened for expression of MMP-9 by PCR to help decide which cell line to use in functional experiments. Total RNA was isolated using the RNeasy® Mini Kit Spin Method and the RNA was reverse transcribed using first strand cDNA synthesis by Superscript II reverse transcriptase (Invitrogen) and oligo $(dT)_{15}$ primers (Promega). The cDNA was then amplified for the detection of the MMP-9 gene and the products viewed on an agarose gel. Expression of β-actin, a house-keeping gene found in all cells, was used as an internal standard for RNA integrity and accuracy of loading. The primers used for detection of MMP-9 were as previously described by Menshikov et al[397]. To confirm the result of the initial PCR, an additional PCR was carried out with a different primer set as described by Ricca et al[398].

The luciferase reporter vector containing 2192bp (-2181 to +11) of the MMP-9 promoter was a kind gift from Professor Shu Ye. The DNA construct contained the T allele at the −1562 polymorphic site (rs3918242). The sequence was verified by sequencing and used as a template to generate three different reporter constructs (the T or C allele at -1562 with either the A or T allele at -1803 (rs3918241) by site directed mutagenesis (SDM; QuikChange Site Directed Mutagenesis Kit, Stratagene) and the results of these were also verified by sequencing. Five of each construct were made by SDM and DNA (mini-prep) from two of each was sequenced to check the allele sequence. One of each construct showing the correct haplotype was selected for use in the reporter assays and DNA (maxi-prep) sequenced with additional primers to check a wider area of sequence.

Selected cell lines were plated and incubated for 24 hours before being transiently transfected with either 100ng or 200ng/well of firefly luciferase reporter construct using Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions.

Cells in each well were incubated with lipofectamine, MMP-9 promoter/luciferase construct and Opti-MEM at 37˚C for 3-4 hours. The DNA/lipofectamine mixture was then removed and the medium was changed. Cells were incubated for an additional 48 hours and luciferase activity was determined as per the manufacturer's instructions. Cells were co-transfected with 20ng PRLSV40 a *Renilla* luciferase expressing vector to serve as a reference for transfection efficiency to standardise results. Experiments were performed in triplicate and repeated at least three times. In addition, for controls, the same amount of empty PGL3 Basic vector was transfected into cells.

Reporter assays were performed 48 hours after transient transfection of cells. The cells were lysed and luminescence was measured using the Dual-Glo™ Luciferase Assay System (Promega) and an automated chemiluminescence detector. Reporter activity was measured as the ratio of firefly to *Renilla* luciferase activity to give RLU (relative light units) and results normalised to the plate by dividing each result (average of the triplicate repeats) by the average RLU for the plate. Results from independent experiments were averaged. One-way ANOVA and Tukey HSD post hoc testing with SPSSv17 software were used for the statistical analysis of the effect of the MMP-9 promoter SNPs in reporter assays.

3.2 General Materials

3.2.1 Cell Lines

BT474 – oestrogen receptor positive, progesterone receptor positive, Her-2 positive

H1299 – non-small cell lung carcinoma cell line, partial deletion of TP53 gene resulting in no expression of p53 protein accounting in part for their proliferative propensity

MCF-7 – oestrogen receptor positive, progesterone receptor positive, hormone responsive, poorly metastatic, adenocarcinoma

MDA-MB-231 – oestrogen receptor negative, highly invasive, one of the most aggressive breast cancer cell lines, adenocarcinoma

MRC5 – lung cancer fibroblast cell line

NIH-3T3 – mouse embryonic fibroblast

SKBR3 – metastatic breast cancer cell line, oestrogen receptor negative, Her-2 over-expressing

T47D – oestrogen receptor positive, progesterone receptor positive, hormone responsive

ZR75.1– oestrogen receptor positive, progesterone receptor positive, hormone responsive

3.3 General Methods

3.3.1 Preparation of Nucleic Acids

3.3.1.1 DNA Extraction from Whole Blood

DNA was extracted from whole blood samples using the salting out method of extraction and was prepared over the period of time from 2005-2007 by myself and Human Genetics DNA bank staff, an additional batch of blood samples for DNA extraction was outsourced to Tepnel.

Whole blood samples, received in 15ml Vacutainers™ and frozen at -80˚C for long term storage, were defrosted and mixed on a rotary mixer in the 4˚C cold laboratory. Each blood sample was transferred into a labeled 50ml conical tube. A clothes peg was placed on the bottom of the upturned Vacutainer™ and rested in the top of the conical tube allowing for natural drainage of blood from the Vacutainer™ and minimal blood loss. The red cells were then removed by washing three times with erythrocyte lysis buffer (ELB); 30ml ELB was added to each 10ml blood sample (15ml for blood samples with a very low volume (~2.5ml)), the samples were placed on ice on the rotary mixer for at least 15 minutes, spun at 1500rpm for 10 minutes at 4˚C, the supernatant discarded in Virkon™ solution, the pellet loosened by gently flicking the bottom of the tube and the process repeated twice more. After the final wash the supernatant was poured off into Virkon™ and any excess ELB was blotted with a white paper towel. 2ml nuclei lysis buffer (NLB) was added to each tube (for low volume samples: 5ml whole blood use 2ml NLB, 2.5ml whole blood use 1.5ml NLB) and the pellets washed into labeled 15ml conical tubes containing 1ml NLB, ensuring the pellet is resuspended. The white cell pellet was lysed by addition of 250µl of 10% SDS to each sample and the samples shaken to disperse the SDS. 150µl protease were added to each sample and the samples placed in a shaking incubator and left mixing gently at 37˚C overnight to allow for protein digestion. The samples were removed from the incubator and allowed to cool to room temperature. 1ml saturated sodium chloride was added to each sample and the samples shaken vigourously by hand for 15 seconds to dehydrate and precipitate out the digested proteins (for low volume

samples: 5ml whole blood use 750µl, 2.5ml whole blood use 500µl) before being spun in the centrifuge at 4000rpm for 20 minutes at room temperature. The clear supernatant was removed and placed in a fresh, labeled 15ml conical tube. The DNA was then precipitated using cold absolute ethanol, adding twice the volume of cold ethanol to each sample and mixing by gentle inversion until the DNA strands were visible and a pellet formed. The DNA pellet was washed in 70% ethanol by removing from the 15ml conical tube using 200µl tip on a Gilson pipette and placing in a labelled microcentrifuge tube containing 1 ml of 70% ethanol. The samples were spun in a microcentrifuge on the highest speed for 15 seconds and the ethanol removed from the tubes. The DNA was allowed to dry. Finally the DNA pellet was dissolved in 1ml TE buffer overnight at room temperature ready for storage (for low volume samples: 5ml whole blood use 500µl, 2.5ml whole blood use 300µl-500µl TE buffer depending on pellet size). Samples were stored at -20˚C, or -80˚C for very long term storage.

NLB, ELB, 10% SDS, protease, saturated salt and TE buffer were made in advance by Human Genetics staff (9.4).

3.3.1.2 Total RNA Extraction from Cell Culture Preparations; RNeasy® Mini Kit Spin Method (Quiagen)

Cells were grown on a 90mm culture dish until confluent, the medium was aspirated from cells and 10ml ice cold PBS added to wash the cells. The PBS was removed to waste and another 1ml PBS added. Using a rubber plunger the cells were scraped from the top to the bottom of the dish. The cells and PBS were collected in a 1.5ml eppindorf and spun at 3000rpm (900g) for 4 minutes. The PBS was removed to waste and the cell pellet snap frozen and stored at -80˚C until needed.

RNA was extracted from the cell culture preparations using the RNeasy® Mini Kit Spin Method according to the manufacturer's instructions.

The RNA pellet was thawed then loosened by gently flicking the tube. 350µl of Buffer RLT (with 1% β-mercaptoethanol) was added and the sample vortexed to mix. The sample

was then homogenized by vortexing for 1 minute. 1 volume (350µl) of 70% ethanol was added to the homogenized lysate and mixed by pipetting. The sample, including any precipitate that may have formed, was added to an RNeasy® spin column placed in a 2ml collection tube, the lid closed and the sample centrifuged for 15 seconds at 8000g (10,000rpm). The flow through was discarded.

An optional on-column DNase digest was performed with the RNase-free DNase Set. 350µl Buffer RW1 was added to the RNeasy® spin column, the lid closed and the sample centrifuged for 15 seconds at 8000g (10,000 rpm) to wash the spin column membrane and the flow-through discarded. 10µl DNase I stock solution was added to 70µl of Buffer RDD in a microcentrifuge tube, mixed by gently inverting the tube then centrifuging briefly before being added to the RNeasy® spin column membrane and left at room temperature for 15 minutes. 350µl Buffer RW1 was then added to the RNeasy® spin column, the lid closed, the column centrifuged for 15 seconds at 8000g (10,000rpm) and the flow-though discarded.

To wash the spin column membrane, 500µl of Buffer RPE was added to the RNeasy® spin column and centrifuged for 60 seconds at 8000g (10,000rpm) and the flow through discarded. This was repeated, centrifuging for 2 minutes the second time. The RNeasy® spin column was then placed in a 1.5ml collection tube and 30µl RNase-free water added directly to the spin column membrane and left to stand for 1 minute to pre-wet the membrane before centrifuging for 1 minute at 8000g (10,000rpm) to elute the RNA.

3.3.1.3 Measurement of Nucleic Acid Concentrations

To measure the concentration of nucleic acid (DNA and RNA) 1µl of the sample was analysed on a NanoDrop® ND-1000 Spectrophotometer according to the manufacturers instructions.

3.3.2 cDNA Synthesis; First-strand cDNA Synthesis by SUPER SCRIPT II Reverse Transcriptase

1µg target RNA, 1µl 0.5µg/ml oligo dT(15) (Promega) and 1µl 10 mM dNTP mix (Promega) were added to a PCR tube and made up to 12 µl with dH_2O. The sample was placed on a PCR machine and heated to 65˚C for 5 minutes, then put on ice for ~1min to ensure the oligo dT(15) annealed to the poly(A) tail of the RNA. 4µl RT Buffer (5x1st strand Buffer, Invitrogen)), 2µl 0.1M DTT (Invitrogen) and 1µl RNase out (Invitrogen), made as a mastermix, was added and mixed by pipetting. The sample was incubated at 42˚C for 2 minutes and then 1µl SScript II RT (Invitrogen) added and mixed by pipetting. The sample was then incubated at 42˚C for 50 minutes, allowing synthesis of the complementary strand, and heated at 70˚C for 15 minutes to inactivate the enzyme. cDNA was stored at -20˚C short term (weeks), or -80˚C long term.

3.3.3 Polymerase Chain Reaction (RT-PCR)

RNA extracted from cell lines was reverse transcribed as described above (Sections 3.3.1.2 and 3.3.2) and 2µl of the cDNA synthesis products were diluted in a 50µl reaction mixture containing 1x Buffer, 0.2mM dNTPs, 1.5mM $MgCl_2$ (for MMP-9 PCRs, 3mM for β-actin), 0.001U/µl GoTaq DNA polymerase and 0.4µM.

MMP-9 PCR conditions were as follows: 94˚C for 3 minutes, followed by 35 cycles of 94˚C for 30 seconds, 58˚C for 30 seconds then 72˚C for 30 seconds and 72˚C for 2 minutes final extension and 10˚C final hold.

β-Actin PCR conditions were as follows: 94˚C for 3 minutes then 20 cycles of 94˚C for 30 seconds, 64˚C for 30 seconds, 72˚C for 30 seconds and 72˚C for 2 minutes for final extension and 10˚C final hold for β-actin.

The oligonucleotide primers for MMP-9 were as described[397] and synthesised for us by Eurofins MWG Operon (Ebersberg, Germany) and were as follows: MMP-9 F: 5' –GTG CGT CTT CCC CTT CAC TTT CCT- 3' MMP-9 R: 5' –GGA ATG ATC TAA GCC CAG CG- 3' giving a 199bp fragment.

The presence of MMP-9 in the panel of cell lines was confirmed with a second set of MMP-9 PCR primers[398] synthesized for us by Eurofins MWG Operon (Ebersberg, Germany) and were 5' –CAA GGA TAC AGT TTG TTC CTC G- 3' and 5' –CAG AGA ATC GCC AGT ACT TCC- 3' giving a 460bp fragment. The cycle was as follows: 94˚C 3 minutes start, 38 cycles of 94˚C for 1 minute, 57˚C for 1 minute, 72˚C for 1 minute, then 72˚C for 2 minutes at end with final hold 10˚C.

Negative controls (samples lacking cDNA) were included in each assay. The PCR products were subjected to electrophoresis in an agarose gel with SafeView™.

3.3.4 Analysis of Nucleic Acids by Agarose Gel Electrophoresis

PCR products were resolved using agarose gel electrophoresis in TAE (Tris-Acetate-EDTA) Buffer. Different concentration gels were used depending on the size of the fragments to be viewed. The 184 bp β-actin and 199bp MMP-9 fragments were analysed on a 2% gel and 460bp MMP-9 fragments were analysed on a 1.5% gel. All gels were run at 70mA for 45minutes. Plasmids were viewed using a 0.6% gel.

To make the gels, agarose was measured in a conical flask and 75ml 1xTAE Buffer added to the conical flask. The flask was then heated in a microwave oven on full power until all agarose had dissolved. 3.75µl SafeView™, a safe nucleic acid stain for the detection of double-stranded DNA, single-stranded DNA and RNA in agarose gels, was then added to the flask (0.05µl/ml TAE) and the gel poured into the gel tray with the gel comb in position. The gel was allowed to set and the comb removed. 15µl of PCR product was loaded into the wells with 3µl of orange G loading dye. 5µl 100bp marker with 3µl loading dye was also loaded onto the gel for easy molecular weight reference. The gel was subsequently viewed with a UV light.

3.3.5 Cloning procedures

The luciferase reporter plasmid for MMP-9 contains 2192bp (-2181 to +11) of the human MMP-9 promoter cloned into the promoterless plasmid vector (pGL3 Basic) containing the coding sequence for firefly luciferase and was kindly provided by Professor Shu Ye[394]. This

plasmid contained the T allele of the SNP rs3918242 at -1562bp. The DNA construct was verified by sequencing (Geneservice).

Site Directed Mutagenesis; QuikChange™ Site-Directed Mutagenesis Kit (Stratagene)

Additional point mutations (Figure 9) were introduced into the luciferase reporter plasmid using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene).

Two complementary oligonucleotide primers containing the desired mutation (alternate alleles of rs3918241 and rs3918242) were designed and synthesised by Eurofins MWG Operon (Ebersberg, Germany).

Primer design requirements: Both mutagenic primers must contain the desired mutation and anneal to the same sequence on opposite strands of the plasmid, primers should be 25 and 45 bases in length and the melting temperature T_m of the primers should be greater than or equal to 78˚C and the primers optimally should have a minimum GC content of 40% and should terminate in one or more C or G bases.

Sample reaction: 5µl of 10x reaction buffer, 1µl (50ng) of dsDNA template, 1µl (125ng) of each oligonucleotide primer, 1µl (125ng), 1µl dNTP mix (1µl of each A, C, G, T), ddH_2O to a final volume of 50µl and 1µl *PfuTurbo* DNA polymerase (2.5 U/µl). Samples were then heated to 95˚C for 30 seconds, then 18 cycles of 95˚C for 30 seconds, 55˚C for 1 minute, 68˚C for 18 minutes (2min/kb of plasmid length). The samples were transfered to ice and 1µl of *Dpn* I restriction enzyme (10U/µl) added directly to each amplification reaction. Each reaction was mixed gently and thoroughly by pipetting up and down several times and then spun in a microcentrifuge for 1 minute and immediately incubated at 37˚C for 1 hour to digest the parental DNA, then placed on ice.

The SDM primers used were as follows: rs3918241 F: 5’ –TGT AAA GGA AGT TAA TTA TCT CC- 3’; rs3918241 R: 5’ –GGA GAT AAT TAA CTT CCT TTA CA- 3’,rs3918242 F: 5’ –GGC GTG GTG GCG CAC GCC TAT AAT ACC- 3’; rs3918242 R: 5’ –GGT ATT ATA GGC GTG CGC CAC CAC GCC- 3’

3.3.5.1 Preparation of LB Broth

A stock of LB broth was maintained by Cancer Sciences Division Staff.

3.3.5.2 Preparation of Selective Agar Plates

A stock of ampicillin agar plates was maintained by Cancer Sciences Division staff.

3.3.5.3 Transformations

15ml falcon tubes were pre-chilled. 50µl chemically competent XL-1 E.coli cells were thawed on ice. The end was clipped off a pipette tip (cell saver pipette tip). The vial of cells was stirred to encourage thawing. 50µl of the competent cells were added to each falcon using the cell saver pipette tip. 0.5µl DNA was added to each tube and stirred with the pipette tip and the tube flicked gently to mix. The falcon tubes were then left on ice for 20 minutes, heat shocked at 42˚C for exactly 45 seconds and placed back on ice for 2 minutes. 250µl of broth was added to each tube. The tubes were then left in the shaker at 200rpm, 37˚C for 1 hour 30 minutes. The tubes were taken out of the shaker, the contents streaked onto the selective LB (100µg/ml Ampicillin) agar plate and the plates placed in the incubator overnight (37˚C).

3.3.5.4 Selection and Culture of Colonies

A single colony was selected from the freshly streaked selective ampicillin plate and inoculated in 5ml LB broth with 5µl 1000mg/ml ampicillin (final conc 100µg/ml) and incubated overnight at 37˚C with vigorous shaking (220rpm).

The bacterial cells were harvested by centrifugation for 10 minutes at 1500rpm and the supernatant was removed by inverting the tube until all medium had drained out.

If growing up cells for DNA Maxiprep, the contents of the universal was poured into a 500ml conical flask with with 125ml broth with 125µl ampicillin and incubated (overnight) at 37˚C with vigorous shaking (220rpm).

3.3.5.5 DNA Extraction from Transformed Bacterial Cultures; QIAprep Spin Miniprep Kit (Qiagen)

Plasmid DNA was extracted using the QIAprep Spin Miniprep kit according to the manufacturer's instructions.

The pelleted bacterial cells were resuspended in 250µl Buffer P1 and transferred to a microcentrifuge tube, 250µl Buffer P2 was added and mixed thoroughly by inverting the tube 4-6 times, 350µl Buffer N3 was added and mixed thoroughly by inverting the tube 4-6 times and then centrifuged for 10 minutes at 13,000 rpm (17,900g) to form a compact white pellet. The supernatant was decanted into the QIAprep spin column, centrifuged for 30-60 seconds and the flow-through discarded. The spin column was washed by adding 0.75ml Buffer PE and centrifuging for 30-60 seconds, the flow through discarded and the column centrifuged for a further minute to remove any residual wash buffer. Finally the QIAprep spin column was placed in a clean 1.5ml microcentrifuge tube and 50µl Buffer EB (10mM Tris-Cl, pH 8.5) or water added to the centre of the spin column and the column left to stand for one minute, then centrifuged for one minute to elute the DNA.

3.3.5.6 DNA Extraction from Transformed Bacterial Cultures; QIAfilter Plasmid Maxi Kit (QIAgen)

Plasmid DNA was extracted using the QIAfilter Plasmid Maxi Kit according to the manufacturer's instructions.

The culture media was poured into two labelled 50ml falcon tubes and pelleted by centrifugation at 6000g for 15 minutes, the medium was poured off to waste and the first pellet was resuspended completely in 10ml Buffer P1 by vortexing, transferred into the second falcon and vortexed to break up the second pellet. 10ml Buffer P2 was added and mixed thoroughly by vigorously inverting the tube 4-6 times, then incubated at room temperature for 5 minutes. During the incubation the QIAfilter Cartridge was prepared and placed in a convenient tube. 10ml Buffer P3 was added to the lysate and mixed thoroughly by inverting vigorously 4-6 times and the lysate immediately poured into the

barrel of the QIAfilter Cartridge and incubated at room temperature for 10 minutes. During the incubation the QIAGEN-tip 500 was equilibrated by applying 10ml Buffer QBT and the column allowed to empty by gravity flow. The cap was removed from the QIAfilter Cartridge outlet nozzle and the plunger inserted gently into the QIAfilter Maxi Cartridge and the cell lysate filtered into the previously equilibrated QIAGEN-tip. The cleared lysate was allowed to filter by gravity flow and the tip then washed with 2x30ml Buffer QC. The Buffer QC was allowed to pass through the QIAGEN-tip by gravity flow. Once the Buffer QC had passed through the filter the QIAGEN-tip was placed over a fresh labelled falcon tube containing 10.5ml isopropanol. 15ml Buffer QF was added to the QIAGEN-tip to elute the DNA and allowed to pass through the filter by gravity flow. The tubes were centrifuged for 4000rpm at 4˚C for 30 minutes to pellet the DNA. The isopropanol was poured away and 1ml 70% ethanol added to the tube and vortexed. The contents of the tube was transferred to a microcentrifuge tube and centrifuged for 5 minutes at 10500xg. The 70% ethanol was removed and the tube left, with the cap off to dry and then resuspended in 1xTE diluted in 300µl dH_2O overnight, stored in the fridge.

3.3.5.7 Sequencing of Plasmid DNA

All sequencing orders were placed online at sequencing@geneservice.co.uk. 10µl of 3.2pmol/µl primer and 15µl at 100ng/µl of sample were required. The MMP-9 promoter region of the initial plasmid sent was sequenced using the RV3 and GL2 primers and the SDM-created constructs were sequenced with the RV3 primer to check the region where mutations were introduced and with the GL2 and MMP9_Int_F (designed by Dr Blaydes, synthesized by Eurofins MWG Operon (Ebersberg, Germany)) primers to check a wider region to ensure no other base changes had been introduced in the three reporter constructs selected for functional experiments.

3.3.6 Transfection Procedures

All cell culture procedures were performed in a filtered hood in sterile conditions. Good laboratory practice was performed at all times and good asceptic conditions maintained

when handling cell cultures and reagents. All liquid waste (cells) was treated with 10% Trigene prior to disposal.

3.3.6.1 Preparation of Cells

MCF-7, MDA-MB-231, MRC5 and NIH-3T3 were maintained in Dulbecco's modified Eagle's medium, DMEM (Invitrogen), and H1299 cells maintained in Roswell Park Memorial Institute medium, RPMI 1640 (Invitrogen), both mediums were supplemented with 10% foetal calf serum (Autogen Bioclear) and penicillin/streptomycin solution (1%) (Cambrex) and L-Glutamine solution (1%) (Cambrex). Cells were incubated at 37˚C and 10% CO_2 for those in DMEM or 5% CO_2 for those in RPMI. All cell lines had been tested for Mycoplasma contaminations and found to be negative. Generally cell cultures were re-fed every other day with culture medium until confluent.

Cells in culture flasks were maintained by periodic 'splitting'. When confluent cells were trypsinised according to the Standard Operating Procedure, JPB2_cell culture, and about 80% of the cells were removed and an equal volume of fresh medium is added, depending on experimental requirements.

Retrieval of frozen cells: Stocks of cell cultures were stored in vapour phase liquid nitrogen tanks. To retrieve cells, they were transferred to a 37˚C water bath or the CO_2 incubator then rapidly thawed in 10mls of complete medium (DMEM or RPMI) in a 15ml falcon tube. Cells were spun at 1000rpm for 3 minutes, the supernatant removed to waste and the pellet resuspended in 10mls fresh medium (5mls if small pellet) and the cell suspension transferred to a T-75 (or T-25 if vial was sparse) tissue culture flask and incubated at the required CO_2 concentration.

Contact-dependent co-cultures were established by plating $1.5x10^6$ MRC-5 and $1.5x10^6$ MCF-7 cells in a 90mm culture dish and made up to 10mls with DMEM and incubated at 37˚C and 10% CO_2 for 48 hours.

3.3.6.2 Transfection with Plasmid DNA

Cells were plated 24 hours in advance of transfection experiments. The medium was aspirated and the cells washed with 10ml HBSS (T75 culture flask) and the HBSS aspirated. One ml of Trypsin-EDTA in HBSS was added to the culture flask and the cells agitated to encourage detachment from the flask. 8ml medium was pipetted over the cells and pipetted up and down to break up any clumps that may have formed. The cell suspension was transferred to a falcon tube, leaving 1-2mls depending on confluency of the cells in the flask to maintain a stock of cells (made up to 10mls with fresh medium and replaced in incubator). The falcon tube was inverted to mix and 40µl of the cell suspension was pipetted into a clean haemocytometer and the cell number counted (average of three grids), giving the number of cells $x10^4$/ml. $2x10^4$ cells/well were transferred into a 96-well cell culture plate and the volume made up to ~150µl with fresh medium. The cell culture plates were incubated at 37˚C for 24 hours.

Prior to transfections the plated cells were examined under a light microscope to ensure cells were ~80% confluent. Fresh dilutions of plasmids were made as required before each transfection experiment (either 100ng/µl or 200ng/µl for plasmid DNA and 20ng/µl for the *Renilla* luciferase expressing plasmid PRLSV40) and centrifuged at 4˚C, 10.5rpm, for 10 minutes. Separate mastermixes were made for each plasmid being transfected of 200ng reporter vector and 20ng of PRLSV40 in 25µl Opti-MEM per reaction. A second mastermix of 25µl Opti-MEM plus 0.6µl Lipofectamine 2000 reagent per reaction was made and incubated at room temperature for 5 minutes in the filtered hood. 25µl per reaction of the lipofectamine/Opti-MEM mix was then added to each of the reporter/Opti-MEM mixes and incubated at room temperature in the filtered hood for 20 minutes. The culture medium on the plated cells was carefully taken down to a volume of 100µl. 50µl of transfection mix was added per well and mixed. The cells were then incubated for 3-4 hours in CO_2, the transfection mix was removed and the cells re-fed with 100µl fresh medium and then incubated for ~48hours. Transfections were performed in triplicate for each construct and repeated at least three times to ensure reproducibility of

the results. Total amounts of transfected plasmid were equalised by addition of the empty vector pGL3-noluc (pGL3 Basic vector with the luciferase gene removed).

3.3.7 Reporter Gene Assay

3.3.7.1 Luciferase Reporter Assay; Dual-Glo™ Luciferase Assay System (Promega)

Cells were removed from the incubator. Medium was taken down to 50µl/well. 50µl Dual Glo Luciferase reagent was added per well, pipetted up and down to mix and incubated at room temperature for 10 minutes. The lysed cells were transferred to an optiplate, covered with a sticky film and processed on a chemiluminescence detector. 50µl 1x Stop and Glo reagent was added to each well and processed on the plate reader. The reporter activity was measured as the ratio of firefly to Renilla luciferase to normalise results. Results from independent transfection experiments were normalised, averaged and standard error calculated. One-way ANOVA and Tukey HSD post hoc testing with SPSSv17 software were used for the statistical analysis.

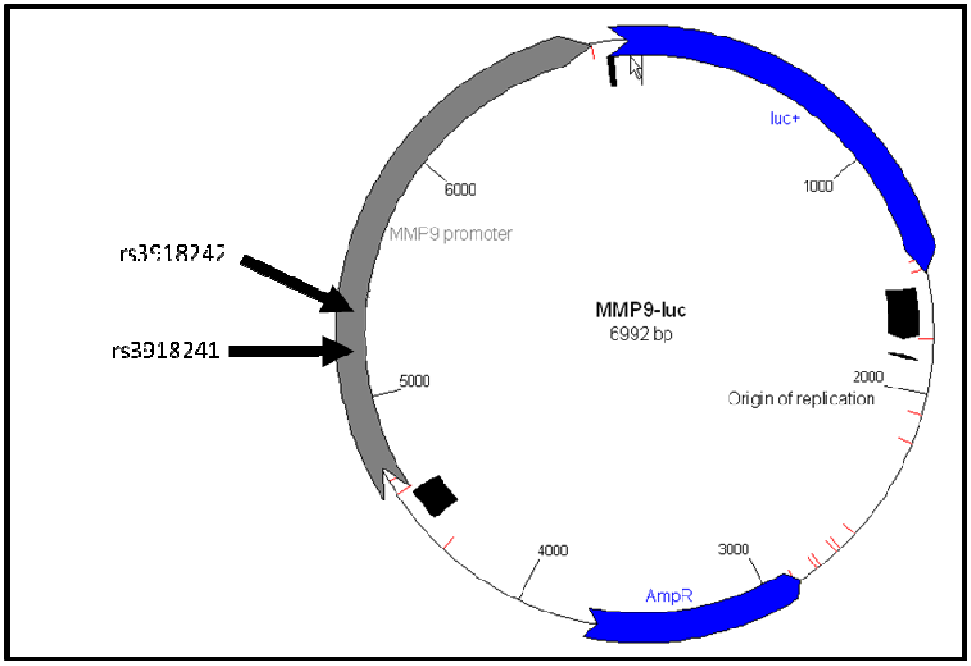


Figure 9: Diagram of pGL3-MMP-9 promoter construct. Plasmids were created for the four possible haplotype combinations of the 2 SNPs (rs3918241 T/A and rs3918242 C/T) using site directed mutagenesis of the original plasmid.

Chapter 4 Results

4.1 Analysis Results

4.1.1 Overview

SNPs from 30 candidate genes were selected and genotyped. The genotype data was screened to remove poor quality data and the remaining SNPs associated with grade, lymph node status, oestrogen receptor status, presence of distant metastases and lymphovascular invasion using the Cochran-Armitage trend test and 2-by-3 tables of disease by genotype. The effect of these SNPs on risk was investigated using the Log rank test to compare Kaplan-Meier survival curves.

4.1.2 Analysis QC results

After performing quality control checks on the data, 73 SNPS and 58 individuals were removed from the analysis due to poor genotyping. After population stratification a further 44 individuals were removed leaving 899 for the analysis.

4.1.3 Patient Cohort

The phenotypic characteristics of the breast cancer cases that were genotyped and passed quality control are shown in Table 3, grouped according to their oestrogen receptor status and grade. Of the 899 cases successfully genotyped, 26 had missing phenotype data. Our cohort of young onset breast cancer patients contains more high grade, oestrogen receptor negative tumours than would be expected in breast cancer cohorts including individuals aged 40 years and over at diagnosis[399], having nearly 10% more ER-negative tumours compared with women aged 40 years and older at diagnosis[399] (30.9% in POSH versus 21.4% in women aged ≥ 40 years).

Table 3: Summary statistics– phenotypic characteristics of the 899 individuals successfully genotyped according to oestrogen receptor status and grade (% of total). 26 individuals were missing phenotype data

	ER+	*ER-*	*Unknown*	*Total*
Grade 1	60	3	1	**64**
Grade 2	276	28	2	**306**
Grade 3	267	239	7	**513**
Unknown	12	3	1	**16**
Total	**615**	**273**	**11**	**899**

4.1.4 Association with Phenotype

The SNPs that passed QC were investigated to determine any association with different aspects of tumour biology using Cochran-Armitage trend test and 2-by-3 tables of disease by genotype, using Plink[156]. The frequency of both alleles in the patient cohort was recorded as well as the chi-square, asymptotic p-value and an estimated odds ratio for the minor allele for each SNP. An odds ratio above 1 indicates a positive association with the phenotype and a value below one a negative association. A corrected P value of 0.0006 or less indicated a significant result.

4.1.4.1 Association with Grade

Comparison of phenotypic extremes of grade (grade 1 vs. grade 3) were compared to ensure significant results and gave evidence that two SNPs in ESR1, and a SNP in IGF1 and LSP1 were significantly associated with grade, with ESR1 and IGF1 SNPs being associated with a higher grade, and the LSP1 SNP with lower grade (Table 4). None of these associations remained significant after Bonferroni correction for the 82 independent tests.

Table 4- Association with grade

GENE	*SNP*	*A1*	*Frequency*		*A2*	*CHISQ*	*P*	*OR*
			Grade 3	**Grade 1**				
ESR1	rs2228480	A	0.2178	0.1032	G	9.031	0.00265	2.421
ESR1	rs3798577	C	0.5088	0.373	T	8.275	0.00402	1.741
IGF-1	rs2373721	G	0.25	0.1406	C	7.509	0.00614	2.037
LSP1	rs661348	C	0.3694	0.4921	T	7.146	0.00752	0.6046
ATM	rs4987876	T	0.0765	0.1406	G	6.108	0.01346	0.506
DAPK1	rs11141901	G	0.2173	0.3125	A	5.855	0.01553	0.611

4.1.4.2 Association with Lymph Node Involvement

Several SNPs in genes including CASP8, ERBB2, CYP1B1, ATM, MMP-9 and IGF-1 appear to be associated with lymph node status, however none of the differences are significant (Table 5).

Table 5-Association with lymph node involvement

GENE	*SNP*	*Minor Allele*	*Frequency*		*Major Allele*	*CHISQ*	*P*	*OR*
			LN+	LN-				
CASP8	rs6760993	G	0.5196	0.4622	A	5.712	0.01685	1.259
ERBB2	rs1058808	C	0.3054	0.3578	G	5.371	0.02047	0.7891
CYP1B1	rs10175368	T	0.2978	0.2481	C	5.256	0.02188	1.285
ATM	rs1800889	T	0.04039	0.06463	C	5.172	0.02295	0.6092
MMP-9	rs2250889	G	0.05991	0.03667	C	5.016	0.02511	1.674
IGF-1	rs5742714	C	0.06522	0.09413	G	4.986	0.02556	0.6714

4.1.4.3 Association with Oestrogen Receptor Status

Comparisons of genotypes of patients with oestrogen receptor positive and negative breast cancer produced strong evidence that SNPs in FGFR2 and TOX3 are associated with oestrogen receptor positive breast cancer and two SNPs in the ATM gene were associated with oestrogen receptor negative breast cancer (Table 6). The two FGFR2 SNPs and one ATM SNP remain significantly associated after conservative correction for multiple tests.

Table 6-Association with Oestrogen receptor status

GENE	*SNP*	*Minor Allele*	*Frequency*		*Major Allele*	*CHISQ*	*P*	*OR*
			ER+	ER-				
FGFR2	rs2981582	T	0.478	0.3595	C	21.59	0.00000338	1.632
FGFR2	rs1219648	G	0.4853	0.3741	A	18.91	0.00001372	1.578
ATM	rs1801516	A	0.1092	0.1769	G	14.6	0.0001331	0.5704
TOX3	rs1420546	C	0.3649	0.2854	T	10.41	0.001256	1.438
ATM	rs3092991	G	0.1254	0.1807	A	9.444	0.002119	0.6503
DAPK1	rs9410612	T	0.3068	0.3668	A	6.215	0.01267	0.7641

4.1.4.4 Association with Distant Metastasis

Three SNPs in the MMP-9 gene were significantly associated with the presence of distant metastases (Table 7), however only one of the associations remained significant after Bonferroni correction for the 82 independent tests. When the data was stratified into ER-

positive and ER-negative groups these associations increased in significance in the ER-positive cohort, with three SNPs in the MMP-9 gene (rs6065912, rs3918241 and rs3918261) maintaining significance after correction for multiple tests (Table 9), but were of no significance in the ER-negative group (Table 8). Two SNPs in ATM and one in IGF-1 were significantly associated with the presence of distant metastasis in ER-negative patients before Bonferroni correction for the 82 independent tests.

Table 7-Association with distant metastasis

GENE	*SNP*	*Minor Allele*	*Frequency*		*Major Allele*	*CHISQ*	*P*	*OR*
			Mets	**No Mets**				
MMP-9	rs6065912	A	0.325	0.1323	G	12.31	0.00045	3.159
MMP-9	rs3918241	A	0.325	0.1464	T	9.747	0.001796	2.806
MMP-9	rs3918261	G	0.325	0.1471	A	9.653	0.001891	2.793
IGF-1	rs2373721	G	0.425	0.2312	C	8.153	0.0043	2.458
DAPK1	rs1041326	T	0.275	0.1295	C	7.181	0.007368	2.55
DAPK1	rs11141901	G	0.375	0.2226	A	5.19	0.02272	2.095

Table 8– Association with Distant metastasis in ER-negative patients

GENE	*SNP*	*Minor Allele*	*Frequency*		*Major Allele*	*CHISQ*	*P*	*OR*
			Mets	**No Mets**				
ATM	rs1801516	A	0.5	0.1687	G	8.859	0.002916	4.929
ATM	rs3092991	G	0.5	0.1729	A	8.5	0.003552	4.783
IGF-1	rs2373721	G	0.5833	0.2283	C	8.199	0.004191	4.732
IGF-1	rs1520220	G	0.4167	0.1552	C	5.937	0.01483	3.889
IGF-1	rs2946834	A	0.6	0.3	G	4.161	0.04137	3.5
RHOC	rs2999156	G	0.75	0.4604	C	3.955	0.04674	3.516
RHOC	rs12144044	A	0.5	0.2528	C	3.742	0.05306	2.955

Table 9- Association with Distant metastasis in ER-positive patients

GENE	*SNP*	*Minor Allele*	*Frequency*		*Major Allele*	*CHISQ*	*P*	*OR*
			Mets	**No Mets**				
MMP-9	rs6065912	A	0.3929	0.1307	G	15.93	0.00006582	4.304
MMP-9	rs3918261	G	0.3929	0.148	A	12.62	0.00038200	3.725
MMP-9	rs3918241	A	0.3929	0.1494	T	12.37	0.00043570	3.684
DAPK1	rs1041326	T	0.3214	0.1326	C	8.243	0.00409100	3.099
FN1	rs10207245	A	0.2143	0.4311	T	5.258	0.02184000	0.3599
MMP-7	rs11225309	A	0.3929	0.2125	C	5.244	0.02203000	2.399

4.1.4.5 Association with Lymphovascular Invasion

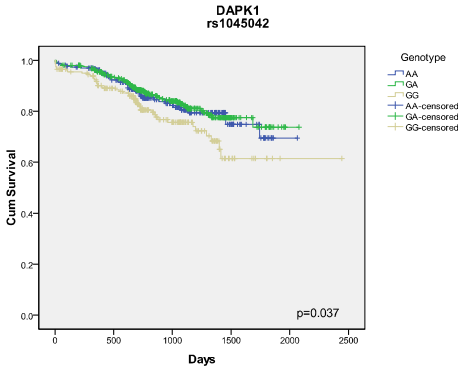
Two SNPs in DAPK1 appear to be associated with lymphovascular invasion, whereas a SNP in the MMP-7 gene has a protective effect (Table 10). None of these associations remained significant after Bonferroni correction for the 82 independent tests.

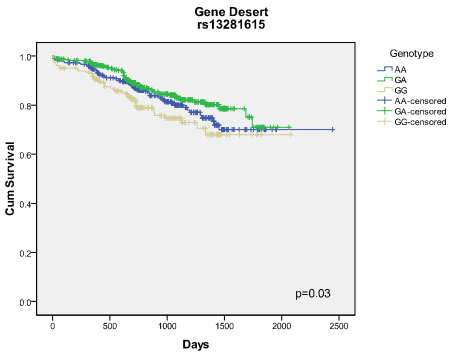
Table 10-Association with lymphovascular invasion

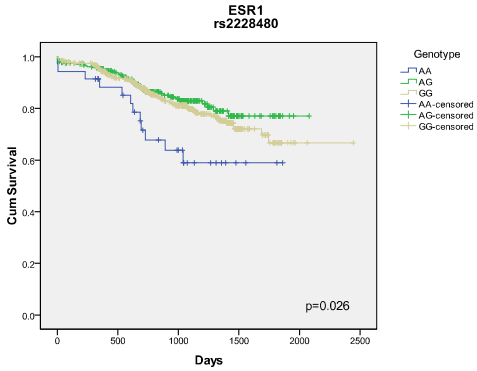
GENE	*SNP*	*Minor Allele*	*Frequency*		*Major Allele*	*CHISQ*	*P*	*OR*
			LVI	**No LVI**				
DAPK1	rs10746815	G	0.4016	0.3349	A	7.8	0.005226	1.333
DAPK1	rs1041326	T	0.1572	0.1118	C	7.072	0.00783	1.482
MMP-7	rs1943779	C	0.252	0.2966	T	4.07	0.04364	0.7989
DAPK1	rs11141901	G	0.248	0.2081	A	3.715	0.05393	1.255
MAP3K1	rs889312	C	0.2882	0.3311	A	3.506	0.06114	0.8179
ESR1	rs3020403	G	0.2987	0.3402	C	3.192	0.07398	0.8258
DAPK1	rs9410612	T	0.3465	0.3054	A	3.143	0.07626	1.206

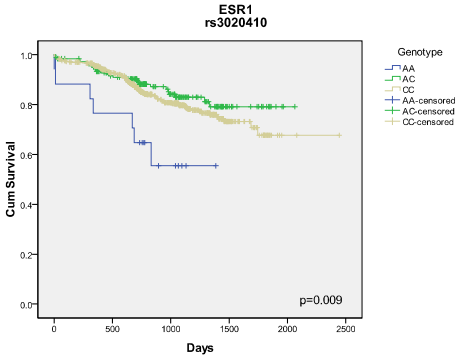
4.1.5 Association with Survival

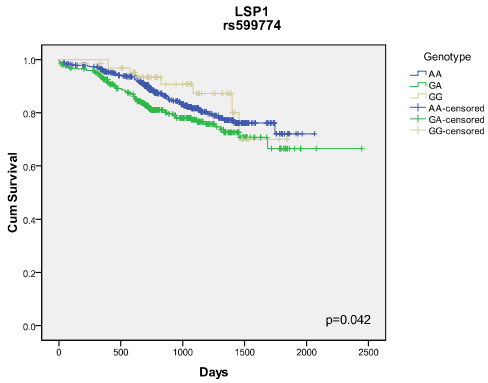
The association of the SNPs passing QC with survival was investigated using distant disease free survival as a surrogate for overall survival. Kaplan-Meier survival curves for the three genotype combinations for each SNP were compared using the Log rank test. Before correction for multiple testing, significant associations were seen with seven SNPs representing five genes, one SNP in DAPK1 (rs1045042), the previously reported SNP in a gene desert (rs13281615), two SNPs in ESR1 (rs2228480 and rs3020410), one SNP in LSP1 (rs599774), one SNP in MAP3K1 (rs889312) and one SNP in MMP-7 (rs1943779). Association of the MMP-9 SNP rs3918241 with distant disease free survival was more significant in ER positive patients than ER negative patient but did not quite reach significance.

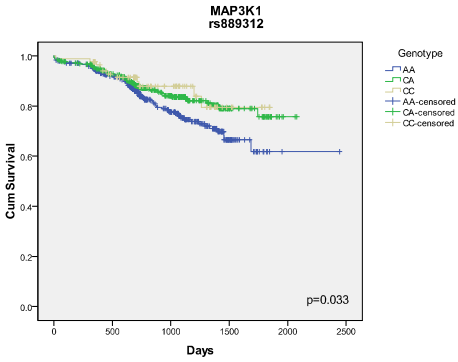












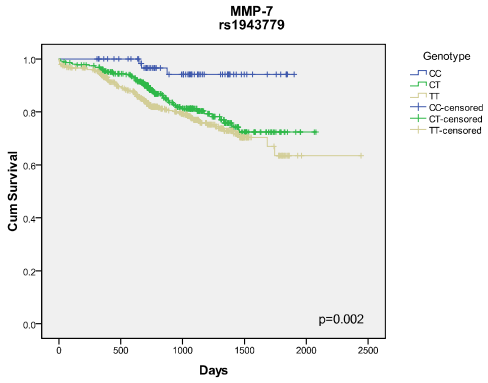


Figure 10: SNPs affecting prognosis (p<0.05) identified by Kaplan-Meier analysis. Survival curves for rare homozygotes are shown as yellow lines, heterozygotes by green lines and common homozygotes by blue lines. *P* values shown are derived from the log-rank test.

4.2 Functional Study Results

4.2.1 Overview

The functional significance of the MMP-9 promoter SNPs rs3918241 and rs3918242 were investigated with a series of transient transfection experiments into a range of cell lines. The SNP rs3918242 has previously been shown to have functional significance[394] and rs3918241 is close by in the promoter region and demonstrated a significant association with distant metastasis in this study. Our panel of cell lines was screened to see which expressed MMP-9 mRNA and those cell lines selected for transfection experiments with a luciferase reporter vector containing one of the four haplotypes of the two promoter SNPs.

4.2.2 Cell Line RT-PCR

RT-PCR showed MMP-9 expression in the H1299, MCF-7, SKBR3 and ZR75.1 cell lines, but not in the T47D, BT474, MDA-MB-231 or MRC-5 cell lines (Figure 11) and this was confirmed with the second primer set (Figure 12). No increase in expression was seen on co-culture of MCF-7 cells with MRC-5 fibroblasts.

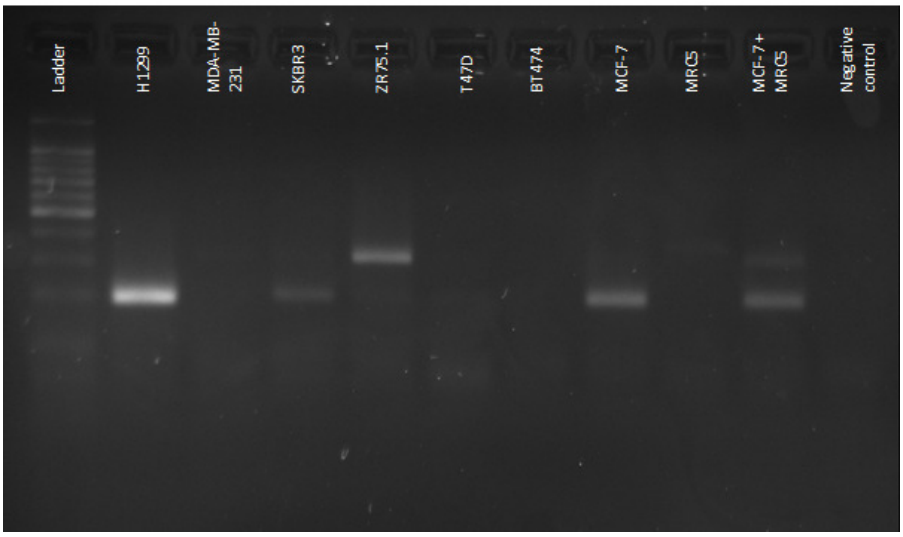


Figure 11: MMP-9 expression identified by RT-PCR - primer set 1

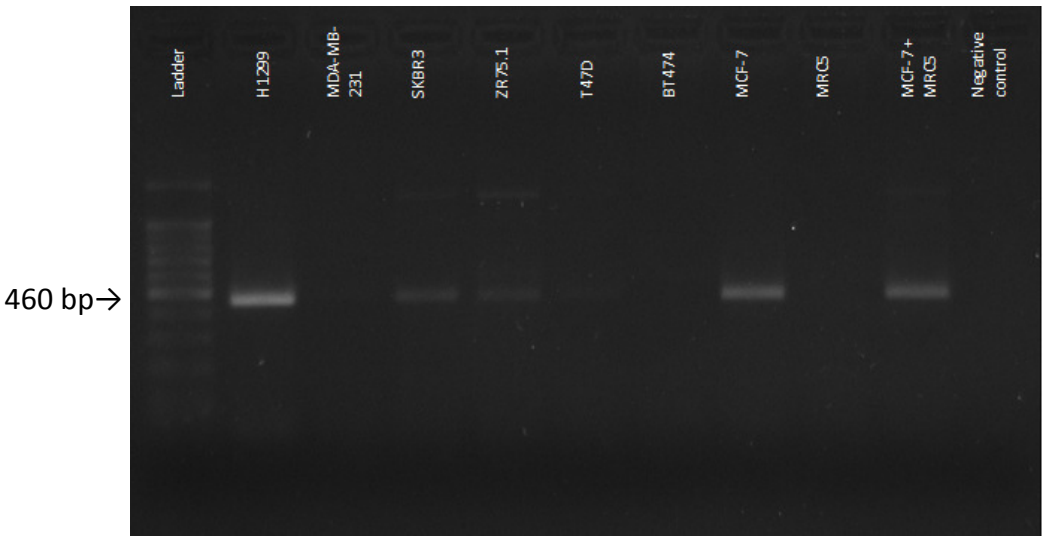


Figure 12: MMP-9 expression identified by RT-PCR - primer set 2

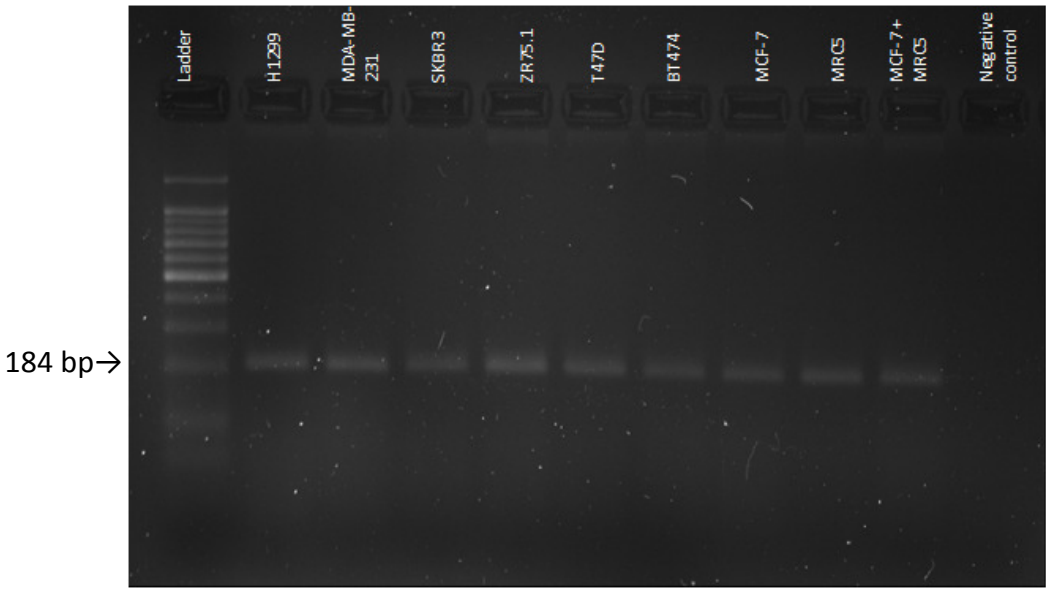


Figure 13: B-actin expression identified by RT-PCR, used as an internal standard for RNA integrity and accuracy of loading.

4.2.3 Luciferase Reporter Assays

We transfected the luciferase reporter vector into both H1299 and MCF-7 cells as both of these cell lines showed expression of MMP-9 by RT-PCR. We also transfected the luciferase reporter vector into the mouse fibroblast cell line NIH-3T3 and the human lung fibroblast cell line MRC5, as well as the metastatic breast cancer cell line MDA-MB-231. RT-PCR did not show a positive expression of MMP-9 in the MDA-MB-231 cell line,

however literature suggested this cell line expressed MMP-9[400;401] and there is a great deal of literature reporting the significance of fibroblast MMP-9 expression in breast cancer[84].

Cells were co-transfected with Renilla Luciferase to control for transfection efficiency and the ratio of firefly to Renilla luciferase calculated to give RLU. The triplicate repeats for each plasmid were averaged and results normalised to the plate by dividing the average result for each plasmid by the average RLU for all transfections on that plate.

4.2.3.1 H1299 Transfections

There were no significant differences between any of the plasmids when transiently transfected into H1299 cells (p=0.061). The minor allele (A) of the SNP rs3918241 gives slightly higher levels of activity than the wild type (T) in this position, as does the wild type (C) allele of rs3918242.

4.2.3.2 MCF-7 Transfections

There were no significant differences between any of the plasmids when transiently transfected into the oestrogen receptor positive MCF-7 cell line (p=0.249). Results show a slightly increased activity with the minor allele (A) at rs3918241 especially in the context of the wild type C allele at rs3918242.

4.2.3.3 MDA-MB-231s Transfections

Statistically significant differences in activity were observed on transfection of the oestrogen receptor negative, highly metastatic MDA-MB-231 cell line with our plasmids. The C allele of the SNP rs3918242 yields approximately a 1.5 fold increase in activity over the T allele and is marginally increased with the presence of the T allele at rs3918241. With the T allele present at rs3918242, the SNP rs3918241 appears to have no effect.

4.2.3.4 NIH-3T3 Transfections

There were no significant differences between any of the plasmids when transiently transfected into NIH-3T3 cells (p=0.256), however a similar pattern was observed as with

the MCF-7 and H1299 cell lines, with the AC haplotype giving the highest levels of activity in reporter assays.

4.2.3.5 MRC-5 Transfections

The MRC-5 cell line did not transfect well enough with our plasmids to give reliable results.

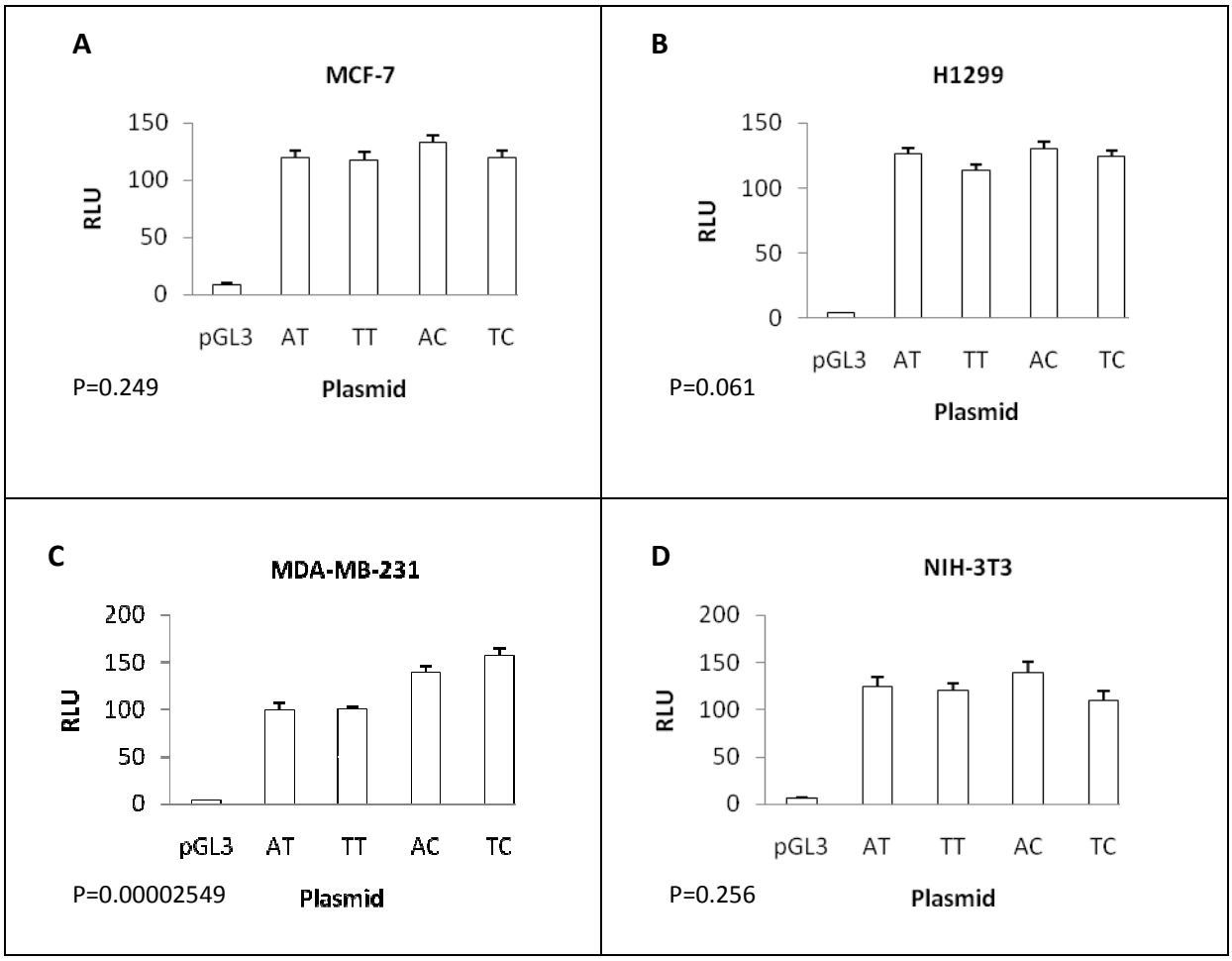


Figure 14: Effects of MMP-9 promoter polymorphisms on MMP-9 expression in MCF-7 (A) H1299 (B) MDA-MB-231 (C) and NIH3T3 (D) cell lines. Plasmids are referred to by the allele present at rs3918241 and rs3918242 respectively. Cells were transfected with 200ng/well of luciferase reporter construct and 20ng PRLSV40 to serve as a reference for transfection efficiency. Reporter activity was measured as the ratio of firefly to *Renilla* luciferase activity to give RLU (relative light units). Results of each transfection were normalised to the plate by dividing each result (an average of the triplicate repeats) by the average RLU for the plate. Results from indepedantly repeated experiments were averaged ((A) and (B) n=9 (C) and (D) n=4). Results shown +/- standard error of the mean. P-values were generated using one-way ANOVA.

Chapter 5 Discussion

5.1 Discussion

Breast cancer affects around 46 000 people in the UK each year and is a leading cause of death amongst women in the western world. There is a clear genetic component to breast cancer with approximately 30% of breast cancers being familial. Despite this, only a fraction of the genes involved in breast cancer susceptibility have been identified. With mutations in the known high (BRCA1, BRCA2, TP53) and intermediate (ATM, BRIP, CHEK2, NSB1, PALB2, PTEN, RAD50) penetrance susceptibility genes, accounting for less than 50% of familial breast cancer cases, a lot remains to be elucidated. Recent GWAS have begun to identify several low penetrance susceptibility alleles that may account for some of unknown genetic susceptibility. Our study investigated whether these common genetic variants might predispose to a certain type of cancer by selecting SNPs in candidate genes and using statistical tests to look for associations with certain tumour phenotypes. We discovered SNPs in various genes that showed associations with different tumour phenotypes both confirming previous associations and identifying novel associations, including three SNPs in the MMP-9 gene associated with distant metastasis. Distant metastasis is the most life-threatening aspect of cancer and remains the overwhelming cause of death in breast cancer patients[157] and MMP-9 has been clearly implicated in progression of breast cancer. We therefore persued a more detailed investigation into the functional significance of the MMP-9 promoter SNPs rs3918241 and rs3918242 as the ability to more accurately predict the risk of distant metastasis will have significant therapeutic implications.

5.1.1 Identifying Common Genetic Variants

Common genetic variants depend on very large scale studies to ensure statistical significance of a suspected SNP association. Genome wide association studies have the power to identify associations between a disease and common genetic variants with low penetrance. However sample size and replication of results are concerns and the

detection of further susceptibility loci will require genome-wide studies with more complete coverage and using ever larger numbers of cases and controls and there is doubt as to whether such large studies are possible, or worthwhile[402] as they may still not produce associations with sufficient predictive power to be of practical use. There is increasing concern that in modern research, false findings may account for the majority or even the vast majority of published research claims[403] and would be worthless for risk prediction[50]. Many GWAS have been underpowered to detect real associations. It has been suggested that more than 2000-5000 samples are needed to identify significant association in these studies[404].

A large number of new susceptibility alleles have been identified and each confers only a very small increase in breast cancer risk. The high frequency of these variants in the population and their low predictive value currently makes them of little use in clinical practice, however either alone or in combination with each other, and with the identification of further alleles, a combination of these may give a much increased odds ratio and thus become appropriate for predictive genetic testing[21]. It is thought it will be possible to identify those individuals that are susceptible to breast cancer by their genotype profile and prevent disease by targeting certain interventions to the at risk individuals. This will provide improvements in the efficacy of population-based programs of interventions for cancers such as mammography by targeting women who are at the greatest risk for breast cancer according to genotype.

Although each SNP may only contribute a small increase in risk, the overall effect of combinations of SNPs on breast cancer risk may be substantial. This polygenic effect of SNPs may be additive or synergistic and may account for most genetic risk. Recently, attempts have been made to identify breast cancer risk that is conferred by SNP-SNP interactions,[44] the strategy developed identified interactions suggesting cross-talk between genes from different pathways such as DNA repair and metabolism pathways and has the potential to identify more complex interactions that may include three or more SNPs in cross-talk between different cancer pathways providing insights into the

multi-genic nature of cancer and the effect of cell functions on breast cancer development. This knowledge may ultimately improve breast cancer risk management.

Our study identified several associations of SNPs with different tumour phenotypes, including three (rs2981582, rs1219648, and rs1801516) in two different genes that remained significant after stringent correction for multiple testing. We confirmed previous associations of SNPs in the FGFR2 gene (rs2981582, rs1219648) with ER-positive breast cancer[211]. Also three SNPs were identified in MMP-9 that were strongly associated with the presence of distant metastasis, one of which was significant after correction for multiple testing in ER-positive patients.

The potential role of SNPs in guiding treatment decisions is promising, however little is known about the roles that these susceptibility alleles play. It is likely that the smaller the odds ratio for a given association, the more likely it is that environmental factors will predominate. Therefore improved understanding of genetic risk factors and their interactions with the environment may allow more accurate predictions of disease and facilitate prevention through measures directed towards people at higher than average risk.

5.1.2 Functional Significance of Common Genetic Variants

We identified SNPs in the promoter region of MMP-9 that were significantly associated with the presence of distant metastasis and given the extensive literature documenting an association of MMP-9 with tumour progression we decided to investigate this further. The proposed role of MMP-9 in breast cancer has been based mainly on observations of high levels of expression in malignant tumours and association of elevated MMP-9 levels with poor prognostic features such as higher grade. We performed a series of transient transfection experiments to see if a SNP was responsible for some of the increase in MMP-9 observed in cancer. We investigated the effect of two promoter SNPs, rs3918241 and rs3918242. Although rs3918242 did not type in our cohort, it has been previously reported to have a fuctional effect on MMP-9 expression[394] and rs3918241 is close by in

the promoter region, and was significantly associated with distant metastasis before Bonferroni correction in the whole cohort and significantly associated with distant metastasis after Bonferroni correction in the ER-positive cohort in this study. Our results indicated that the common, wild type, allele (C) of SNP rs3918242 leads to an increased production of MMP-9. This contradicts the only study to date on the functional effect of this SNP which reports a 1.5 fold increase in MMP-9 expression with the minor (T) allele of this SNP. Although not significant, our data also suggests that the minor (A) allele of rs3918241 is associated with an increased production of MMP-9 in MCF-7, H1299 and NIH-3T3 cells.

5.1.2.1 The Matrix Metalloproteinases

The MMPs are zinc dependent proteases whose primary function is extracellular matrix degradation and remodelling. These actions can impact on the behaviour of tumour cells by creating a space into which tumour cells can migrate, promoting metastasis. MMPs can also cleave growth factors, cell surface receptors, cell adhesion molecules and chemokines/cytokines,[405-408] modifying the activity of these signalling molecules.[409] Matrix metalloproteinases may also regulate angiogenesis both positively and negatively.[157;410;411]

Despite several MMP family members being specifically implicated in the progression of breast cancer, the regulation of MMP expression and activity in tumours appears complex and the manner in which MMPs collectively contribute to breast tumour progression is not well understood[56].

MMPs are mostly secreted into the matrix as latent zymogens (proenzymes) by various stromal and epithelial cell types, including mesenchymal cells, T cells, monocytes, macrophages, neutrophils, keratinocytes and tumour cells. The zymogen is cleaved into an active enzyme, usually in the pericellular or extracellular space. Activation of MMPs in the local micro-environment can result in discrete alterations in tissue architecture[412].

Endogenous MMP inhibitors are the tissue inhibitors of metalloproteinases (TIMPs). TIMPs are produced by the same cells that produce MMPs and act by forming complexes with the MMPs. Under normal physiological conditions MMPs are present in tissues at low levels, usually in the latent form, and are responsible for normal tissue turnover[413] with their local activity in tissues being regulated by the TIMPs.

The function of the MMPs is tightly regulated by several different mechanisms[407;409;414]: at the level of transcription, at the point of activation from the precursor zymogens, interaction with specific ECM components, modulation of MMP mRNA half-life by growth factors and cytokines has been documented[415] and inhibition by the TIMPs[56;415;416]. If MMP-9 production is excessive, TIMP inhibition can be insufficient to control the MMPs which can cause an imbalance in the ECM breakdown and repair system[413].

Several MMPs have been implicated in the development and progression of cancer but the majority of the literature focuses on the gelatinases (MMP-2 and MMP-9) which have been implicated in several aspects of tumourigenesis, particularly invasion and metastasis.[417]

MMP-9 is regulated primarily at the transcriptional level. Its expression is regulated tightly and specifically[418;419] under the control of a 2.2kb upstream regulatory sequence that is known to harbour binding sites for AP-1 (activator protein 1), NFκB (nuclear factor κB), SP-1 (specificity protein 1) and PEA3/Ets (PEA3-polyoma enhancer activator 3).[420] Data by Yao *et al*[417] suggests there is also a heregulin-β1-response element in the MMP-9 promoter region from -84bp to -241bp. Multiple signalling pathways can induce MMP-9 gene transcription by activating sequence specific transcription factors which bind to these cis-elements on the MMP-9 promoter[418;419;421] in turn promoting the further recruitment of chromatin remodelling complexes, co-activators and transcriptional machinery to induce MMP-9 expression.[421] Transcription of the gene generates a 2.5kb transcript that is translated into the latent 92kDa protein product, which is activated by several enzymes including stromelysin, MMP-2 and cathepsin G by the removal of 73 amino acids at the N-terminus.[420]

5.1.2.2 The Role of MMP-9 in Breast Cancer

For distant metastasis to occur tumour cells must detach from the primary tumour site, intravasate into the blood or lymphatic system and extravasate at distant sites of the body. This invasive process is regarded as a multi-step phenomenon. In 1986 Lance Liotta[422] proposed the three step theory of invasion hypothesis that describes the series of biochemical events occurring during invasion of the ECM by tumour cells. The first step is tumour cell attachment via cell surface receptors which specifically bind to components of the matrix such as laminin and fibronectin. The second step is local proteolytic degradation of basement membrane and the extracellular matix (including degradation of the attachment components) by hydrolytic enzymes secreted by, induced by, or activated by the anchored tumour cell. The third step is tumour cell locomotion into the region of the matrix modified by proteolysis. Continued invasion of the matrix may take place by cyclic repetition of these three steps [422;423]. Malignant cells continually stimulate the host stromal and vascular cells to carry out physiological invasion and this activation of the local invasive environment seems to create a permissive field for the malignant cell[76;88;138].

Therefore the localised degradation of the ECM by enzymes, step two of the three step theory of invasion, is crucial for tumour cells to be able to detach from the primary tumour, invade the stromal tissue, enter the circulation, arrest at the peripheral vascular bed, extravasate, invade the target organ interstitium and parenchyma, and form a metastatic colony and migrate[56;415;424;425] playing a key role in the process of invasion and subsequent metastasis.

Various classes of proteolytic enzymes such as the matrix-metalloproteinases (MMPs), serine proteinases and cathepsins have all been implicated in the proteolytic process [170;351;426-429]. However, some components are very resistant to proteolytic attacks, and are degraded only by matrix metalloproteinases, particularly the interstitial collagens[170]. Type IV collagen is the main protein of the basement membrane, an insoluble continuous but flexible structure beneath the epithelial linings and around vascular structures which is impermeable to large proteins[430], and it forms the first barriers through which the

invading and metastasizing cells must penetrate[55]. The basement membrane becomes locally permeable to cell movement only during tissue remodelling. Type IV collagen is an important substrate because it constitutes the structural scaffolding upon which the other components of the matrix are assembled. Degradation of this structural protein is believed to be preferentially performed by two specific type IV collagenases/gelatinases- namely, the 72kDa (MMP-2) and the 92kDa (MMP-9) forms[56;431].

Therefore, MMP-9 is of crucial importance in the process of invasion and metastasis since production of basement membrane-degrading enzymes by cancer cells is necessary for invasive growth[55;414;422;432] and direct evidence of its role in tumour progression has been derived from transfection experiments in which the MMP-9 gene in non-metastatic cells endowed them with the ability to metastasize[433].

Differential proteinase expressions have been identified in breast cancer suggesting that distinct proteinase profiles may be involved at different stages of the metastatic cascade, depending on the surrounding components of the ECM[434] and it has been described that production and activation of MMP-9 occurs during the late cancerous stage[77;359;435] which could explain its impact on prognosis in clinically detected invasive breast tumours [157]. It is possible that a somatic mutation leading to an increased level of MMP-9 will alter the balance between host proteases and protease inhibitors and increase the likelihood of a patient's tumour becoming invasive.

Numerous studies have been conducted to localise expression of MMP-9 however it remains controversial as to which cells secrete MMP-9 and which cells regulate its activation in cancer. The expression of MMP-9 in breast tumour tissue has been examined by different methods with varying results, including ELISA[436], northern blotting[437], zymography[417;435;438;439], in situ hybridization[440] and immunohistochemistry[56;440]. Many reports show that MMP-9 is expressed in tumour cells, however it has become apparent that production of MMP-9 is not solely a characteristic of malignant tissue as secretion of MMP-9 has been also been demonstrated in several other cell types[364].

Many studies have shown that MMP-9 is expressed in breast cancer tissue, but there are conflicting results regarding the identity of the cells that express this protease. MMP-9 has continually been identified to some extent in breast tumour cells[55;56;364;434]. However the most significant expression appears to be in the stromal cells, in particular the inflammatory cells, predominantly the macrophages[56;80;364;414;424;434;440-442] and fibroblasts[55;364;434] and to a small degree in endothelial cells[55;414;424] with no MMP-9 expression seen in normal breast tissue[56;364].

There is very little literature on the expression of MMP-9 in breast cancer cell lines, and reports are conflicting. In agreement with our results, RT-PCR has identified MMP-9 expression in SKBR3[417] and MCF-7[443] and no expression seen in MDA-MB-231 cells[99]. In contrast to our results MMP-9 mRNA expression has also been identified in MDA-MB-231 at high[400] and low[401] levels and in T47D cells but not in MCF-7, SKBR3 or ZR75.1[99].

In many cancers MMP-9 expression was not localised to the tumour cells themselves but to the surrounding stromal cells suggesting that tumour cell-host cell communication is critical to the regulation of MMP-9 regulated tumour cell progression and in contrast to the above studies, several reports have shown that direct cell-cell contact is required for MMP-9 expression[99;425;444] and a malignant phenotype[445].

Some studies have found that direct cell-cell interactions between tumour cells and their stroma are essential in the induction of MMP-9 expression in breast tumour-derived stromal fibroblasts. Co-culture of MCF-7 cells with breast tumour derived fibroblasts induced MMP-9 activity[99] in the tumour fibroblasts, shown by RNase protection assays. This was not seen in our results which demonstrated a high level of expression of MMP-9 in MCF-7 cells alone, and no increase in activity upon co-culture of MCF-7 cells with the MRC5 fibroblast cell line. However, many reports have demonstrated that this contact dependant stromal expression of MMP-9 is dependent on the malignancy of the tumour cell line[104;425;444]. An increased invasiveness of malignant cells has been seen on co-cultivation with macrophages due to a TNFα-dependant MMP induction in the macrophage, this induction was not seen with a benign cell line[80]. Therefore co-culture of

the more malignant cell line MDA-MB-231 with human fibroblasts may show an increase in MMP-9 expression and in fact this has been previously shown to increase production of MMP-9 when compared with breast cancer single cell culture or fibroblast single cell culture experiments as demonstrated by zymography and Western immunoblot analysis.[446] Some reports, however, have suggested that direct cell-cell contact is not necessary and that MMP-9 expression is induced in the tumour cell via a soluble factor released from the fibroblasts.[446;447]

With so many studies demonstrating that MMP-9 expression is dependent on the malignancy of the cell line, it fits that MMP-9 expression be associated with a more malignant phenotype[433;441;445;448] with increased metastatic potential[364;415;417;449-456] and so MMP-9 levels may well be a good prognostic marker in cancer.

To date only a few studies have investigated the prognostic value of MMP-9 in breast cancer and, as with the localisation of expression of MMP-9, results are not consistent. The majority of the studies use immunohistochemical techniques. MMP-9 expression has been associated with several aggressive factors[437] such as a higher grade [359;457], presence of distant metastases[157;449], lymph node metastasis[368], and poor survival[157;437;457]. A poor survival in hormone responsive small tumours with an increased stromal expression was also reported.[414] In contrast to the above studies, the same group also reported that increased tumour cell expression of MMP-9 was associated with smaller tumours and a low recurrence rate, favouring survival. When stratified into subgroups based on lymph node involvement (positive or negative) an increased MMP-9 expression was correlated with a favourable prognosis[56] and lower grade and reduced lymphovascular and vascular invasion[434] in the lymph node negative patients. However, several reports have found no association between an elevated MMP-9 expression and prognostic factors.[364;458-460] Macrophage MMP-9 expression has also been implicated in tumour progression by increasing angiogenesis, tumour invasion and metastasis.[83;461]

The inconsistent results of association with disease progression from Immunohistochemistry studies may be explained in part by the use of different

antibodies. Antibodies may be specific for different forms of the MMP-9 protein, for example they may be specific for latent, active, or total MMP-9, or all three. Also, some antibodies may recognise only uncomplexed MMP-9 whereas others may only recognise MMP-9 complexed with TIMP. It is often not made clear which forms of MMP-9 are recognised by the antibodies, making it difficult to compare studies. The importance of the form of the protein recognised by the antibodies was demonstrated by Garbett et al[462] who found the latent form of MMP-9 in both tumour and background breast tissues but only the active form of this enzyme in tumour tissue of paired samples. Davies et al[359] also found that there was a higher proportion of active enzyme with increasing tumour grade.

The antibodies used by Scorilas et al[56] and Baker et al[434] recognize only the latent form of the enzyme and this may explain the apparently positive effect of MMP-9 expression in lymph node negative patients, since it is likely that initially MMP-9 presenting in tumours from node negative patients is in the inactive form and as the tumours become more aggressive the balance shifts towards the active form. This shift to the active form could lead to metastatic progression and increased malignancy, for example via enhanced angiogenic activity and therefore the latent MMP-9 staining disappears[414]. This explanation fits with the evidence of increasing MMP-9 expression in the later stages of carcinogenesis.[77;435]

The increase in positive stromal MMP-9 expression found to predict poor survival in the hormone-responsive, small tumours (i.e. in a group with usually very good prognosis) may be a consequence of interaction with hormones.[414] ERα activation has been shown to lead to tumour progression by stimulating cell growth and invasiveness through increased expression of MMPs, including MMP-9[463] and other studies have also confirmed the relationship between steroid hormones and increased gelatinase activity.[464;465] However there is conflicting evidence showing oestrogen induces a decrease in intracellular and secreted protein levels of MMP-9 and tamoxifen increases protein levels in oestrogen receptor positive MCF-7 cells.[466] Studies have shown that MMP-9 may act as both a positive and negative regulator of angiogenesis[467] by generating pro-angiogenic factors,

such as VEGF, as well as the anti-angiogenic factors such as angiostatin[407;411;468] and therefore Tamoxifen may be associated with an increase in production of anti-angiogenic fragments via modulation of MMP-9.[466;466] Therefore, although an association of of MMP-9 over expression with breast cancer progression is well documented, it is becoming evident that its role in neovascularisation of tumour and proliferation of cancer cells may also be of great significance in tumour growth and progression, and regulation of the activity of these proteinases may to be of great importance at different stages in tumour progression.

To investigate the effect of MMP-9 promoter SNPs on MMP-9 expression levels we performed a series of transient reporter gene transfection experiments with a plasmid containing approximately 2kb of the MMP-9 promoter with one of four allele combinations of two SNPs, a potentially functional SNP, rs3918242 in the MMP-9 promoter which has been reported to lead to an increased expression of MMP-9 and a SNP 241bp upstream that we identified as being strongly associated with distant metastasis, especially in ER-positive patients. Plasmids were transfected into five cell lines but significant results were only obtained with the highly invasive, metastatic MDA-MB-231 breast cancer cell line. Results indicated that the minor allele of rs3918242 caused a decrease in MMP-9 expression compared to the wild type in a highly metastatic cell line conflicting the only published literature on the functional role of this SNP in which the minor allele was demonstrated to cause an increase in MMP-9 expression in macrophages by altering affinity of a puatitve repressor binding site for the repressor protein although to my knowledge there is no published literature validating this finding. TFSEARCH did not show any changes in transcription factor binding sites with the minor allele.

There is a significant amount of literature that has identified an association of the minor allele of this SNP with several diseases however its role in breast cancer has not been extensively investigated and results are conflicting and of low significance with most sample sizes being fairly small. The minor allele (T) has been associated with a significant increase in progression and invasion in a small case control study in the Iranian

population[469] and a moderately increased risk for the TT genotype, or a protective effect of the CC genotype, was observed in a Swedish cohort, however the rarity of risk genotypes limits the prognostic significance[392]. In contrast, the CT/TT genotype was found to have a marginally better prognosis compared to the CC genotype in an Australian cohort, and the T allele was significantly more frequent in patients with ER+ tumours and wild type TP53[365] and no effect in a South Brazillian cohort[470]. A Polish study[471] found the C>T polymorphism had no influence on expression level and suggested that the putative repressor that may bind to the sequence with the C allele is not expressed in breast cancer cells. A further study evaluating effect of this SNP and metastatic spread of breast cancer LN- and LN+ patients of mixed ethnicity found the CT genotype was associated with LN+ disease and this association strengthened when the cohort was stratified to look at Caucasians only. However, there was no association of the T allele with node status in the mixed ethnicity group after the Caucasians were removed suggesting the discrepancies reported between associations in the literature may, in part, be explained by ethnic differences[472]. Further investigation is needed to clarify the role, if any, that this SNP may play in breast cancer.

Although the results did not reach significance, the ER-positive MCF-7 cell line and the NIH-3T3 fibroblast cell line both showed an increase in MMP-9 expression with the A allele (minor allele) of r3918241. This SNP is located in the core of a sequence of a GATA motif which may bind GATA transcription factors increasing MMP-9 transcription[473]. PROMO, a virtual lab for identification of putative transcription factor binding sites in DNA sequences, uses transcription factor binding site data from the TRANSFAC database and indicates the A allele of rs391841 introduces a GATA1 binding site, and TFSEARCH, which also uses data from the TRANSFAC database to predict transcription factor binding sites, predicts the A allele creates an extra GATA-1, Oct-1, GATA-2 and GATA-x binding motifs and ESEfinder shows the A allele may enhance promoter activity and a recent paper has suggested that this may be responsible for increases in expression observed in non-atopic asthma in children[473].

It is known that metastasis requires degradation of type IV collagen (cleaved by MMP-9) however, once past the basement membrane, tumour cells needs to invade the interstitial stroma (types I and III collagen) to reach vasculature and disseminate at distant sites. Types I and III collagen are degraded by MMP-1[357], therefore expression of additional MMPs may also be of importance and the effect of MMP-9 on metastastis may be limited if expression of additional MMPs are low, so it could be more beneficial to investigate relative levels of several MMPs rather than of individual MMPs.

5.1.2.3 The Potential Use of MMP-9 as a Target for Therapy

As a consequence of studies focusing almost exclusively on cancer cells, nearly all of the currently used cancer therapeutic agents target the cancer cells that, due to their inherent genomic instability, possess an evolutionary advantage and frequently acquire therapeutic resistance.[474] In recent years it has become clear that the tumour microenvironment clearly plays a role in the development and progression of cancer. This knowledge has significant implications in cancer therapy since these cells are thought to be more genetically stable than the tumour cells they are less likely to develop an acquired resistance to cancer therapy. A better understanding of the cross-talk between tumour and stroma are needed to help understand the biology of aggressive cells of metastatic tumours. Various features of the metastatic tumour stroma have been identified as potential targets for therapy such as the tumour associated macrophages[475] and rigorous validation of these targets over the next few years may lead to new classes of anticancer agents such as anti-angiogenic, anti-stromal and anti-metastatic drugs becoming common place in clinical practice. Although eradication of a tumour is preferable, it is not always possible with aggressive chemotherapy which can be harmful to the patient. Instead it may be that cancers can be managed long term in a similar way to chronic diseases, with drugs constraining the malignant potential of the cancer cells. Our investigation found that an increase in MMP-9 was associated with an increase in metastasis so this may be a potential target for anticancer therapy. The importance of MMPs in tumour invasion is widely acknowledged. The most compelling evidence of the involvement of MMP-9 in tumour invasion came from a study showing that the overproduction of this

metalloproteinase in non-metastatic rat embryo cells conferred a metastatic phenotype on these cells[433]. Since then, other studies have confirmed its crucial role in tumour development and recent studies have demonstrated that MMP-2 and MMP-9 are expressed not only by cancer cells, but also by surrounding host cells [425;431;476;477] and several studies have shown an association of MMP-9 with various prognostic factors[437;457] suggesting the contribution of the host-derived gelatinase to tumour progression. It has also been shown that it may be possible to revert the malignant phenotype by correcting environmental cues and normalising signal transduction cascades, demonstrating that the microenvironment can dominate over the malignant genotype.[478] Therefore, tumour metastasis may be reduced by inhibiting MMP-9 activity whether the tumour cell itself secretes MMP-9 or whether it is produced by the stromal cells, suggesting the usefulness of MMP-9 inhibitors for protection from metastasis in anticancer therapy. [157;423;479]

Current therapeutic strategies using synthetic inhibitors of MMPs have already demonstrated the clinical potential of regulating protease activity.[364;480]

It is more efficient to decrease MMP-9 levels by preventing transcription of a low number of gene copies than to neutralise a high number of proteolytically active molecules that are secreted by cells over-expressing MMP-9. However, MMP-9 is upregulated via multiple signalling pathways and some inhibitors may block more than one pathway to repress MMP-9 expression.[417] Several transcription factors known to regulate MMP-9 gene expression have pleiotropic effects, for example AP-1 or NF-κB antagonists can inhibit phosphorylation of c-Jun and the expression of several inflammatory genes such as IL-2, IFN-γ and TNF-α therefore these inhibitors will likely block other members of the MMP family that are either under transcriptional control of these cytokines or contain recognition motifs in their promoter that are specific for these transcriptional elements, for example MMP-1, -3, -13 and -14 which are induced by TNF-α and contain AP-1 responsive elements meaning specificity of inhibitors is crucial in achieving success. [481]

Many matrix metalloproteinase inhibitors (MMPI's) have been investigated for anticancer properties. MMPI's are thought to inhibit primary tumour invasion and metastasis[60] but

most are broad spectrum MMP inhibitors such as Batimastat and the second generation Marimastat and have many musculoskeletal side effects with inflammation of the tendons and joints. Some MMPs have a protective role against cancer and therefore targeting the whole family will undoubtedly fail therapeutically which may explain the limited efficacy seen in the clinic. [423] There are recent reports of gelatinase (MMP-2 and MMP-9) inhibitors in animal models and these have demonstrated clinical potential preventing the migration of tumour cells, tumour growth and invasion in animal models and prolonging survival in xenograft-bearing animals[482].

In the future improved diagnostics and public awareness will lead to an increased number of patients presenting with earlier stage disease, a cohort of which will be at high risk for tumour progression. Patients who are at a high risk of cancer progression and development of metastases may benefit from therapies specifically targeting MMP-9 expression while sparing the expression of the 'protective' collagenases[420]. Stratifying patients into sub groups in this way will help treatment decisions and allow a more personalised targeted treatment of breast cancer. Scorilas et al[56] found that increased MMP-9 levels were associated with prognosis in breast cancer and suggested that the over-expression of MMP-9 in breast cancer may be also used as a marker to subdivide node negative breast cancer patients in order to determine optimal treatment modality. This could be useful clinically as 20-35% of node-negative breast cancer patients relapse[377], and determination of MMP-9 expression may help to separate patients with better prognosis from those with a poor prognosis and identify patients which may benefit from different treatment modalities that may include protease inhibitors.[56]

This study found that the SNPs in the promoter of the MMP-9 gene that were associated with distant metastasis were in fact more significant in ER+ patients, reaching significance after correction for multiple testing, and this could be very useful in identifying a subgroup of ER+ patients at high risk of developing distant metastases who may benefit from adjuvant anti-metastatic agents. This could be particularly useful for those ER-positive patients who do not respond to hormonal therapy.

Approximately 70% of all primary breast tumours express ER-α, known to be important in the development and course of the disease. The SERM (selective oestrogen receptor modulator) Tamoxifen targets and inhibits ER- α and is the standard endocrine treatment for hormone receptor positive breast cancer in both the initial adjuvant therapy and as treatment of patients with metastatic disease. It has been used for more than 30 years to treat, and more recently, to prevent breast cancer.[483] However, even among ER+ tumours, response to Tamoxifen is variable. When diagnosed at an early stage adjuvant systemic tamoxifen therapy can cure approx 10% of the patients[484]. In recurrent disease approx 50% of patients have no benefit from Tamoxifen (intrinsic resistance)[485;486]. From the other half of the patients who initially respond to therapy a large proportion will eventually develop progressive disease due to acquired resistance.[485] Tamoxifen resistance is a major cause of death in patients with recurrent breast cancer. Current clinical factors can correctly predict therapy response in only half of the treated patients.[485]

In recent years several studies have been published suggesting that response to anticancer therapies such as Tamoxifen depends not only on the characteristics of the tumour but also on characteristics of the host, for example a genetic polymorphism in the CYP2D6 gene is associated with intrinsic resistance to Tamoxifen predicting altered Tamoxifen metabolism and a poorer outcome than expected with the wild-type genotype[483] thus the future of personalised treatment approaches will involve analysis of both the tumour and host characteristics.

Umar et al 2009[485] identified a putative protein profile associated with Tamoxifen resistance, and the top discriminating protein was the extracellular matrix metalloproteinase inducer EMMPRIN. Levels of EMMPRIN were higher in the therapy resistant tumours and associated with earlier disease progression. EMMPRIN was validated in an independent patient cohort and was significantly associated with resistance to Tamoxifen therapy and a shorter time to progression upon Tamoxifen

treatment in recurrent breast cancer. This further supports the role of matrix metalloproteinases in breast cancer progression.

Chapter 6 Conclusions

Until recently cancer research has been focussed on the tumour cell itself, now the focus is moving towards the role of the microenvironment of the tumour and how the environment may be 'permissive' for cancer growth, progression and metastasis. It is likely that an individual's genetic make-up may influence susceptibility to cancer and also the type and aggressiveness of the tumour that may develop. Identifying genetic variants that may predispose to cancer or to certain types of cancer will be invaluable in cancer treatment and maybe even prevention.

This study aimed to identify SNPs that may predispose to cancers with a certain phenotype, such as oestrogen receptor status, or indicate aggressiveness, for example grade or propensity to metastasise. We confirmed previously reported associations of the TOX3 gene and SNPs rs2981582 and rs1219648 in the FGFR2 gene with ER-positive disease and identified a novel association of rs1801516 in the ATM gene with ER-negative disease. We also identified several SNPs in the MMP-9 gene that were associated with distant metastasis. As metastasis is the major cause of morbidity and death for cancer patients and there is a wealth of literature associating increased MMP-9 levels with cancer, we investigated this further looking at the effect of two MMP-9 promoter SNPs on expression levels using Luciferase reporter assays. The minor allele of rs3918242 decreased MMP-9 expression 1.5 fold compared to the wild type in the highly metastatic MDA-MB-231 cell line, whereas the minor allele of rs3918241 increased expression of MMP-9 in the ER-positive MCF-7 cell line and the NIH-3T3 fibroblast cell line, although not significantly. The minor allele of rs3918241 introduces a GATA1 binding site which may be responsible for the increase in expression observed in ER-positive cells. The minor allele of rs3918242 has been reported to alter a repressor binding site[394]. Further investigation is needed to identify which of these SNPs, or indeed any of the other SNPs they are in LD with, are responsible for the functional effect observed.

When stratifying patients on the basis of ER status we found that the three MMP-9 promoter SNPs, including rs3918241, significantly associated with the presence of distant

metastasis were even more significant amongst the ER-positive group, remaining significant after stringent correction for multiple testing. Sex steroids play a dominant role in breast carcinogenesis and there is evidence of an interaction of oestrogen and MMP-9 however the mechanisms are still largely unknown.

It would be of interest to genotype rs3918242 in our cohort and repeat the association with distant metastasis to establish how significant the effect of this SNP is and if the minor allele is associated with the presence of metastasis or, as our functional studies indicate, is protective against metastasis.

Tamoxifen is known to increase MMP-9 expression in hormone responsive MCF-7 cells[466] so repeating the transfection experiments in MCF-7 cells in the presence of Tamoxifen may help increase the significance of the results.

Electrophoretic mobility shift assays could be used to assess whether either of the promoter SNPs investigated here alter nuclear factor binding to help identify if either of these SNPs are responsible for an alteration in MMP-9 expression.

The mean survival time in our cohort was only 2.4 years so association of the SNPs with survival was not studied in depth here, however association with survival could be further investigated using Cox regression analysis to identify any SNPs that may have a significant effect on risk, over and above the known risk factors including grade, ER status, tumour diameter and pathological nodal status.

In summary, the search for risk alleles may identify alleles contributing only a very small overall risk that may not be of use clinically, but in combination may prove useful in identifying individuals with an increased risk of breast cancer. It may be more useful to further investigate how SNPs relate to prognostic features allowing stratification of breast cancer patients based on the type or aggressiveness of cancer they may develop. Current markers correctly predict therapy response in only half of treated patients. Approximately one third of ER-positive tumours are refractory to Tamoxifen[486] and in recurrent disease approximately half of all patients treated will have no benefit from Tamoxifen (intrinsic

resistance)[485;486] and of those that initially respond, a high percentage will develop an acquired resistance. Identifying further biomarkers that may help predict response to treatment or even act as a target for drugs will help guide treatment decisions, for example by identifying a subset of ER-positive patients who may be likely to develop metastatic disease, will determine patients who could benefit from treatment with adjuvant anti-metastatic therapy giving a more personalised approach to therapy.

Chapter 7 Future Work

This study has identified some novel associations of SNPs in genes with various aspects of breast tumour biology, however all of these associations need to be validated. In order to validate these results SNPs that reached significance in this study, or even borderline significance, need to be replicated in another cohort. The remaining 2000 patients in the POSH study would be an ideal cohort in which to validate the results as it is a unique, large, well characterised cohort with data on survival, pathology data and treatment received for each patient allowing the same associations with tumour characteristics to be made as in the present study.

This study also identified a strong association between SNPs in the promoter region of the MMP-9 gene with the presence of distant metastasis. The literature is very conflicting regarding the cellular localisation of MMP-9 (Figure 15). In several cancers, including colon[431] and skin[487], there is a lot of evidence for a significant stromal expression of MMP-9, particularly in the tumour associated macrophages. In breast cancer several studies have identified stromal expression to varying extents[359;364;368;414;442], although some studies have shown strongest expression in the carcinoma cells[55;56;157;377]. Many of the previous studies have used immunohistochemistry to study the expression of MMP-9 and the different expression patterns observed might be explained by the use of different antibodies with different specificities for the active, latent or complexed forms of MMP-9, or all three and often it is not made clear which the antibodies regognise making comparisons between studies difficult. The majority of studies have also been conducted in a relatively small number of samples only.

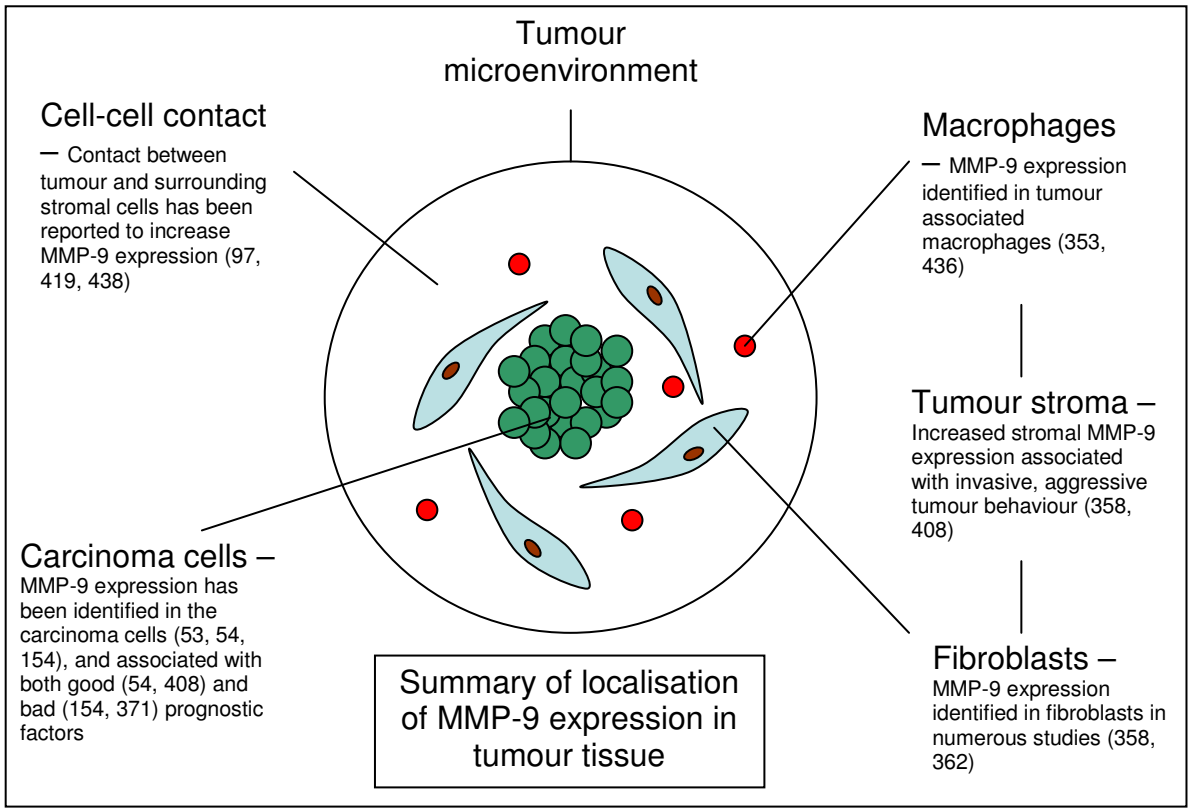


Figure 15: Schematic representation of the tumour microenvironment summarising literature documenting the cellular localisation of MMP-9

In order to establish the cellular localisation and expression levels of MMP-9 protein, immunohistochemistry with specific antibodies needs to be performed in a large number of breast tissue samples. The POSH cohort will again provide a very good cohort for this, with a large number of breast tissue samples available to use for such a study. Breast tissue samples from approximately 1000 of the POSH patients have been called and over 600 so far have been assembled on tissue microarrays. Formalin fixed paraffin embedded (FFPE) tissue gives good preservation of tissue morphology so it is usually relatively easy to specify the cellular localisation of the stain in this tissue. Antibodies to MMP-9 would need to be tested on whole tumour sections in a range of breast cancers from the pathology laboratory (approximately 20-30 cases) to review the staining patterns and optimise the conditions. Once happy that the stain works optimally the TMAs could be stained and scored on intensity of expression. Using tissue samples from this cohort will then give us the opportunity to correlate the expression data with various aspects of

tumour biology such as oestrogen receptor status, propensity to metastasise and survival and also to associate with genotype data to investigate how expression levels vary with the presence of the MMP-9 promoter SNPs. As treatment such as chemotherapy and adjuvant hormonal therapy can alter MMP-9 expression[377] it will be of benefit to know the treatment received for each patient when making comparisons. This work would require the help of an experienced histopathologist to optimise the stain and agree a scoring system. In situ hybridisation could also be performed to investigate MMP-9 DNA or mRNA levels in the tumour tissue to compare with the protein levels identified.

Several studies have documented cell-cell contact is needed between tumour and stroma cells to induce MMP-9 expression in the host[425]. Although co-culture of MRC5 and MCF-7 did not appear to increase MMP-9 expression in RT-PCR, co-culture of fibroblasts with a more invasive cell line such as MDA_MB_231 might show an increase in MMP-9 as it has been demonstrated that co-cultures with more malignant cell lines only promote MMP-9 expression[104;425]. Previous studies have also shown that orthotopic fibroblasts are required to induce an increase in MMP-9[101] so co-culture with breast fibroblasts may significantly alter MMP-9 expression, which can be identified by RT-PCR or qRT-PCR. Transfections with the plasmids created in this study could be performed in either the cancer cell line or the fibroblast cell line, and after 24hrs co-cultured with either untransfected fibroblasts or cancer cells respectively to investigate the effect of co-culture on MMP-9 expression, and allow identification of the cell type responsible for the most significant increase in MMP-9. These experiments would demonstrate the important role of the host cells in tumourigenesis and the effect of the promoter polymorphisms in host cells.

An increased expression of MMP-9 has been associated with breast cancer in numerous studies. Carcinoma cell expression of MMP-9 has been associated with both with good[56;414] and bad[157;377] prognostic factors. Stromal cell expression of MMP-9 has been associated with poor prognostic factors[414] therefore there could be very cell specific effects of MMP-9 and of the promoter SNP. It would be of interest to perform

transfections with our four plasmids in a macrophage cell line to see if these cell specific differences account for the contradictory result observed to that previously reported[394].

Functional studies were conducted to investigate the role of two promoter SNPs, one previously reported and one in very strong linkage with it identified in our association study. We found a significant increase in MMP-9 expression with the wild type allele of the previously reported SNP rs3918242 meaning the minor allele, leading to a decrease in MMP-9 expression, was associated with the presence of distant metastasis. This unexpected result is difficult to interpret in terms of the current concept of the molecular mechanisms involved in tumour invasion and metastasis, and conflicts the results of the only published data on its functional effect. However there is data clearly challenging the classical beliefs that MMP-9 is solely a pro-metastatic enzyme[488;489], promoting tumour invasion exclusively by modulating the remodelling of the ECM, for example MMPs may also regulate angiogenesis either positively through their ability to activate proangiogenic factors such as VEGF, or negatively via the generation of angiogenesis inhibitors such as angiostatin, endostatin and tumstatin[407;488;490] (Figure 16) therefore our result could be explained by MMP-9 having an anti-angiogenic effect. It has been shown that oestrogen decreases MMP-9 expression levels and Tamoxifen increases them[466]. Since Tamoxifen is known to have anti-angiogenic effects by generating endostatin via MMP-9 production[466], this supports the hypothesis of MMP-9 playing an anti-angiogenic role. The potential role of MMP-9 on angiogenesis could be investigated by transfecting the MDA_MB_231 cell line with MMP-9 plasmids and measuring levels of angiostatin and endostatin produced by the cells. Mouse xenograft models using breast cancer cells overexpressing MMP9 from a stably transfected plasmid could also be used however these would be complex to set up.

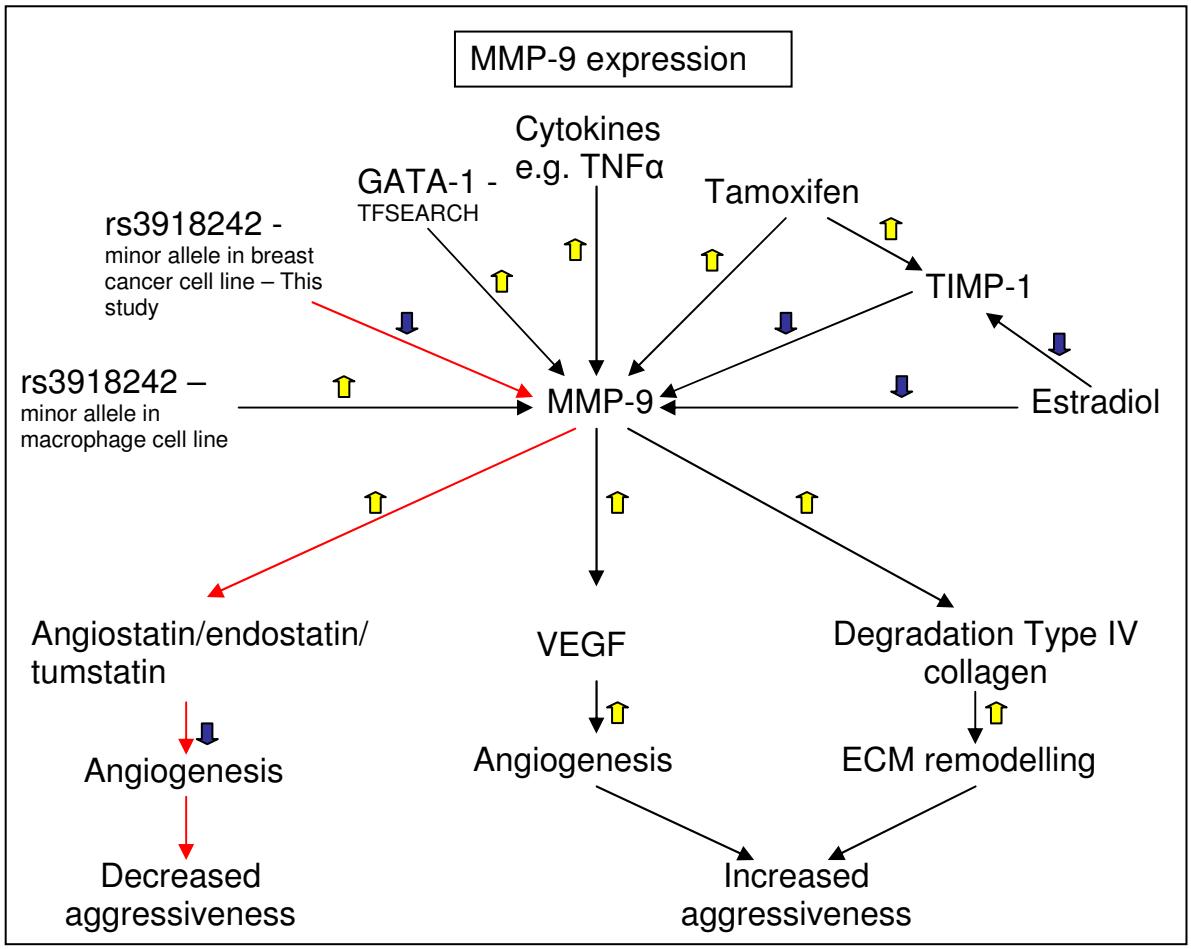


Figure 16: Network diagram of factors affecting MMP-9 expression and their potential consequences. Yellow arrows indicate an increase in expression and blue arrows a decrease. Red arrows indicate the possible effect of the promoter SNP rs3918242 identified in this work.

Investigation of the MMP-9 promoter region using the TFSEARCH database revealed that the promoter SNP rs3918241 introduced a GATA-1 binding site in the promoter sequence. To investigate the effect of GATA-1 in relation to our promoter SNPs GATA-1 in breast cancer cell lines could be inhibited using siRNA to see if the effect is still observed in the absence of GATA-1 or cells could be co-transfected with both the plasmid reporter vectors created in this study and a GATA-1 expressing plasmid to see if this has any SNP-dependent effects. In addition, chromatin immunoprecipitation (ChiP) experiments or electrophoretic mobility shift assays (EMSA) could be performed to investigate the location of DNA binding sites for GATA-1 protein and determine the protein-DNA interactions that may be occurring.

Chapter 8 Papers and Abstracts

8.1 Papers

- Named author on a paper outlining the POSH study protocol[396].

- Named author on paper titled 'The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer'[491].

8.2 Abstracts and Posters

- Poster presented at the Cancer Sciences Division Conference, June 2007 and the British Society for Human Genetics, Sept 2007 titled 'Inherited mutations in BRCA1 are much more frequent than in BRCA2 amongst young breast cancer cases'.

- An abstract on the work carried out in year 2, titled 'An inherited genetic polymorphism in the oestrogen receptor gene is associated with poor prognosis in young breast cancer patients' was presented at the Cancer Sciences Division Conference, July 2008.

- Poster presented at the National Cancer Research Institute conference titled 'An inherited genetic polymorphism in the oestrogen receptor gene is associated with poor prognosis in young breast cancer patients' October 2008.

- Oral presentation titled 'The Role of Genetic Factors in Breast Cancer Aetiology' was given at the Faculty of Medicine, Health and Life Sciences Postgraduate Conference, University of Southampton, June 2009 for which the prize of 'best oral presentation' was awarded.

Chapter 9 Appendices

9.1 Appendix 1 - Sequenom

Sequenom offer a genotyping service called iPLEX Gold. This uses the MassARRAY® platform developed by Sequenom to specifically meet the requirements of moderate to high throughput genetic analysis and is suitable for genotyping tens to hundreds of SNPs in hundreds to thousands of samples. It offers cost effective, high quality custom SNP genotyping combining single-base primer extension biochemistry and MALDI-TOF mass spectrometry. The service includes primer and assay design through to the production of genotypes. All steps involved are highly automated and are tracked using a laboratory management system with bar coding.

9.1.1.1 SNP analysis using Sequenom MassARRAY Genotyping

Assay Design – The MassARRAY® Designer software can automatically design both PCR and MassEXTEND® primers for multiplexed assays using sequence information from public databases. The primer design sequence is automatically transferred into an oligonucleotide requisition form for ordering of the primer.

multiplex PCR – this is the first step of the genotyping process. It is carried out to generate short PCR products (> 100 bp) containing one SNP or insertion deletion.

Excess dNTPs are then removed from the reaction by incubation with 0.3 U shrimp alkaline phosphatase (USB).

MassEXTEND® Reaction - MassEXTEND® is a primer extension reaction in which an oligonucleotide primer anneals immediately upstream of the polymorphic site being genotyped. The extension of the primer is according to the sequence of the variant site and can be a single complementary base or a series of complementary bases. This results in an allele-specific difference in mass between extension products.

The genotype data is generated by analysis of minisequencing reaction products by mass spectrometry.

Genotype Calling and Results - The extension product is then spotted onto a 384 well spectroCHIP® and the spectoCHIP placed into the MALDI-TOF (Matrix Assisted Laser Desorption/Ionization Time-of-Flight) mass spectrometer. Oonce inside, a laser is fired onto the pads and an electrical charge is applied, attracting the extended oligonucleotides to a detector. The time taken to reach the detector (TOF, time of flight) is solely based upon the mass of the oligonucleotide. This mass difference allows the data analysis software to differentiate between SNP alleles. The resulting spectra is converted into genotype using a software system called the SpectroTYPER-RT (RT for real-time). Through the assay design process, the software is informed of the expected mass of all possible extension products. If a peak is seen where expected the software calls the allele, but if no peak is seen, the laser is fired again until a spectra is achieved containing the expected peaks.

Sequenom offer different 'tiers' that vary slightly in assay design, validation and genotyping product. For our study we opted for tier 2. This does not require that DNA samples undergo a functional QC, there will be up to 2 attempts made to generate a data point, and customers will be charged on a possible data point basis rather than delivered genotype basis and for all assay designs and materials, regardless of whether all the SNPs type successfully.

9.2 Appendix 2 - Sequenom Tier Descriptions

Assay Design and Validation

Tier 1: Assays are both designed and validated.

A.) Assays are designed using Spectrodesigner for the first two attempts. Manual intervention (if necessary) is used for the third attempt.

B.) Validation involves testing the assays on control samples for:

1. Robustness. Each assay is tested to ensure that it is capable of generating at least 90% data on reliable DNA samples.
2. Reproducible. Each assay is tested to ensure that it generates reproducible results (at least 99.3%)
3. Accurate. Each assay is tested to ensure that it generates accurate results. Depending on the project accuracy may be tested by:
    a. comparison with genotype results from independent platforms
    b. evaluation of Mendelian inheritance
    c. evaluation of Hardy-Weinberg values

C.) Tier 1 is required if the client desired to have guaranteed data delivery in production. That is, if we promise to deliver at least 90% of the possible data for a given assay and at least 95% of the data for a given project, we must ascertain beforehand that the assay is capable of meeting those standards.

D.) As the number of production DNA samples increase, it may actually be less expensive for the client to choose Tier 1 as opposed to Tier 2. This is because non-functioning assays will be removed and not applied in production.

E.) Customer is charged only for assays that validate; they are not charged for failed assays.

Tier 2: Assays are designed and applied directly into production

A.) Saves time, as production can begin immediately

B.) Useful when an appropriate validation cohort is not available

C.) May be more cost effective; especially with projects involving fewer DNA samples

D.) Cannot be used in conjunction with Tier 1 genotyping (guaranteed delivery)

E.) Customer is charged for all assay designs and materials, regardless of whether they end up working.

Tier 3: Customer designs are used.

A.) Allows customer to transfer MassARRAY designs from their own labs or other service providers to Sequenom.

B.) Allows customer to use designs from previous service projects within the Services Department.

1. If assays were validated (Tier 1) previously, Tier 1 genotyping (guaranteed delivery) is possible.
2. If assays were developed under Tier 2 previously, Tier 1 genotyping (guaranteed delivery) is not possible.

C.) Customer is charged a modest assay management fee for the necessary bioinformatics work

Genotyping Production

Tier 1: Guaranteed Delivery

A.) Requires that assays were developed under Tier 1 above
B.) Requires that DNA samples undergo a functional QC
C.) Guarantees that the customer will get at least 90% of the data on a per assay basis (ie no assays will be delivered with less than 90% of the possible data).
D.) Guarantees that the customer will get at least 95% of the possible data for the entire project (DNA samples x validated assays x .95).
E.) Customer is charged only for data that is delivered.

Tier 2: Attempted Genotype Pricing

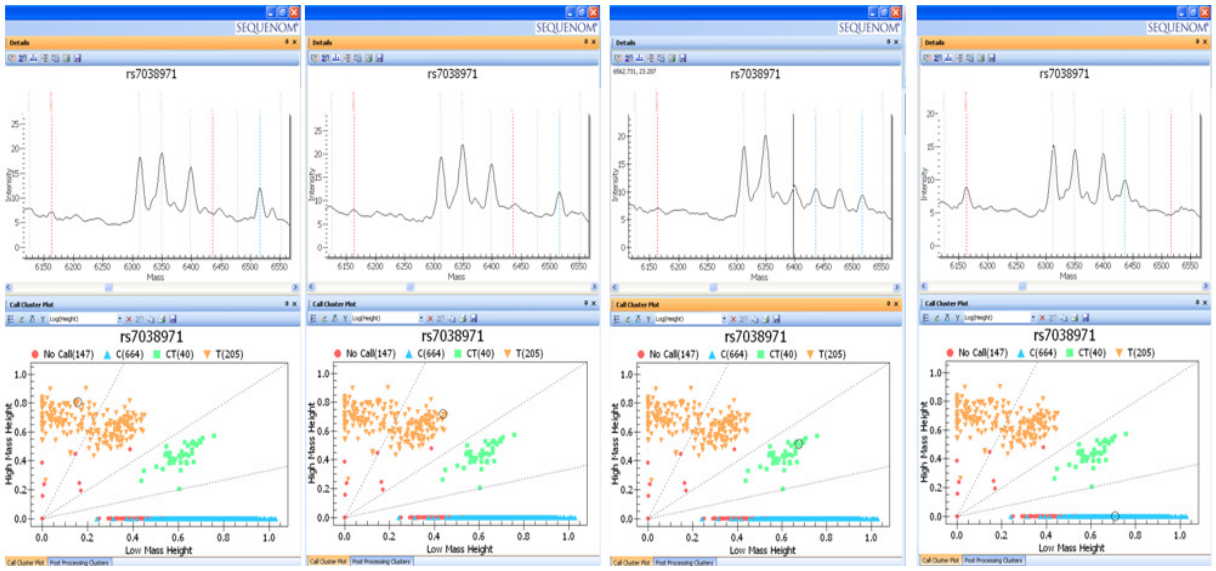
A.) Can use any tier of assay development above
B.) Does not require that DNA samples undergo a functional QC
C.) Makes no guarantee regarding data delivery
D.) Services will make up to two attempts to generate a data point.
E.) Customers are charged on a possible data point basis rather than delivered genotype basis. Thus, there is no difference between the estimated project cost and the maximum project cost as with Tier 1 genotyping.
F.) Turnaround time will typically be faster than with Tier 1 genotyping
G.) Cost will typically be cheaper than with Tier 1 genotyping.
H.) For many applications, this option may offer the greatest value.

Tier 3: Single Pass Genotype Pricing

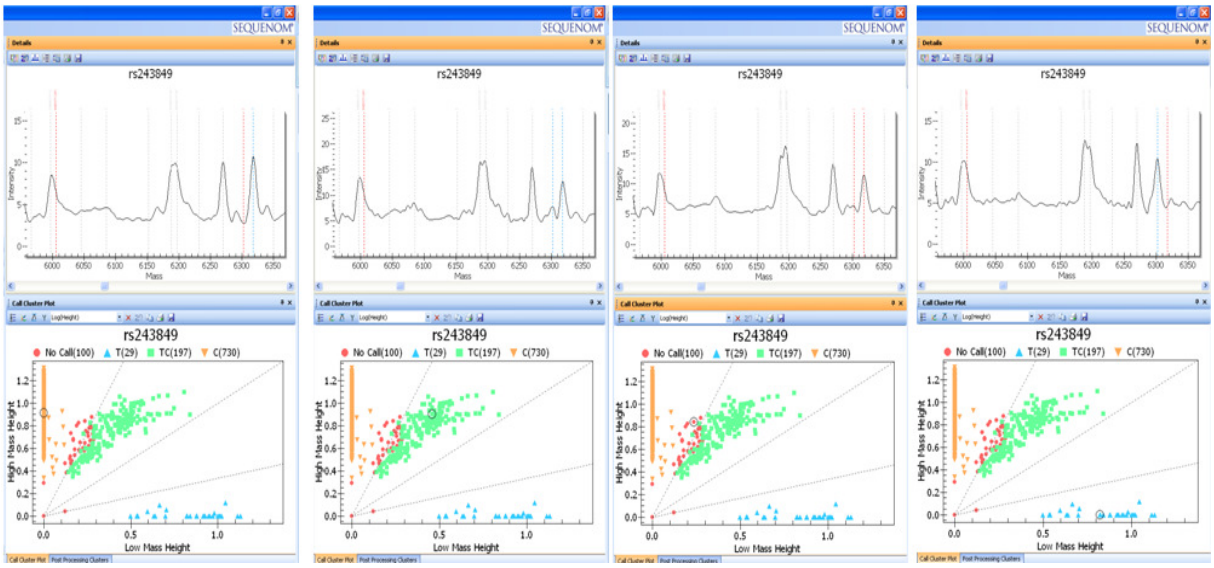
A.) Can use any tier of assay development above
B.) Does not require that DNA samples undergo a functional QC
C.) Makes not guarantee regarding data delivery
D.) Services will make a single attempt to generate a data point.
E.) Customers are charged on a possible data point basis rather than delivered genotype basis. Thus, there is no difference between the estimated project cost and the maximum project cost as with Tier 1 genotyping.
F.) Turnaround time will typically be faster than with either Tier 1 or Tier 2 genotyping
G.) Cost will be cheaper than with Tier 1 or Tier 2 genotyping.
H.) This option is recommended for quick first pass scans, where missing data will not adversely impact results.

9.3 Appendix 3 - Sequenom Cluster Plots

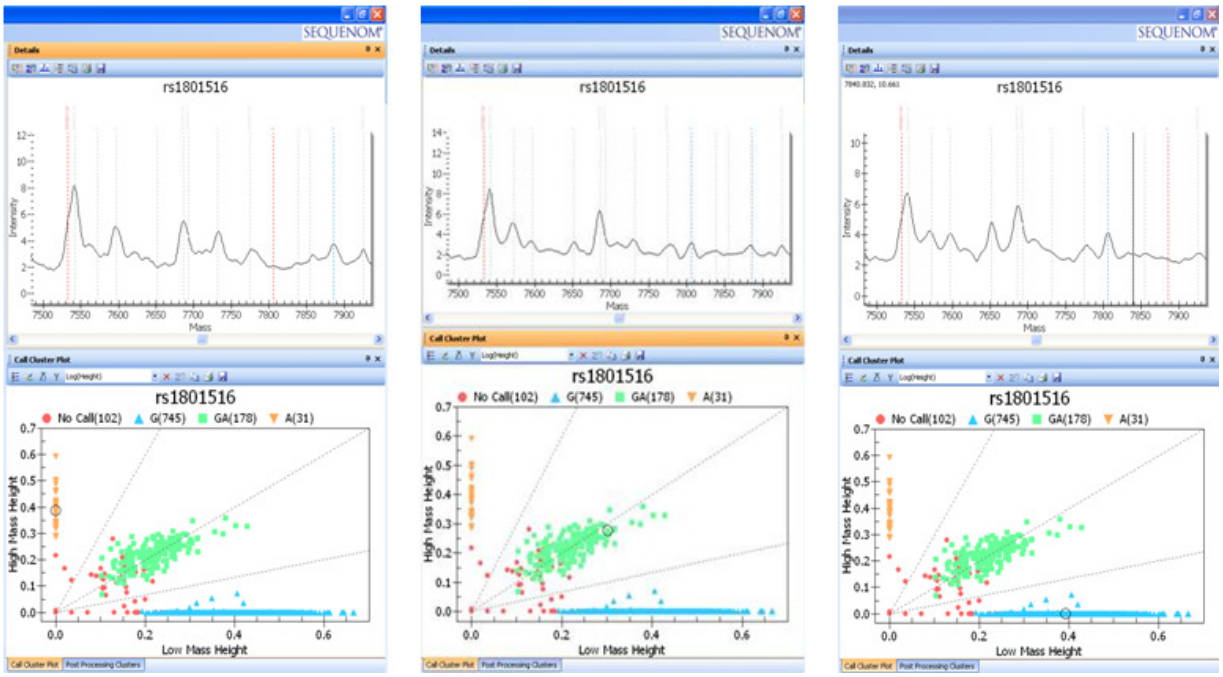
rs7038971 – the high mass homozygous cluster does not look like a heterozygous on the spectrum. The low mass peak(s) are significantly lower than in the true hets, in green.



rs243849 – low mass allele peak does not meet peak criteria and extension primer conversion appears complete



rs1801516 – Both peaks are present for het calls and homozygote calls are clear.



9.4 Appendix 4 - DNA Extraction Reagents

Nuclei lysis buffer – (10mM Tris, 400mM NaCl, 2mM sodium EDTA)To make up 500ml; 5ml 1.0M tris pH 8.0, 40ml 5.0M sodium chloride, 2ml 0.5M EDTA pH 8.0, 453ml UHQ StillPlus water. Stored at room temperature.

Erythrocyte lysis buffer x10 – To make up 500ml; 5.05gm potassium hydrogen carbonate, 41gm ammonium chloride, 1ml 0.5M EDTA pH 8.0, make up to 500ml with UHQ StillPlus water (made in advance by Human Genetics staff, autoclaved for 15mins at 121˚C. Stored at room temperature until needed. Place in fridge overnight to cool to 4˚C ready for use the next day and dilute 1 in 10 for use.

10% SDS (sodium dodecyl sulphate)- 10gm SDS in 100ml UHQ StillPlus water, stored at room temperature.

Protease – (40mg/ml) 800mg made up to 20ml with UHQ StillPlus water, stored in -20˚C freezer.

Saturated salt (6M sodium chloride) – 35gm NaCl in 100ml UHQ StillPlus water, stored at room temperature.

Ethanol 70% - 350ml ethanol plus 150ml water, stored at room temperature.

TE buffer – (10mM Tris/1.0mM EDTA buffer pH 7.5) To make up 500ml; 5ml 1.0M Tris pH 7.5, 1ml 0.5M EDTA pH 8.0, 494ml UHQ StillPlus water (usually made up in advance by Human Genetics staff, autoclaved for 15mins at 121˚C in 80ml volumes). Stored at room temperature.

2% Virkon™ - 40gm dissolved in 200ml tap water

Chapter 10 References

(1) 2010 Available from: URL:http://info.cancerresearchuk.org/cancerstats/types/breast/incidence/

(2) Parkin DM. International variation. *Oncogene* 2004; 23(38):6329-6340.

(3) von Hansemann D. Über asymmetrische Zellteilung in Epithelkrebsen und deren biologische Bedeutung. *Virchow's Arch Path Anat* 1890; 119:299-326.

(4) Available from: URL:http://www.statistics.gov.uk/cci/nugget.asp?id=915

(5) Parkin DM, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. *CA Cancer J Clin* 2005; 55(2):74-108.

(6) Nelson NJ. Migrant Studies Aid the Search for Factors Linked to Breast Cancer Risk. *J Natl Cancer Inst* 2006; 98(7):436-438.

(7) Boveri T. Concerning the Origin of Malignant Tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J Cell Sci* 2008; 121(Supplement_1):1-84.

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

(9) Slijepcevic P. Familial breast cancer: recent advances. *Acta Medica Academica* 2007; 36:38-43.

(10) Vogelstein B, Kinzler KW. The multistep nature of cancer. *Trends Genet* 1993; 9(4):138-141.

(11) Vasen HFA, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. *Gastroenterology* 1999; 116(6):1453-1456.

(12) Wooster R, Weber BL. Breast and ovarian cancer. *N Engl J Med* 2003; 348(23):2339-2347.

(13) Evans DG, Shenton A, Woodward E, Lalloo F, Howell A, Maher ER. Penetrance estimates for BRCA1 and BRCA2 based on genetic testing in a Clinical Cancer Genetics service setting: risks of breast/ovarian cancer quoted should reflect the cancer burden in the family. *BMC Cancer* 2008; 8:155.

(14) Willems PJ. Susceptibility genes in breast cancer: more is less? *Clinical Genetics* 2007; 72:493-496.

(15) Watson P, Lynch HT. Cancer risk in mismatch repair gene mutation carriers. *Familial Cancer* 2001; 1(1):57-60.

(16) Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG et al. The sequence of the human genome. *Science* 2001; 291(5507):1304-1351.

(17) Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 1990; 250(4988):1684-1689.

(18) Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994; 266(5182):66-71.

(19) Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* 1995; 378(6559):789-792.

(20) The International HapMap Project. *Nature* 2003; 426(6968):789-796.

(21) Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature 447(7148):1087-93,* 2007.

(22) Easton DF. How many more breast cancer predisposition genes are there? *Breast Cancer Res* 1999; 1(1):14-17.

(23) Dimri GP. In Search of Breast Cancer Culprits: Suspecting the Suspected and the Unsuspected. *Breast Cancer: Basic and Clinical Research* 2008; 2008(BCBCR-1-Dimri).

(24) Easton DF, Eeles RA. Genome-wide association studies in cancer. *Hum Mol Genet* 2008;ddn287.

(25) 2010 Available from: URL:http://www.abpi.org.uk/publications/pdfs/Target/TargetBreastCancer.pdf

(26) Ziegler RG, Hoover RN, Pike MC, Hildesheim A, Nomura AM, West DW et al. Migration patterns and breast cancer risk in Asian-American women. *J Natl Cancer Inst* 1993; 85(22):1819-1827.

(27) Deapen D, Liu L, Perkins C, Bernstein L, Ross RK. Rapidly rising breast cancer incidence rates among Asian-American women. *Int J Cancer* 2002; 99(5):747-750.

(28) MacMahon B. A Biologic Framework For The Risk Factors Of Breast Cancer. *Advances In Oncology* 1994; 10(4):3-9.

(29) Nathanson KL, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. *Nat Med* 2001; 7(5):552-556.

(30) Hartman M, Lindstrom L, Dickman PW, Adami HO, Hall P, Czene K. Is breast cancer prognosis inherited? *Breast Cancer Research 9(3):R39,* 2007.

(31) Tutt A, Ashworth A. The relationship between the roles of BRCA genes in DNA repair and cancer predisposition. *Trends Mol Med* 2002; 8(12):571-576.

(32) Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005; 434(7035):917-921.

(33) James PA, Doherty R, Harris M, Mukesh BN, Milner A, Young MA et al. Optimal selection of individuals for BRCA mutation testing: a comparison of available methods. *J Clin Oncol* 2006; 24(4):707-715.

(34) Walsh T, King MC. Ten genes for inherited breast cancer. *Cancer Cell* 2007; 11(2):103-105.

(35) Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 2002; 31(1):33-36.

(36) Smith P, McGuffog L, Easton DF, Mann GJ, Pupo GM, Newman B et al. A genome wide linkage search for breast cancer susceptibility genes. *Genes Chromosomes Cancer* 2006; 45(7):646-655.

(37) Baynes C, Healey CS, Pooley KA, Scollen S, Luben RN, Thompson DJ et al. Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk. *Breast Cancer Res* 2007; 9(2):R27.

(38) Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 2007; 39(3):352-358.

(39) Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. *J Natl Cancer Inst* 2006; 98(19):1382-1396.

(40) Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics 39(7):870-4,* 2007.

(41) Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics 39(7):865-9,* 2007.

(42) Hedenfalk I, Ringner M, Ben-Dor A, Yakhini Z, Chen Y, Chebil G et al. Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc Natl Acad Sci U S A* 2003; 100(5):2532-2537.

(43) Nordgard SH, Johansen FE, Alnaes GI, Naume B, Borresen-Dale AL, Kristensen VN. Genes harbouring susceptibility SNPs are differentially expressed in the breast cancer subtypes. *Breast Cancer Res* 2007; 9(6):113.

(44) Onay VU, Briollais L, Knight JA, Shi E, Wang Y, Wells S et al. SNP-SNP interactions in breast cancer susceptibility. *BMC Cancer* 2006; 6:114.

(45) Savas S, Ozcelik H. Phosphorylation states of cell cycle and DNA repair proteins can be altered by the nsSNPs. *BMC Cancer* 2005; 5:107.

(46) Savas S, Taylor IW, Wrana JL, Ozcelik H. Functional nonsynonymous single nucleotide polymorphisms from the TGF-beta protein interaction network. *Physiol Genomics* 2007; 29(2):109-117.

(47) Chakravarti A. It's raining SNPs, hallelujah? *Nat Genet* 1998; 19(3):216-217.

(48) Fagerholm R, Hofstetter B, Tommiska J, Aaltonen K, Vrtel R, Syrjakoski K et al. NAD(P)H:quinone oxidoreductase 1 NQO1[ast]2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. *Nat Genet* 2008; 40(7):844-853.

(49) Schroth W, Antoniadou L, Fritz P, Schwab M, Muerdter T, Zanger UM et al. Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. *J Clin Oncol* 2007; 25(33):5187-5193.

(50) Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer. *N Engl J Med* 2008; 358(26):2796-2803.

(51) Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol* 2001; 21(1):1-18.

(52) Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered* 2003; 56(1-3):73-82.

(53) Lakhani SR, van d, V, Jacquemier J, Anderson TJ, Osin PP, McGuffog L et al. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. *J Clin Oncol* 2002; 20(9):2310-2318.

(54) Garber JE, Goldstein AM, Kantor AF, Dreyfus MG, Fraumeni JF, Jr., Li FP. Follow-up study of twenty-four families with Li-Fraumeni syndrome. *Cancer Res* 1991; 51(22):6094-6097.

(55) Soini Y, Hurskainen T, Hoyhtya M, Oikarinen A, Autio-Harmainen H. 72 KD and 92 KD type IV collagenase, type IV collagen, and laminin mRNAs in breast cancer: a study by in situ hybridization. *J Histochem Cytochem* 1994; 42(7):945-951.

(56) Scorilas A, Karameris A, Arnogiannaki N, Ardavanis A, Bassilopoulos P, Trangas T et al. Overexpression of matrix-metalloproteinase-9 in human breast cancer: a potential favourable indicator in node-negative patients. *Br J Cancer* 2001; 84(11):1488-1496.

(57) Available from: URL:http://www.cancerresearchuk.org/

(58) Mendoza M, Khanna C. Revisiting the seed and soil in cancer metastasis. *Int J Biochem Cell Biol* 2009; 41(7):1452-1462.

(59) Saad S, Gottlieb DJ, Bradstock KF, Overall CM, Bendall LJ. Cancer cell-associated fibronectin induces release of matrix metalloproteinase-2 from normal fibroblasts. *Cancer Res* 2002; 62(1):283-289.

(60) Mendes O, Kim HT, Stoica G. Expression of MMP2, MMP9 and MMP3 in breast cancer brain metastasis in a rat model. *Clin Exp Metastasis* 2005; 22(3):237-246.

(61) Fokas E, Engenhart-Cabillic R, Daniilidis K, Rose F, An HX. Metastasis: the seed and soil theory gains identity. *Cancer Metastasis Rev* 2007; 26(3-4):705-715.

(62) Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat Rev Cancer* 2003; 3(6):453-458.

(63) Knight WA, Livingston RB, Gregory EJ, McGuire WL. Estrogen receptor as an independent prognostic factor for early recurrence in breast cancer. *Cancer Res* 1977; 37(12):4669-4671.

(64) Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res* 2007; 9(1):R6.

(65) Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. Molecular portraits of human breast tumours. *Nature* 2000; 406(6797):747-752.

(66) Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America 98(19):10869-74,* 2001.

(67) Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003; 100(14):8418-8423.

(68) van ', V, Dai H, van d, V, He YD, Hart AA, Mao M et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415(6871):530-536.

(69) van de Vijver MJ, He YD, Van't Veer LJ, Dai H, Hart AA, Voskuil DW et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347(25):1999-2009.

(70) Perou CM, Sorlie T, Eisen MB, van de RM, Jeffrey SS, Rees CA et al. Molecular portraits of human breast tumours. *Nature 406(6797):747-52,* 2000.

(71) Martin KJ, Kritzman BM, Price LM, Koh B, Kwan CP, Zhang X et al. Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res* 2000; 60(8):2232-2238.

(72) Iwao K, Matoba R, Ueno N, Ando A, Miyoshi Y, Matsubara K et al. Molecular classification of primary breast tumors possessing distinct prognostic properties. *Hum Mol Genet* 2002; 11(2):199-206.

(73) Ahr A, Karn T, Solbach C, Seiter T, Strebhardt K, Holtrich U et al. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* 2002; 359(9301):131-132.

(74) Ahr A, Holtrich U, Solbach C, Scharl A, Strebhardt K, Karn T et al. Molecular classification of breast cancer patients by gene expression profiling. *J Pathol* 2001; 195(3):312-320.

(75) Condeelis J, Pollard JW. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell* 2006; 124(2):263-266.

(76) Park CC, Bissell MJ, Barcellos-Hoff MH. The influence of the microenvironment on the malignant phenotype. *Mol Med Today* 2000; 6(8):324-329.

(77) Liotta LA, Kohn EC. The microenvironment of the tumour-host interface. *Nature* 2001; 411(6835):375-379.

(78) Gregoire M, Lieubeau B. The role of fibroblasts in tumor behavior. *Cancer Metastasis Rev* 1995; 14(4):339-350.

(79) Hagemann T, Wilson J, Kulbe H, Li NF, Leinster DA, Charles K et al. Macrophages induce invasiveness of epithelial cancer cells via NF-kappa B and JNK. *J Immunol* 2005; 175(2):1197-1205.

(80) Hagemann T, Robinson SC, Schulz M, Trumper L, Balkwill FR, Binder C. Enhanced invasiveness of breast cancer cell lines upon co-cultivation with macrophages is due to TNF-alpha dependent up-regulation of matrix metalloproteases. *Carcinogenesis* 2004; 25(8):1543-1549.

(81) Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet* 2001; 357(9255):539-545.

(82) Silberstein GB. Tumour-stromal interactions. Role of the stroma in mammary development. *Breast Cancer Res* 2001; 3(4):218-223.

(83) Siveen KS, Kuttan G. Role of macrophages in tumour progression. *Immunol Lett* 2009; 123(2):97-102.

(84) Bhowmick NA, Neilson EG, Moses HL. Stromal fibroblasts in cancer initiation and progression. *Nature* 2004; 432(7015):332-337.

(85) Beachy PA, Karhadkar SS, Berman DM. Tissue repair and stem cell renewal in carcinogenesis. *Nature* 2004; 432(7015):324-331.

(86) Dvorak HF. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N Engl J Med* 1986; 315(26):1650-1659.

(87) Coussens LM, Werb Z. Inflammation and cancer. *Nature* 2002; 420(6917):860-867.

(88) Brown LF, Guidi AJ, Schnitt SJ, Van De Water L, Iruela-Arispe ML, Yeo TK et al. Vascular stroma formation in carcinoma in situ, invasive carcinoma, and metastatic carcinoma of the breast. *Clin Cancer Res* 1999; 5(5):1041-1056.

(89) Queen MM, Ryan RE, Holzer RG, Keller-Peck CR, Jorcyk CL. Breast cancer cells stimulate neutrophils to produce oncostatin M: potential implications for tumor progression. *Cancer Res* 2005; 65(19):8896-8904.

(90) Wong SY, Hynes RO. Tumor-lymphatic interactions in an activated stromal microenvironment. *J Cell Biochem* 2007; 101(4):840-850.

(91) Gonzalez LO, Pidal I, Junquera S, Corte MD, Vazquez J, Rodriguez JC et al. Overexpression of matrix metalloproteinases and their inhibitors in mononuclear inflammatory cells in breast cancer correlates with metastasis-relapse. *Br J Cancer* 2007; 97(7):957-963.

(92) Bingle L, Brown NJ, Lewis CE. The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies. *J Pathol* 2002; 196(3):254-265.

(93) Deryugina EI, Quigley JP. Matrix metalloproteinases and tumor metastasis. *Cancer Metastasis Rev* 2006; 25(1):9-34.

(94) Sappino AP, Skalli O, Jackson B, Schurch W, Gabbiani G. Smooth-muscle differentiation in stromal cells of malignant and non-malignant breast tissues. *Int J Cancer* 1988; 41(5):707-712.

(95) Orimo A, Weinberg RA. Stromal fibroblasts in cancer: a novel tumor-promoting cell type. *Cell Cycle* 2006; 5(15):1597-1601.

(96) Sloane BF, Yan S, Podgorski I, Linebaugh BE, Cher ML, Mai J et al. Cathepsin B and tumor proteolysis: contribution of the tumor microenvironment. *Semin Cancer Biol* 2005; 15(2):149-157.

(97) Kalluri R, Zeisberg M. Fibroblasts in cancer. *Nat Rev Cancer* 2006; 6(5):392-401.

(98) Ryan MC, Orr DJ, Horgan K. Fibroblast stimulation of breast cancer cell growth in a serum-free system. *Br J Cancer* 1993; 67(6):1268-1273.

(99) Singer CF, Kronsteiner N, Marton E, Kubista M, Cullen KJ, Hirtenlehner K et al. MMP-2 and MMP-9 expression in breast cancer-derived human fibroblasts is differentially regulated by stromal-epithelial interactions. *Breast Cancer Res Treat* 2002; 72(1):69-77.

(100) Camps JL, Chang SM, Hsu TC, Freeman MR, Hong SJ, Zhau HE et al. Fibroblast-mediated acceleration of human epithelial tumor growth in vivo. *Proc Natl Acad Sci U S A* 1990; 87(1):75-79.

(101) Yashiro M, Ikeda K, Tendo M, Ishikawa T, Hirakawa K. Effect of organ-specific fibroblasts on proliferation and differentiation of breast cancer cells. *Breast Cancer Res Treat* 2005; 90(3):307-313.

(102) Anderson IC, Mari SE, Broderick RJ, Mari BP, Shipp MA. The angiogenic factor interleukin 8 is induced in non-small cell lung cancer/pulmonary fibroblast cocultures. *Cancer Res* 2000; 60(2):269-272.

(103) Lewis CE, Pollard JW. Distinct role of macrophages in different tumor microenvironments. *Cancer Res* 2006; 66(2):605-612.

(104) Chen JJ, Lin YC, Yao PL, Yuan A, Chen HY, Shun CT et al. Tumor-associated macrophages: the double-edged sword in cancer progression. *J Clin Oncol* 2005; 23(5):953-964.

(105) Leek RD, Lewis CE, Whitehouse R, Greenall M, Clarke J, Harris AL. Association of macrophage infiltration with angiogenesis and prognosis in invasive breast carcinoma. *Cancer Res* 1996; 56(20):4625-4629.

(106) van Ravenswaay Claasen HH, Kluin PM, Fleuren GJ. Tumor infiltrating cells in human cancer. On the possible role of CD16+ macrophages in antitumor cytotoxicity. *Lab Invest* 1992; 67(2):166-174.

(107) Nath D, Hartnell A, Happerfield L, Miles DW, Burchell J, Taylor-Papadimitriou J et al. Macrophage-tumour cell interactions: identification of MUC1 on breast cancer cells as a potential counter-receptor for the macrophage-restricted receptor, sialoadhesin. *Immunology* 1999; 98(2):213-219.

(108) O'Sullivan C, Lewis CE. Tumour-associated leucocytes: friends or foes in breast carcinoma. *J Pathol* 1994; 172(3):229-235.

(109) Volodko N, Reiner A, Rudas M, Jakesz R. Tumour-associated macrophages in breast cancer and their prognostic correlations. *The Breast* 1998; 7(2):99-105.

(110) Lin EY, Nguyen AV, Russell RG, Pollard JW. Colony-stimulating factor 1 promotes progression of mammary tumors to malignancy. *J Exp Med* 2001; 193(6):727-740.

(111) Pollard JW. Tumour-educated macrophages promote tumour progression and metastasis. *Nat Rev Cancer* 2004; 4(1):71-78.

(112) Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends Immunol* 2002; 23(11):549-555.

(113) Stephens TC, Currie GA, Peacock JH. Repopulation of gamma-irradiated Lewis lung carcinoma by malignant cells and host macrophage progenitors. *Br J Cancer* 1978; 38(5):573-582.

(114) McBride WH. Phenotype and functions of intratumoral macrophages. *Biochim Biophys Acta* 1986; 865(1):27-41.

(115) Johnson JP, Riethmuller G, Schirrmacher V. Tumor immunology: Paul Ehrlich's heritage. *Immunol Today* 1989; 10(8):S35-S37.

(116) Lin EY, Pollard JW. Role of infiltrated leucocytes in tumour growth and spread. *Br J Cancer* 2004; 90(11):2053-2058.

(117) Adams TE, Alpert S, Hanahan D. Non-tolerance and autoantibodies to a transgenic self antigen expressed in pancreatic beta cells. *Nature* 1987; 325(6101):223-228.

(118) Daniel D, Chiu C, Giraudo E, Inoue M, Mizzen LA, Chu NR et al. CD4+ T cell-mediated antigen-specific immunotherapy in a mouse model of cervical cancer. *Cancer Res* 2005; 65(5):2018-2025.

(119) Balkwill F, Charles KA, Mantovani A. Smoldering and polarized inflammation in the initiation and promotion of malignant disease. *Cancer Cell* 2005; 7(3):211-217.

(120) Stetler-Stevenson WG, Aznavoorian S, Liotta LA. Tumor cell interactions with the extracellular matrix during invasion and metastasis. *Annu Rev Cell Biol* 1993; 9:541-573.

(121) Sternlicht MD, Lochter A, Sympson CJ, Huey B, Rougier JP, Gray JW et al. The stromal proteinase MMP3/stromelysin-1 promotes mammary carcinogenesis. *Cell* 1999; 98(2):137-146.

(122) Lewis CE, Leek R, Harris A, McGee JO. Cytokine regulation of angiogenesis in breast cancer: the role of tumor-associated macrophages. *J Leukoc Biol* 1995; 57(5):747-751.

(123) O'Sullivan C, Lewis CE, Harris AL, McGee JO. Secretion of epidermal growth factor by macrophages associated with breast carcinoma. *Lancet* 1993; 342(8864):148-149.

(124) Bingle L, Lewis CE, Corke KP, Reed MW, Brown NJ. Macrophages promote angiogenesis in human breast tumour spheroids in vivo. *Br J Cancer* 2006; 94(1):101-107.

(125) Mantovani A, Bottazzi B, Colotta F, Sozzani S, Ruco L. The origin and function of tumor-associated macrophages. *Immunol Today* 1992; 13(7):265-270.

(126) Lin EY, Gouon-Evans V, Nguyen AV, Pollard JW. The macrophage growth factor CSF-1 in mammary gland development and tumor progression. *J Mammary Gland Biol Neoplasia* 2002; 7(2):147-162.

(127) Leek RD, Landers RJ, Harris AL, Lewis CE. Necrosis correlates with high vascular density and focal macrophage infiltration in invasive carcinoma of the breast. *Br J Cancer* 1999; 79(5-6):991-995.

(128) Lee AH, Happerfield LC, Bobrow LG, Millis RR. Angiogenesis and inflammation in ductal carcinoma in situ of the breast. *J Pathol* 1997; 181(2):200-206.

(129) Goede V, Brogelli L, Ziche M, Augustin HG. Induction of inflammatory angiogenesis by monocyte chemoattractant protein-1. *Int J Cancer* 1999; 82(5):765-770.

(130) Tsutsui S, Yasuda K, Suzuki K, Tahara K, Higashi H, Era S. Macrophage infiltration and its prognostic implications in breast cancer: the relationship with VEGF expression and microvessel density. *Oncol Rep* 2005; 14(2):425-431.

(131) Lee AH, Happerfield LC, Bobrow LG, Millis RR. Angiogenesis and inflammation in invasive carcinoma of the breast. *J Clin Pathol* 1997; 50(8):669-673.

(132) Ohno S, Ohno Y, Suzuki N, Kamei T, Koike K, Inagawa H et al. Correlation of histological localization of tumor-associated macrophages with clinicopathological features in endometrial cancer. *Anticancer Res* 2004; 24(5C):3335-3342.

(133) Martin MD, Carter KJ, Jean-Philippe SR, Chang M, Mobashery S, Thiolloy S et al. Effect of Ablation or Inhibition of Stromal Matrix Metalloproteinase-9 on Lung Metastasis in a Breast Cancer Model Is Dependent on Genetic Background. *Cancer Res* 2008; 68(15):6251-6259.

(134) van Kempen LC, Coussens LM. MMP9 potentiates pulmonary metastasis formation. *Cancer Cell* 2002; 2(4):251-252.

(135) Belotti D, Calcagno C, Garofalo A, Caronia D, Riccardi E, Giavazzi R et al. Vascular endothelial growth factor stimulates organ-specific host matrix metalloproteinase-9 expression and ovarian cancer invasion. *Mol Cancer Res* 2008; 6(4):525-534.

(136) Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000; 100(1):57-70.

(137) Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998; 396(6712):643-649.

(138) Wernert N. The multiple roles of tumour stroma. *Virchows Arch* 1997; 430(6):433-443.

(139) Fidler IJ, Kripke ML. Metastasis results from preexisting variant cells within a malignant tumor. *Science* 1977; 197(4306):893-895.

(140) Hart IR, Fidler IJ. Role of organ selectivity in the determination of metastatic patterns of B16 melanoma. *Cancer Res* 1980; 40(7):2281-2287.

(141) Paget S. The Distribution of Secondary Growths in Cancer of The Breast. The Lancet 133[3421], 571-573. 23-3-1889.

(142) Ribatti D, Mangialardi G, Vacca A. Stephen Paget and the 'seed and soil' theory of metastatic dissemination. *Clin Exp Med* 2006; 6(4):145-149.

(143) Steeg PS, Theodorescu D. Metastasis: a therapeutic target for cancer. *Nat Clin Pract Oncol* 2008; 5(4):206-219.

(144) Aplin AE, Howe AK, Juliano RL. Cell adhesion molecules, signal transduction and cell growth. *Curr Opin Cell Biol* 1999; 11(6):737-744.

(145) Hynes RO, Lively JC, McCarty JH, Taverna D, Francis SE, Hodivala-Dilke K et al. The diverse roles of integrins and their ligands in angiogenesis. *Cold Spring Harb Symp Quant Biol* 2002; 67:143-153.

(146) Nikolopoulos SN, Blaikie P, Yoshioka T, Guo W, Giancotti FG. Integrin beta4 signaling promotes tumor angiogenesis. *Cancer Cell* 2004; 6(5):471-483.

(147) Naylor MJ, Li N, Cheung J, Lowe ET, Lambert E, Marlow R et al. Ablation of beta1 integrin in mammary epithelium reveals a key role for integrin in glandular morphogenesis and differentiation. *J Cell Biol* 2005; 171(4):717-728.

(148) Reddig PJ, Juliano RL. Clinging to life: cell to matrix adhesion and cell survival. *Cancer Metastasis Rev* 2005; 24(3):425-439.

(149) Juliano RL, Varner JA. Adhesion molecules in cancer: the role of integrins. *Curr Opin Cell Biol* 1993; 5(5):812-818.

(150) Kurschat P, Mauch C. Mechanisms of metastasis. *Clin Exp Dermatol* 2000; 25(6):482-489.

(151) Kaplan RN, Psaila B, Lyden D. Bone marrow cells in the 'pre-metastatic niche': within bone and beyond. *Cancer Metastasis Rev* 2006; 25(4):521-529.

(152) Hunter K, Welch DR, Liu ET. Genetic background is an important determinant of metastatic potential. *Nat Genet* 2003; 34(1):23-24.

(153) Bowen RL, Duffy SW, Ryan DA, Hart IR, Jones JL. Early onset of breast cancer in a group of British black women. *Br J Cancer* 2008; 98(2):277-281.

(154) Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K et al. Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study. *JAMA* 2006; 295(21):2492-2502.

(155) Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 2006; 66(21):10292-10301.

(156) Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics 81(3):559-75,* 2007.

(157) Vizoso FJ, Gonzalez LO, Corte MD, Rodriguez JC, Vazquez J, Lamelas ML et al. Study of matrix metalloproteinases and their inhibitors in breast cancer. *Br J Cancer* 2007; 96(6):903-911.

(158) Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* 2006; 34(Database issue):D590-D598.

(159) Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT et al. The UCSC Genome Browser Database. *Nucleic Acids Res* 2003; 31(1):51-54.

(160) Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A et al. The UCSC genome browser database: update 2007. *Nucleic Acids Res* 2007; 35(Database issue):D668-D673.

(161) A haplotype map of the human genome. *Nature* 2005; 437(7063):1299-1320.

(162) Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. *Genome Res* 2005; 15(11):1592-1593.

(163) Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; 449(7164):851-861.

(164) Nantermet PV, Xu J, Yu Y, Hodor P, Holder D, Adamski S et al. Identification of genetic pathways activated by the androgen receptor during the induction of proliferation in the ventral prostate gland. *J Biol Chem* 2004; 279(2):1310-1322.

(165) Lippman M, Bolan G, Huff K. The effects of androgens and antiandrogens on hormone-responsive human breast cancer in long-term tissue culture. *Cancer Res* 1976; 36(12):4610-4618.

(166) Kuenen-Boumeester V, Van der Kwast TH, Claassen CC, Look MP, Liem GS, Klijn JG et al. The clinical significance of androgen receptors in breast cancer and their relation to histological and cell biological parameters. *Eur J Cancer* 1996; 32A(9):1560-1565.

(167) Lillie EO, Bernstein L, Ursin G. The role of androgens and polymorphisms in the androgen receptor in the epidemiology of breast cancer. *Breast Cancer Res* 2003; 5(3):164-173.

(168) Hall RE, Aspinall JO, Horsfall DJ, Birrell SN, Bentel JM, Sutherland RL et al. Expression of the androgen receptor and an androgen-responsive protein, apolipoprotein D, in human breast cancer. *Br J Cancer* 1996; 74(8):1175-1180.

(169) Lea OA, Kvinnsland S, Thorsen T. Improved Measurement of Androgen Receptors in Human Breast Cancer. *Cancer Res* 1989; 49(24_Part_1):7162-7167.

(170) Nelson AR, Fingleton B, Rothenberg ML, Matrisian LM. Matrix metalloproteinases: biologic activity and clinical implications. *J Clin Oncol* 2000; 18(5):1135-1149.

(171) Agoff SN, Swanson PE, Linden H, Hawes SE, Lawton TJ. Androgen receptor expression in estrogen receptor-negative breast cancer. Immunohistochemical, clinical, and prognostic associations. *Am J Clin Pathol* 2003; 120(5):725-731.

(172) Schippinger W, Regitnig P, Dandachi N, Wernecke KD, Bauernhofer T, Samonigg H et al. Evaluation of the prognostic significance of androgen receptor expression in metastatic breast cancer. *Virchows Arch* 2006; 449(1):24-30.

(173) Gonzalez LO, Corte MD, Vazquez J, Junquera S, Sanchez R, Alvarez AC et al. Androgen receptor expresion in breast cancer: relationship with clinicopathological characteristics of the tumors, prognosis, and expression of metalloproteases and their inhibitors. *BMC Cancer* 2008; 8:149.

(174) Taplin ME, Bubley GJ, Shuster TD, Frantz ME, Spooner AE, Ogata GK et al. Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *N Engl J Med* 1995; 332(21):1393-1398.

(175) Rebbeck TR, Kantoff PW, Krithivas K, Neuhausen S, Blackwood MA, Godwin AK et al. Modification of BRCA1-associated breast cancer risk by the polymorphic androgen-receptor CAG repeat. *Am J Hum Genet* 1999; 64(5):1371-1377.

(176) Hopp TA, Weiss HL, Hilsenbeck SG, Cui Y, Allred DC, Horwitz KB et al. Breast cancer patients with progesterone receptor PR-A-rich tumors have poorer disease-free survival rates. *Clin Cancer Res* 2004; 10(8):2751-2760.

(177) Leygue E, Dotzlaw H, Watson PH, Murphy LC. Identification of novel exon-deleted progesterone receptor variant mRNAs in human breast tissue. *Biochem Biophys Res Commun* 1996; 228(1):63-68.

(178) Yeates C, Hunt SM, Balleine RL, Clarke CL. Characterization of a truncated progesterone receptor protein in breast tumors. *J Clin Endocrinol Metab* 1998; 83(2):460-467.

(179) De V, I, Huggins GS, Hankinson SE, Lescault PJ, Boezen M, Colditz GA et al. A functional polymorphism in the promoter of the progesterone receptor gene associated with endometrial cancer risk. *Proc Natl Acad Sci U S A* 2002; 99(19):12263-12268.

(180) Andersen TI, Heimdal KR, Skrede M, Tveit K, Berg K, Borresen AL. Oestrogen receptor (ESR) polymorphisms and breast cancer susceptibility. *Hum Genet* 1994; 94(6):665-670.

(181) Anderson TI, Wooster R, Laake K, Collins N, Warren W, Skrede M et al. Screening for ESR mutations in breast and ovarian cancer patients. *Hum Mutat* 1997; 9(6):531-536.

(182) Iwase H, Greenman JM, Barnes DM, Hodgson S, Bobrow L, Mathew CG. Sequence variants of the estrogen receptor (ER) gene found in breast cancer patients with ER negative and progesterone receptor positive tumors. *Cancer Lett* 1996; 108(2):179-184.

(183) Lemieux P, Fuqua S. The role of the estrogen receptor in tumor progression. *J Steroid Biochem Mol Biol* 1996; 56(1-6 Spec No):87-91.

(184) Roodi N, Bailey LR, Kao WY, Verrier CS, Yee CJ, Dupont WD et al. Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. *J Natl Cancer Inst* 1995; 87(6):446-451.

(185) Schubert EL, Lee MK, Newman B, King MC. Single nucleotide polymorphisms (SNPs) in the estrogen receptor gene and breast cancer susceptibility. *J Steroid Biochem Mol Biol* 1999; 71(1-2):21-27.

(186) Hill SM, Fuqua SA, Chamness GC, Greene GL, McGuire WL. Estrogen receptor expression in human breast cancer associated with an estrogen receptor gene restriction fragment length polymorphism. *Cancer Res* 1989; 49(1):145-148.

(187) Yaich L, Dupont WD, Cavener DR, Parl FF. Analysis of the PvuII restriction fragment-length polymorphism and exon structure of the estrogen receptor gene in breast cancer and peripheral blood. *Cancer Res* 1992; 52(1):77-83.

(188) Boyapati SM, Shu XO, Ruan ZX, Cai Q, Smith JR, Wen W et al. Polymorphisms in ER-alpha gene interact with estrogen receptor status in breast cancer survival. *Clin Cancer Res* 2005; 11(3):1093-1098.

(189) Rapuri PB, Gallagher JC, Knezetic JA, Haynatzka V. Estrogen receptor alpha gene polymorphisms are associated with changes in bone remodeling markers and treatment response to estrogen. *Maturitas* 2006; 53(4):371-379.

(190) van Duijnhoven FJ, Bezemer ID, Peeters PH, Roest M, Uitterlinden AG, Grobbee DE et al. Polymorphisms in the estrogen receptor alpha gene and mammographic density. *Cancer Epidemiol Biomarkers Prev* 2005; 14(11 Pt 1):2655-2660.

(191) van Duijnhoven FJ, Peeters PH, Warren RM, Bingham SA, Uitterlinden AG, van Noord PA et al. Influence of estrogen receptor alpha and progesterone receptor polymorphisms on the effects of hormone therapy on mammographic density. *Cancer Epidemiol Biomarkers Prev* 2006; 15(3):462-467.

(192) Hoch RV, Thompson DA, Baker RJ, Weigel RJ. GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int J Cancer* 1999; 84(2):122-128.

(193) Parikh P, Palazzo JP, Rose LJ, Daskalakis C, Weigel RJ. GATA-3 expression as a predictor of hormone response in breast cancer. *J Am Coll Surg* 2005; 200(5):705-710.

(194) Mehra R, Varambally S, Ding L, Shen R, Sabel MS, Ghosh D et al. Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* 2005; 65(24):11259-11264.

(195) van de RM, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol* 2002; 161(6):1991-1996.

(196) Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X et al. Mutation of GATA3 in human breast tumors. *Oncogene* 2004; 23(46):7669-7678.

(197) James JJ, Evans AJ, Pinder SE, Gutteridge E, Cheung KL, Chan S et al. Bone metastases from breast carcinoma: histopathological - radiological correlations and prognostic features. *Br J Cancer* 2003; 89(4):660-665.

(198) Smid M, Wang Y, Klijn JG, Sieuwerts AM, Zhang Y, Atkins D et al. Genes associated with breast cancer metastatic to bone. *J Clin Oncol* 2006; 24(15):2261-2267.

(199) Huijts PE, Vreeswijk MP, Kroeze-Jansema KH, Jacobi CE, Seynaeve C, Krol-Warmerdam EM et al. Clinical correlates of low-risk variants in FGFR2, TNRC9, MAP3K1, LSP1 and 8q24 in a Dutch cohort of incident breast cancer cases. *Breast Cancer Res* 2007; 9(6):R78.

(200) Maestro R, Dei Tos AP, Hamamori Y, Krasnokutsky S, Sartorelli V, Kedes L et al. Twist is a potential oncogene that inhibits apoptosis. *Genes Dev* 1999; 13(17):2207-2217.

(201) Sahlin P, Windh P, Lauritzen C, Emanuelsson M, Gronberg H, Stenman G. Women with Saethre-Chotzen syndrome are at increased risk of breast cancer. *Genes Chromosomes Cancer* 2007; 46(7):656-660.

(202) Martin TA, Goyal A, Watkins G, Jiang WG. Expression of the transcription factors snail, slug, and twist and their clinical significance in human breast cancer. *Ann Surg Oncol* 2005; 12(6):488-496.

(203) Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C et al. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* 2004; 117(7):927-939.

(204) Mironchik Y, Winnard PT, Jr., Vesuna F, Kato Y, Wildes F, Pathak AP et al. Twist overexpression induces in vivo angiogenesis and correlates with chromosomal instability in breast cancer. *Cancer Res* 2005; 65(23):10801-10809.

(205) Gibson S, Widmann C, Johnson GL. Differential involvement of MEK kinase 1 (MEKK1) in the induction of apoptosis in response to microtubule-targeted drugs versus DNA damaging agents. *J Biol Chem* 1999; 274(16):10916-10922.

(206) Widmann C, Johnson NL, Gardner AM, Smith RJ, Johnson GL. Potentiation of apoptosis by low dose stress stimuli in cells expressing activated MEK kinase 1. *Oncogene* 1997; 15(20):2439-2447.

(207) Widmann C, Gerwins P, Johnson NL, Jarpe MB, Johnson GL. MEK kinase 1, a substrate for DEVD-directed caspases, is involved in genotoxin-induced apoptosis. *Mol Cell Biol* 1998; 18(4):2416-2429.

(208) Deng M, Chen WL, Takatori A, Peng Z, Zhang L, Mongan M et al. A role for the mitogen-activated protein kinase kinase kinase 1 in epithelial wound healing. *Mol Biol Cell* 2006; 17(8):3446-3455.

(209) Grose R, Dickson C. Fibroblast growth factor signaling in tumorigenesis. *Cytokine Growth Factor Rev* 2005; 16(2):179-186.

(210) Adnane J, Gaudray P, Dionne CA, Crumley G, Jaye M, Schlessinger J et al. BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene* 1991; 6(4):659-663.

(211) Rebbeck TR, DeMichele A, Tran TV, Panossian S, Bunin GR, Troxel AB et al. Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. *Carcinogenesis* 2009; 30(2):269-274.

(212) van d, V, Peterse JL, Mooi WJ, Wisman P, Lomans J, Dalesio O et al. Neu-protein overexpression in breast cancer. Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer. *N Engl J Med* 1988; 319(19):1239-1245.

(213) Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989; 244(4905):707-712.

(214) Xie D, Shu XO, Deng Z, Wen WQ, Creek KE, Dai Q et al. Population-based, case-control study of HER2 genetic polymorphism and breast cancer risk. *J Natl Cancer Inst* 2000; 92(5):412-417.

(215) Ameyaw MM, Tayeb M, Thornton N, Folayan G, Tariq M, Mobarek A et al. Ethnic variation in the HER-2 codon 655 genetic polymorphism previously associated with breast cancer. *J Hum Genet* 2002; 47(4):172-175.

(216) Folkman J. Angiogenesis-dependent diseases. *Semin Oncol* 2001; 28(6):536-542.

(217) Weinstat-Saslow D, Steeg PS. Angiogenesis and colonization in the tumor metastatic process: basic and applied advances. *FASEB J* 1994; 8(6):401-407.

(218) Schneider BP, Skaar TC, Sledge GW, Badve S, Li L, Flockhart DA. Analysis of angiogenesis genes from paraffin-embedded breast tumor and lymph nodes. *Breast Cancer Res Treat* 2006; 96(3):209-215.

(219) Jacobs EJ, Feigelson HS, Bain EB, Brady KA, Rodriguez C, Stevens VL et al. Polymorphisms in the vascular endothelial growth factor gene and breast cancer in the Cancer Prevention Study II cohort. *Breast Cancer Res* 2006; 8(2):R22.

(220) Guinebretiere JM, Le MG, Gavoille A, Bahi J, Contesso G. Angiogenesis and risk of breast cancer in women with fibrocystic disease. *J Natl Cancer Inst* 1994; 86(8):635-636.

(221) Guidi AJ, Fischer L, Harris JR, Schnitt SJ. Microvessel density and distribution in ductal carcinoma in situ of the breast. *J Natl Cancer Inst* 1994; 86(8):614-619.

(222) Guidi AJ, Schnitt SJ, Fischer L, Tognazzi K, Harris JR, Dvorak HF et al. Vascular permeability factor (vascular endothelial growth factor) expression and angiogenesis in patients with ductal carcinoma in situ of the breast. *Cancer* 1997; 80(10):1945-1953.

(223) Weidner N, Semple JP, Welch WR, Folkman J. Tumor angiogenesis and metastasis--correlation in invasive breast carcinoma. *N Engl J Med* 1991; 324(1):1-8.

(224) Krippl P, Langsenlehner U, Renner W, Yazdani-Biuki B, Wolf G, Wascher TC et al. A common 936 C/T gene polymorphism of vascular endothelial growth factor is associated with decreased breast cancer risk. *Int J Cancer* 2003; 106(4):468-471.

(225) Smith KC, Bateman AC, Fussell HM, Howell WM. Cytokine gene polymorphisms and breast cancer susceptibility and prognosis. *Eur J Immunogenet* 2004; 31(4):167-173.

(226) Jin Q, Hemminki K, Enquist K, Lenner P, Grzybowska E, Klaes R et al. Vascular endothelial growth factor polymorphisms in relation to breast cancer development and prognosis. *Clin Cancer Res* 2005; 11(10):3647-3653.

(227) Langsenlehner U, Wolf G, Langsenlehner T, Gerger A, Hofmann G, Clar H et al. Genetic polymorphisms in the vascular endothelial growth factor gene and breast cancer risk. The Austrian "tumor of breast tissue: incidence, genetics, and environmental risk factors" study. *Breast Cancer Res Treat* 2008; 109(2):297-304.

(228) Schneider BP, Radovich M, Sledge GW, Robarge JD, Li L, Storniolo AM et al. Association of polymorphisms of angiogenesis genes with breast cancer. *Breast Cancer Res Treat* 2008; 111(1):157-163.

(229) Lu H, Shu XO, Cui Y, Kataoka N, Wen W, Cai Q et al. Association of genetic polymorphisms in the VEGF gene with breast cancer survival. *Cancer Res* 2005; 65(12):5015-5019.

(230) Kataoka N, Cai Q, Wen W, Shu XO, Jin F, Gao YT et al. Population-based case-control study of VEGF gene polymorphisms and breast cancer risk among Chinese women. *Cancer Epidemiol Biomarkers Prev* 2006; 15(6):1148-1152.

(231) Pacher M, Seewald MJ, Mikula M, Oehler S, Mogg M, Vinatzer U et al. Impact of constitutive IGF1/IGF2 stimulation on the transcriptional program of human breast cancer cells. *Carcinogenesis* 2007; 28(1):49-59.

(232) Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. *Lancet* 1998; 351(9113):1393-1396.

(233) Schernhammer ES, Holly JM, Pollak MN, Hankinson SE. Circulating levels of insulin-like growth factors, their binding proteins, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2005; 14(3):699-704.

(234) Schernhammer ES, Holly JM, Hunter DJ, Pollak MN, Hankinson SE. Insulin-like growth factor-I, its binding proteins (IGFBP-1 and IGFBP-3), and growth hormone and breast cancer risk in The Nurses Health Study II. *Endocr Relat Cancer* 2006; 13(2):583-592.

(235) Peyrat JP, Bonneterre J, Hecquet B, Vennin P, Louchez MM, Fournier C et al. Plasma insulin-like growth factor-1 (IGF-1) concentrations in human breast cancer. *Eur J Cancer* 1993; 29A(4):492-497.

(236) Tamimi R, Cox D, Kraft P, Pollak M, Haiman C, Cheng I et al. Common genetic variation in IGF1, IGFBP-1, and IGFBP-3 in relation to mammographic density: a cross-sectional study. *Breast Cancer Research* 2007; 9(1):R18.

(237) Canzian F, McKay JD, Cleveland RJ, Dossus L, Biessy C, Rinaldi S et al. Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study. *Br J Cancer* 2006; 94(2):299-307.

(238) Al-Zahrani A, Sandhu MS, Luben RN, Thompson D, Baynes C, Pooley KA et al. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. *Hum Mol Genet* 2006; 15(1):1-10.

(239) Badache A, Hynes NE. Interleukin 6 inhibits proliferation and, in cooperation with an epidermal growth factor receptor autocrine loop, increases migration of T47D breast cancer cells. *Cancer Res* 2001; 61(1):383-391.

(240) Robinson EK, Sneige N, Grimm EA. Correlation of interleukin 6 with interleukin 1alpha in human mammary tumours, but not with oestrogen receptor expression. *Cytokine* 1998; 10(12):970-976.

(241) Bozcuk H, Uslu G, Samur M, Yildiz M, Ozben T, Ozdogan M et al. Tumour necrosis factor-alpha, interleukin-6, and fasting serum insulin correlate with clinical outcome in metastatic breast cancer patients treated with chemotherapy. *Cytokine* 2004; 27(2-3):58-65.

(242) Bachelot T, Ray-Coquard I, Menetrier-Caux C, Rastkha M, Duc A, Blay JY. Prognostic value of serum levels of interleukin 6 and of serum and plasma levels of vascular endothelial growth factor in hormone-refractory metastatic breast cancer patients. *Br J Cancer* 2003; 88(11):1721-1726.

(243) Knupfer H, Preiss R. Significance of interleukin-6 (IL-6) in breast cancer (review). *Breast Cancer Res Treat* 2007; 102(2):129-135.

(244) Zhang GJ, Adachi I. Serum interleukin-6 levels correlate to tumor progression and prognosis in metastatic breast carcinoma. *Anticancer Res* 1999; 19(2B):1427-1432.

(245) Bieche I, Lerebours F, Tozlu S, Espie M, Marty M, Lidereau R. Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature. *Clin Cancer Res* 2004; 10(20):6789-6795.

(246) Asgeirsson KS, Olafsdottir K, Jonasson JG, Ogmundsdottir HM. The effects of IL-6 on cell adhesion and e-cadherin expression in breast cancer. *Cytokine* 1998; 10(9):720-728.

(247) Basolo F, Fiore L, Fontanini G, Conaldi PG, Calvo S, Falcone V et al. Expression of and response to interleukin 6 (IL6) in human mammary tumors. *Cancer Res* 1996; 56(13):3118-3122.

(248) Fontanini G, Campani D, Roncella M, Cecchetti D, Calvo S, Toniolo A et al. Expression of interleukin 6 (IL-6) correlates with oestrogen receptor in human breast carcinoma. *Br J Cancer* 1999; 80(3-4):579-584.

(249) Karczewska A, Nawrocki S, Breborowicz D, Filas V, Mackiewicz A. Expression of interleukin-6, interleukin-6 receptor, and glycoprotein 130 correlates with good prognoses for patients with breast carcinoma. *Cancer* 2000; 88(9):2061-2071.

(250) Slattery ML, Curtin K, Baumgartner R, Sweeney C, Byers T, Giuliano AR et al. IL6, aspirin, nonsteroidal anti-inflammatory drugs, and breast cancer risk in women living in the southwestern United States. *Cancer Epidemiol Biomarkers Prev* 2007; 16(4):747-755.

(251) Vickers MA, Green FR, Terry C, Mayosi BM, Julier C, Lathrop M et al. Genotype at a promoter polymorphism of the interleukin-6 gene is associated with baseline levels of plasma C-reactive protein. *Cardiovasc Res* 2002; 53(4):1029-1034.

(252) Jones SA, Horiuchi S, Topley N, Yamamoto N, Fuller GM. The soluble interleukin 6 receptor: mechanisms of production and implications in disease. *FASEB J* 2001; 15(1):43-58.

(253) Hefler LA, Grimm C, Lantzsch T, Lampe D, Leodolter S, Koelbl H et al. Interleukin-1 and interleukin-6 gene polymorphisms and the risk of breast cancer in caucasian women. *Clin Cancer Res* 2005; 11(16):5718-5721.

(254) Iacopetta B, Grieu F, Joseph D. The -174 G/C gene polymorphism in interleukin-6 is associated with an aggressive breast cancer phenotype. *Br J Cancer* 2004; 90(2):419-422.

(255) Hefler LA, Grimm C, Ackermann S, Malur S, Radjabi-Rahat AR, Leodolter S et al. An interleukin-6 gene promoter polymorphism influences the biological phenotype of ovarian cancer. *Cancer Res* 2003; 63(12):3066-3068.

(256) DeMichele A, Martin AM, Mick R, Gor P, Wray L, Klein-Cabral M et al. Interleukin-6 -174G-->C polymorphism is associated with improved outcome in high-risk breast cancer. *Cancer Res* 2003; 63(22):8051-8056.

(257) Balasubramanian SP, Azmy IA, Higham SE, Wilson AG, Cross SS, Cox A et al. Interleukin gene polymorphisms and breast cancer: a case control study and systematic literature review. *BMC Cancer* 2006; 6:188.

(258) Slattery ML, Curtin K, Giuliano AR, Sweeney C, Baumgartner R, Edwards S et al. Active and passive smoking, IL6, ESR1, and breast cancer risk. *Breast Cancer Res Treat* 2008; 109(1):101-111.

(259) Derynck R, Akhurst RJ, Balmain A. TGF-beta signaling in tumor suppression and cancer progression. *Nat Genet* 2001; 29(2):117-129.

(260) Grainger DJ, Heathcote K, Chiano M, Snieder H, Kemp PR, Metcalfe JC et al. Genetic control of the circulating concentration of transforming growth factor type beta1. *Hum Mol Genet* 1999; 8(1):93-97.

(261) Siegel PM, Shu W, Cardiff RD, Muller WJ, Massague J. Transforming growth factor beta signaling impairs Neu-induced mammary tumorigenesis while promoting pulmonary metastasis. *Proc Natl Acad Sci U S A* 2003; 100(14):8430-8435.

(262) Dunning AM, Ellis PD, McBride S, Kirschenlohr HL, Healey CS, Kemp PR et al. A transforming growth factorbeta1 signal peptide variant increases secretion in vitro and is associated with increased incidence of invasive breast cancer. *Cancer Res* 2003; 63(10):2610-2615.

(263) Malik ST, Naylor MS, East N, Oliff A, Balkwill FR. Cells secreting tumour necrosis factor show enhanced metastasis in nude mice. *Eur J Cancer* 1990; 26(10):1031-1034.

(264) Leek RD, Landers R, Fox SB, Ng F, Harris AL, Lewis CE. Association of tumour necrosis factor alpha and its receptors with thymidine phosphorylase expression in invasive breast carcinoma. *Br J Cancer* 1998; 77(12):2246-2251.

(265) Ruuls SR, Sedgwick JD. Unlinking tumor necrosis factor biology from the major histocompatibility complex: lessons from human genetics and animal models. *Am J Hum Genet* 1999; 65(2):294-301.

(266) Levine AJ, Momand J, Finlay CA. The p53 tumour suppressor gene. *Nature* 1991; 351(6326):453-456.

(267) Hollstein M, Sidransky D, Vogelstein B, Harris CC. p53 mutations in human cancers. *Science* 1991; 253(5015):49-53.

(268) Ho GH, Calvano JE, Bisogna M, Abouezzi Z, Borgen PI, Cordon-Cardo C et al. Genetic alterations of the p14ARF -hdm2-p53 regulatory pathway in breast carcinoma. *Breast Cancer Res Treat* 2001; 65(3):225-232.

(269) Pharoah PD, Day NE, Caldas C. Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. *Br J Cancer* 1999; 80(12):1968-1973.

(270) Dunning AM, Healey CS, Pharoah PD, Teare MD, Ponder BA, Easton DF. A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 1999; 8(10):843-854.

(271) Tommiska J, Eerola H, Heinonen M, Salonen L, Kaare M, Tallila J et al. Breast cancer patients with p53 Pro72 homozygous genotype have a poorer survival. *Clin Cancer Res* 2005; 11(14):5098-5103.

(272) Toyama T, Zhang Z, Nishio M, Hamaguchi M, Kondo N, Iwase H et al. Association of TP53 codon 72 polymorphism and the outcome of adjuvant therapy in breast cancer patients. *Breast Cancer Res* 2007; 9(3):R34.

(273) Sjalander A, Birgander R, Hallmans G, Cajander S, Lenner P, Athlin L et al. p53 polymorphisms and haplotypes in breast cancer. *Carcinogenesis* 1996; 17(6):1313-1316.

(274) Zhang Z, Zhang R. p53-independent activities of MDM2 and their relevance to cancer therapy. *Curr Cancer Drug Targets* 2005; 5(1):9-20.

(275) Lundgren K, Montes de Oca LR, McNeill YB, Emerick EP, Spencer B, Barfield CR et al. Targeted expression of MDM2 uncouples S phase from mitosis and inhibits mammary gland development independent of p53. *Genes Dev* 1997; 11(6):714-725.

(276) Alt JR, Bouska A, Fernandez MR, Cerny RL, Xiao H, Eischen CM. Mdm2 binds to Nbs1 at sites of DNA damage and regulates double strand break repair. *J Biol Chem* 2005; 280(19):18771-18781.

(277) McCann AH, Kirley A, Carney DN, Corbally N, Magee HM, Keating G et al. Amplification of the MDM2 gene in human breast cancer and its association with MDM2 and p53 protein status. *Br J Cancer* 1995; 71(5):981-985.

(278) Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC et al. A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell* 2004; 119(5):591-602.

(279) Bond GL, Hu W, Levine A. A single nucleotide polymorphism in the MDM2 gene: from a molecular and cellular explanation to clinical effect. *Cancer Res* 2005; 65(13):5481-5484.

(280) Boersma BJ, Howe TM, Goodman JE, Yfantis HG, Lee DH, Chanock SJ et al. Association of breast cancer outcome with status of p53 and MDM2 SNP309. *J Natl Cancer Inst* 2006; 98(13):911-919.

(281) Millikan RC, Heard K, Winkel S, Hill EJ, Heard K, Massa B et al. No association between the MDM2 -309 T/G promoter polymorphism and breast cancer in African-Americans or Whites. *Cancer Epidemiol Biomarkers Prev* 2006; 15(1):175-177.

(282) Ma H, Hu Z, Zhai X, Wang S, Wang X, Qin J et al. Polymorphisms in the MDM2 promoter and risk of breast cancer: a case-control analysis in a Chinese population. *Cancer Lett* 2006; 240(2):261-267.

(283) Wilkening S, Bermejo JL, Burwinkel B, Klaes R, Bartram CR, Meindl A et al. The single nucleotide polymorphism IVS1+309 in mouse double minute 2 does not affect risk of familial breast cancer. *Cancer Res* 2006; 66(2):646-648.

(284) Swift M, Morrell D, Massey RB, Chase CL. Incidence of cancer in 161 families affected by ataxia-telangiectasia. *N Engl J Med* 1991; 325(26):1831-1836.

(285) Swift M, Reitnauer PJ, Morrell D, Chase CL. Breast and other cancers in families with ataxia-telangiectasia. *N Engl J Med* 1987; 316(21):1289-1294.

(286) Easton D, Ford D, Peto J. Inherited susceptibility to breast cancer. *Cancer Surv* 1993; 18:95-113.

(287) Khanna KK, Chenevix-Trench G. ATM and genome maintenance: defining its role in breast cancer susceptibility. *J Mammary Gland Biol Neoplasia* 2004; 9(3):247-262.

(288) Khanna KK. Cancer risk and the ATM gene: a continuing debate. *J Natl Cancer Inst* 2000; 92(10):795-802.

(289) Olsen JH, Hahnemann JM, Borresen-Dale AL, Brondum-Nielsen K, Hammarstrom L, Kleinerman R et al. Cancer in patients with ataxia-telangiectasia and in their relatives in the nordic countries. *J Natl Cancer Inst* 2001; 93(2):121-127.

(290) Einarsdottir K, Rosenberg LU, Humphreys K, Bonnard C, Palmgren J, Li Y et al. Comprehensive analysis of the ATM, CHEK2 and ERBB2 genes in relation to breast tumour characteristics and survival: a population-based case-control and follow-up study. *Breast Cancer Res* 2006; 8(6):R67.

(291) Stredrick DL, Garcia-Closas M, Pineda MA, Bhatti P, Alexander BH, Doody MM et al. The ATM missense mutation p.Ser49Cys (c.146C>G) and the risk of breast cancer. *Hum Mutat* 2006; 27(6):538-544.

(292) Thorstenson YR, Roxas A, Kroiss R, Jenkins MA, Yu KM, Bachrich T et al. Contributions of ATM mutations to familial breast and ovarian cancer. *Cancer Res* 2003; 63(12):3325-3333.

(293) Heikkinen K, Rapakko K, Karppinen SM, Erkko H, Nieminen P, Winqvist R. Association of common ATM polymorphism with bilateral breast cancer. *Int J Cancer* 2005; 116(1):69-72.

(294) Angele S, Romestaing P, Moullan N, Vuillaume M, Chapot B, Friesen M et al. ATM haplotypes and cellular response to DNA damage: association with breast cancer risk and clinical radiosensitivity. *Cancer Res* 2003; 63(24):8717-8725.

(295) Tommiska J, Jansen L, Kilpivaara O, Edvardsen H, Kristensen V, Tamminen A et al. ATM variants and cancer risk in breast cancer patients from Southern Finland. *BMC Cancer* 2006; 6:209.

(296) Xu X, Weaver Z, Linke SP, Li C, Gotay J, Wang XW et al. Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells. *Mol Cell* 1999; 3(3):389-395.

(297) Moynahan ME, Chiu JW, Koller BH, Jasin M. Brca1 controls homology-directed DNA repair. *Mol Cell* 1999; 4(4):511-518.

(298) Thompson ME, Jensen RA, Obermiller PS, Page DL, Holt JT. Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression. *Nat Genet* 1995; 9(4):444-450.

(299) Friedman LS, Ostermeyer EA, Szabo CI, Dowd P, Lynch ED, Rowell SE et al. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nat Genet* 1994; 8(4):399-404.

(300) Castilla LH, Couch FJ, Erdos MR, Hoskins KF, Calzone K, Garber JE et al. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat Genet* 1994; 8(4):387-391.

(301) Barker DF, Almeida ER, Casey G, Fain PR, Liao SY, Masunaka I et al. BRCA1 R841W: a strong candidate for a common mutation with moderate phenotype. *Genet Epidemiol* 1996; 13(6):595-604.

(302) Druesne-Pecollo N, Tehard B, Mallet Y, Gerber M, Norat T, Hercberg S et al. Alcohol and genetic polymorphisms: effect on risk of alcohol-related cancer. *Lancet Oncol* 2009; 10(2):173-180.

(303) Dumitrescu RG, Shields PG. The etiology of alcohol-induced breast cancer. *Alcohol* 2005; 35(3):213-225.

(304) Newbold RR, Liehr JG. Induction of uterine adenocarcinoma in CD-1 mice by catechol estrogens. *Cancer Res* 2000; 60(2):235-237.

(305) Han X, Liehr JG. DNA single-strand breaks in kidneys of Syrian hamsters treated with steroidal estrogens: hormone-induced free radical damage preceding renal malignancy. *Carcinogenesis* 1994; 15(5):997-1000.

(306) De V, I, Hankinson SE, Li L, Colditz GA, Hunter DJ. Association of CYP1B1 polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002; 11(5):489-492.

(307) Liehr JG, Ricci MJ. 4-Hydroxylation of estrogens as marker of human mammary tumors. *Proc Natl Acad Sci U S A* 1996; 93(8):3294-3296.

(308) Shimada T, Watanabe J, Kawajiri K, Sutter TR, Guengerich FP, Gillam EM et al. Catalytic properties of polymorphic human cytochrome P450 1B1 variants. *Carcinogenesis* 1999; 20(8):1607-1613.

(309) Zheng W, Xie DW, Jin F, Cheng JR, Dai Q, Wen WQ et al. Genetic polymorphism of cytochrome P450-1B1 and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2000; 9(2):147-150.

(310) Hanna IH, Dawling S, Roodi N, Guengerich FP, Parl FF. Cytochrome P450 1B1 (CYP1B1) pharmacogenetics: association of polymorphisms with functional differences in estrogen hydroxylation activity. *Cancer Res* 2000; 60(13):3440-3444.

(311) Bailey LR, Roodi N, Dupont WD, Parl FF. Association of cytochrome P450 1B1 (CYP1B1) polymorphism with steroid receptor status in breast cancer. *Cancer Res* 1998; 58(22):5038-5041.

(312) Kocabas NA, Sardas S, Cholerton S, Daly AK, Karakaya AE. Cytochrome P450 CYP1B1 and catechol O-methyltransferase (COMT) genetic polymorphisms and breast cancer susceptibility in a Turkish population. *Arch Toxicol* 2002; 76(11):643-649.

(313) Saintot M, Malaveille C, Hautefeuille A, Gerber M. Interactions between genetic polymorphism of cytochrome P450-1B1, sulfotransferase 1A1, catechol-o-methyltransferase and tobacco exposure in breast cancer risk. *Int J Cancer* 2003; 107(4):652-657.

(314) Sillanpaa P, Heikinheimo L, Kataja V, Eskelinen M, Kosma VM, Uusitupa M et al. CYP1A1 and CYP1B1 genetic polymorphisms, smoking and breast cancer risk in a Finnish Caucasian population. *Breast Cancer Res Treat* 2007; 104(3):287-297.

(315) Lee KM, Abel J, Ko Y, Harth V, Park WY, Seo JS et al. Genetic polymorphisms of cytochrome P450 19 and 1B1, alcohol use, and breast cancer risk in Korean women. *Br J Cancer* 2003; 88(5):675-678.

(316) Rylander-Rudqvist T, Wedren S, Granath F, Humphreys K, Ahlberg S, Weiderpass E et al. Cytochrome P450 1B1 gene polymorphisms and postmenopausal breast cancer risk. *Carcinogenesis* 2003; 24(9):1533-1539.

(317) Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr. et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br J Cancer* 2002; 87(11):1234-1245.

(318) Bowlin SJ, Leske MC, Varma A, Nasca P, Weinstein A, Caplan L. Breast cancer risk and alcohol consumption: results from a large case-control study. *Int J Epidemiol* 1997; 26(5):915-923.

(319) Petri AL, Tjonneland A, Gamborg M, Johansen D, Hoidrup S, Sorensen TI et al. Alcohol intake, type of beverage, and risk of breast cancer in pre- and postmenopausal women. *Alcohol Clin Exp Res* 2004; 28(7):1084-1090.

(320) Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. *JAMA* 2001; 286(17):2143-2151.

(321) Lieber CS. Role of oxidative stress and antioxidant therapy in alcoholic and nonalcoholic liver diseases. *Adv Pharmacol* 1997; 38:601-628.

(322) Purohit V, Khalsa J, Serrano J. Mechanisms of alcohol-associated cancers: introduction and summary of the symposium. *Alcohol* 2005; 35(3):155-160.

(323) Triano EA, Slusher LB, Atkins TA, Beneski JT, Gestl SA, Zolfaghari R et al. Class I alcohol dehydrogenase is highly expressed in normal human mammary epithelium but not in invasive breast cancer: implications for breast carcinogenesis. *Cancer Res* 2003; 63(12):3092-3100.

(324) Lilla C, Koehler T, Kropp S, Wang-Gohrke S, Chang-Claude J. Alcohol dehydrogenase 1B (ADH1B) genotype, alcohol consumption and breast cancer risk by age 50 years in a German case-control study. *Br J Cancer* 2005; 92(11):2039-2041.

(325) Inbal B, Cohen O, Polak-Charcon S, Kopolovic J, Vadai E, Eisenbach L et al. DAP kinase links the control of apoptosis to metastasis. *Nature* 1997; 390(6656):180-184.

(326) Raveh T, Droguett G, Horwitz MS, DePinho RA, Kimchi A. DAP kinase activates a p19ARF/p53-mediated apoptotic checkpoint to suppress oncogenic transformation. *Nat Cell Biol* 2001; 3(1):1-7.

(327) Bandyopadhyay S, Zhan R, Wang Y, Pai SK, Hirota S, Hosobe S et al. Mechanism of apoptosis induced by the inhibition of fatty acid synthase in breast cancer cells. *Cancer Res* 2006; 66(11):5934-5940.

(328) Levy D, Plu-Bureau, Decroix Y, Hugol D, Rostene W, Kimchi A et al. Death-associated protein kinase loss of expression is a new marker for breast cancer prognosis. *Clin Cancer Res* 2004; 10(9):3124-3130.

(329) Shohat G, Spivak-Kroizman T, Cohen O, Bialik S, Shani G, Berrisi H et al. The pro-apoptotic function of death-associated protein kinase is controlled by a unique inhibitory autophosphorylation-based mechanism. *J Biol Chem* 2001; 276(50):47460-47467.

(330) Jin Y, Blue EK, Gallagher PJ. Control of death-associated protein kinase (DAPK) activity by phosphorylation and proteasomal degradation. *J Biol Chem* 2006; 281(51):39033-39040.

(331) Gonzalez-Gomez P, Bello MJ, Alonso ME, Lomas J, Arjona D, Aminoso C et al. Frequent death-associated protein-kinase promoter hypermethylation in brain metastases of solid tumors. *Oncol Rep* 2003; 10(4):1031-1033.

(332) Lehmann U, Celikkaya G, Hasemeier B, Langer F, Kreipe H. Promoter hypermethylation of the death-associated protein kinase gene in breast cancer is associated with the invasive lobular subtype. *Cancer Res* 2002; 62(22):6634-6638.

(333) Dulaimi E, Hillinck J, Ibanez dC, I, Al-Saleem T, Cairns P. Tumor suppressor gene promoter hypermethylation in serum of breast cancer patients. *Clin Cancer Res* 2004; 10(18 Pt 1):6189-6193.

(334) Kuo JC, Lin JR, Staddon JM, Hosoya H, Chen RH. Uncoordinated regulation of stress fibers and focal adhesions by DAP kinase. *J Cell Sci* 2003; 116(Pt 23):4777-4790.

(335) Raval A, Tanner SM, Byrd JC, Angerman EB, Perko JD, Chen SS et al. Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. *Cell* 2007; 129(5):879-890.

(336) Yang S, Liu J, Thor AD, Yang X. Caspase expression profile and functional activity in a panel of breast cancer cell lines. *Oncol Rep* 2007; 17(5):1229-1235.

(337) MacPherson G, Healey CS, Teare MD, Balasubramanian SP, Reed MW, Pharoah PD et al. Association of a common variant of the CASP8 gene with reduced risk of breast cancer. *J Natl Cancer Inst* 2004; 96(24):1866-1869.

(338) Frank B, Hemminki K, Wappenschmidt B, Meindl A, Klaes R, Schmutzler RK et al. Association of the CASP10 V410I variant with reduced familial breast cancer risk and interaction with the CASP8 D302H variant. *Carcinogenesis* 2006; 27(3):606-609.

(339) Liotta LA, Stetler-Stevenson WG. Tumor invasion and metastasis: an imbalance of positive and negative regulation. *Cancer Res* 1991; 51(18 Suppl):5054s-5059s.

(340) Helleman J, Jansen MP, Berns EM. Gene expression profiling of treatment resistance: hype or hope for therapeutic target identification. *Int J Gynecol Cancer* 2006; 16 Suppl 2:538-542.

(341) Ioachim E, Charchanti A, Briasoulis E, Karavasilis V, Tsanou H, Arvanitis DL et al. Immunohistochemical expression of extracellular matrix components tenascin, fibronectin, collagen type IV and laminin in breast cancer: their prognostic value and role in tumour invasion and progression. *Eur J Cancer* 2002; 38(18):2362-2370.

(342) Takei H, Iino Y, Horiguchi J, Yokoe T. Immunohistochemical fibronectin staining pattern and prognosis in invasive breast carcinoma. *Oncology* 1995; 52(2):106-111.

(343) Spentzos D, Levine DA, Ramoni MF, Joseph M, Gu X, Boyd J et al. Gene expression signature with independent prognostic significance in epithelial ovarian cancer. *J Clin Oncol* 2004; 22(23):4700-4710.

(344) DeNardo DG, Kim HT, Hilsenbeck S, Cuba V, Tsimelzon A, Brown PH. Global gene expression analysis of estrogen receptor transcription factor cross talk in breast cancer: identification of estrogen-induced/activator protein-1-dependent genes. *Mol Endocrinol* 2005; 19(2):362-378.

(345) Beckers J, Herrmann F, Rieger S, Drobyshev AL, Horsch M, Hrabe de AM et al. Identification and validation of novel ERBB2 (HER2, NEU) targets including genes involved in angiogenesis. *Int J Cancer* 2005; 114(4):590-597.

(346) Vidal F, Aberdam D, Miquel C, Christiano AM, Pulkkinen L, Uitto J et al. Integrin beta 4 mutations associated with junctional epidermolysis bullosa with pyloric atresia. *Nat Genet* 1995; 10(2):229-234.

(347) Shaw LM, Rabinovitz I, Wang HH, Toker A, Mercurio AM. Activation of phosphoinositide 3-OH kinase by the alpha6beta4 integrin promotes carcinoma invasion. *Cell* 1997; 91(7):949-960.

(348) Dajee M, Lazarov M, Zhang JY, Cai T, Green CL, Russell AJ et al. NF-kappaB blockade and oncogenic Ras trigger invasive human epidermal neoplasia. *Nature* 2003; 421(6923):639-643.

(349) Murgia C, Blaikie P, Kim N, Dans M, Petrie HT, Giancotti FG. Cell cycle and adhesion defects in mice carrying a targeted deletion of the integrin beta4 cytoplasmic domain. *EMBO J* 1998; 17(14):3940-3951.

(350) Guo W, Pylayeva Y, Pepe A, Yoshioka T, Muller WJ, Inghirami G et al. Beta 4 integrin amplifies ErbB2 signaling to promote mammary tumorigenesis. *Cell* 2006; 126(3):489-502.

(351) Chambers AF, Matrisian LM. Changing views of the role of matrix metalloproteinases in metastasis. *J Natl Cancer Inst* 1997; 89(17):1260-1270.

(352) Duffy MJ, McCarthy K. Matrix metalloproteinases in cancer: prognostic markers and targets for therapy (review). *Int J Oncol* 1998; 12(6):1343-1348.

(353) Mason J, Yancy H, Lashley K, Jett M, Day A. Comparative study of matrix metalloproteinase expression between African American and Caucasian Women. *Journal of Carcinogenesis* 2004; 3(1):15.

(354) Liabakk NB, Talbot I, Smith RA, Wilkinson K, Balkwill F. Matrix metalloprotease 2 (MMP-2) and matrix metalloprotease 9 (MMP-9) type IV collagenases in colorectal cancer. *Cancer Res* 1996; 56(1):190-196.

(355) Westermarck J, Kahari VM. Regulation of matrix metalloproteinase expression in tumor invasion. *FASEB J* 1999; 13(8):781-792.

(356) Murphy G, Knauper V, Atkinson S, Butler G, English W, Hutton M et al. Matrix metalloproteinases in arthritic disease. *Arthritis Res* 2002; 4(Suppl 3):S39-S49.

(357) Duffy MJ, Maguire TM, Hill A, McDermott E, O'Higgins N. Metalloproteinases: role in breast carcinogenesis, invasion and metastasis. *Breast Cancer Res* 2000; 2(4):252-257.

(358) Rudolph-Owen LA, Chan R, Muller WJ, Matrisian LM. The matrix metalloproteinase matrilysin influences early-stage mammary tumorigenesis. *Cancer Res* 1998; 58(23):5500-5506.

(359) Davies B, Miles DW, Happerfield LC, Naylor MS, Bobrow LG, Rubens RD et al. Activity of type IV collagenases in benign and malignant breast disease. *Br J Cancer* 1993; 67(5):1126-1131.

(360) Lochter A, Srebrow A, Sympson CJ, Terracio N, Werb Z, Bissell MJ. Misregulation of stromelysin-1 expression in mouse mammary tumor cells accompanies acquisition of stromelysin-1-dependent invasive properties. *J Biol Chem* 1997; 272(8):5007-5015.

(361) Sledge GW, Jr., Qulali M, Goulet R, Bone EA, Fife R. Effect of matrix metalloproteinase inhibitor batimastat on breast cancer regrowth and metastasis in athymic mice. *J Natl Cancer Inst* 1995; 87(20):1546-1550.

(362) Wang M, Liu YE, Greene J, Sheng S, Fuchs A, Rosen EM et al. Inhibition of tumor growth and metastasis of human breast cancer cells transfected with tissue inhibitor of metalloproteinase 4. *Oncogene* 1997; 14(23):2767-2774.

(363) Fingleton B. Matrix metalloproteinase inhibitors for cancer therapy:the current situation and future prospects. *Expert Opin Ther Targets* 2003; 7(3):385-397.

(364) Jones JL, Glynn P, Walker RA. Expression of MMP-2 and MMP-9, their inhibitors, and the activator MT1-MMP in primary breast carcinomas. *J Pathol* 1999; 189(2):161-168.

(365) Grieu F, Li WQ, Iacopetta B. Genetic polymorphisms in the MMP-2 and MMP-9 genes and breast cancer phenotype. *Breast Cancer Res Treat* 2004; 88(3):197-204.

(366) Nakopoulou L, Tsirmpa I, Alexandrou P, Louvrou A, Ampela C, Markaki S et al. MMP-2 protein in invasive breast cancer and the impact of MMP-2/TIMP-2 phenotype on overall survival. *Breast Cancer Res Treat* 2003; 77(2):145-155.

(367) Brummer O, Athar S, Riethdorf L, Loning T, Herbst H. Matrix-metalloproteinases 1, 2, and 3 and their tissue inhibitors 1 and 2 in benign and malignant breast lesions: an in situ hybridization study. *Virchows Arch* 1999; 435(6):566-573.

(368) Iwata H, Kobayashi S, Iwase H, Masaoka A, Fujimoto N, Okada Y. Production of matrix metalloproteinases and tissue inhibitors of metalloproteinases in human breast carcinomas. *Jpn J Cancer Res* 1996; 87(6):602-611.

(369) Jinga DC, Blidaru A, Condrea I, Ardeleanu C, Dragomir C, Szegli G et al. MMP-9 and MMP-2 gelatinases and TIMP-1 and TIMP-2 inhibitors in breast cancer: correlations with prognostic factors. *J Cell Mol Med* 2006; 10(2):499-510.

(370) Hirvonen R, Talvensaari-Mattila A, Paakko P, Turpeenniemi-Hujanen T. Matrix metalloproteinase-2 (MMP-2) in T(1-2)N0 breast carcinoma. *Breast Cancer Res Treat* 2003; 77(1):85-91.

(371) Talvensaari-Mattila A, Paakko P, Turpeenniemi-Hujanen T. MMP-2 positivity and age less than 40 years increases the risk for recurrence in premenopausal patients with node-positive breast carcinoma. *Breast Cancer Res Treat* 1999; 58(3):287-293.

(372) Talvensaari-Mattila A, Paakko P, Blanco-Sequeiros G, Turpeenniemi-Hujanen T. Matrix metalloproteinase-2 (MMP-2) is associated with the risk for a relapse in postmenopausal patients with node-positive breast carcinoma treated with antiestrogen adjuvant therapy. *Breast Cancer Res Treat* 2001; 65(1):55-61.

(373) Leppa S, Saarto T, Vehmanen L, Blomqvist C, Elomaa I. A high serum matrix metalloproteinase-2 level is associated with an adverse prognosis in node-positive breast carcinoma. *Clin Cancer Res* 2004; 10(3):1057-1063.

(374) Talvensaari-Mattila A, Paakko P, Hoyhtya M, Blanco-Sequeiros G, Turpeenniemi-Hujanen T. Matrix metalloproteinase-2 immunoreactive protein: a marker of aggressiveness in breast carcinoma. *Cancer* 1998; 83(6):1153-1162.

(375) Talvensaari-Mattila A, Paakko P, Turpeenniemi-Hujanen T. Matrix metalloproteinase-2 (MMP-2) is associated with survival in breast carcinoma. *Br J Cancer* 2003; 89(7):1270-1275.

(376) Johnsen M, Lund LR, Romer J, Almholt K, Dano K. Cancer invasion and tissue remodeling: common themes in proteolytic matrix degradation. *Curr Opin Cell Biol* 1998; 10(5):667-671.

(377) Li HC, Cao DC, Liu Y, Hou YF, Wu J, Lu JS et al. Prognostic value of matrix metalloproteinases (MMP-2 and MMP-9) in patients with lymph node-negative breast carcinoma. *Breast Cancer Res Treat* 2004; 88(1):75-85.

(378) Saleem M, Kweon MH, Johnson JJ, Adhami VM, Elcheva I, Khan N et al. S100A4 accelerates tumorigenesis and invasion of human prostate cancer through the transcriptional regulation of matrix metalloproteinase 9. *Proc Natl Acad Sci U S A* 2006; 103(40):14825-14830.

(379) Ranuncolo SM, Armanasco E, Cresta C, Bal De Kier JE, Puricelli L. Plasma MMP-9 (92 kDa-MMP) activity is useful in the follow-up and in the assessment of prognosis in breast cancer patients. *Int J Cancer* 2003; 106(5):745-751.

(380) Wang FQ, So J, Reierstad S, Fishman DA. Matrilysin (MMP-7) promotes invasion of ovarian cancer cells by activation of progelatinase. *Int J Cancer* 2005; 114(1):19-31.

(381) Kitoh T, Yanai H, Saitoh Y, Nakamura Y, Matsubara Y, Kitoh H et al. Increased expression of matrix metalloproteinase-7 in invasive early gastric cancer. *J Gastroenterol* 2004; 39(5):434-440.

(382) Ajisaka H, Yonemura Y, Miwa K. Correlation of lymph node metastases and expression of matrix metalloproteinase-7 in patients with gastric cancer. *Hepatogastroenterology* 2004; 51(57):900-905.

(383) Engelse MA, Hanemaaijer R, Koolwijk P, van H, V. The fibrinolytic system and matrix metalloproteinases in angiogenesis and tumor progression. *Semin Thromb Hemost* 2004; 30(1):71-82.

(384) Mylona E, Kapranou A, Mavrommatis J, Markaki S, Keramopoulos A, Nakopoulou L. The multifunctional role of the immunohistochemical expression of MMP-7 in invasive breast cancer. *APMIS* 2005; 113(4):246-255.

(385) Price SJ, Greaves DR, Watkins H. Identification of novel, functional genetic variants in the human matrix metalloproteinase-2 gene: role of Sp1 in allele-specific transcriptional regulation. *J Biol Chem* 2001; 276(10):7549-7558.

(386) Yu C, Pan K, Xing D, Liang G, Tan W, Zhang L et al. Correlation between a single nucleotide polymorphism in the matrix metalloproteinase-2 promoter and risk of lung cancer. *Cancer Res* 2002; 62(22):6430-6433.

(387) Miao X, Yu C, Tan W, Xiong P, Liang G, Lu W et al. A functional polymorphism in the matrix metalloproteinase-2 gene promoter (-1306C/T) is associated with risk of development but not metastasis of gastric cardia adenocarcinoma. *Cancer Res* 2003; 63(14):3987-3990.

(388) Yu C, Zhou Y, Miao X, Xiong P, Tan W, Lin D. Functional haplotypes in the promoter of matrix metalloproteinase-2 predict risk of the occurrence and metastasis of esophageal cancer. *Cancer Res* 2004; 64(20):7622-7628.

(389) Xu E, Lai M, Lv B, Xing X, Huang Q, Xia X. A single nucleotide polymorphism in the matrix metalloproteinase-2 promoter is associated with colorectal cancer. *Biochem Biophys Res Commun* 2004; 324(3):999-1003.

(390) Charoenrat P, Khantapura P. The role of genetic polymorphisms in the promoters of the matrix metalloproteinase-2 and tissue inhibitor of metalloproteinase-2 genes in head and neck cancer. *Oral Oncol* 2006; 42(3):257-267.

(391) Zhou Y, Yu C, Miao X, Tan W, Liang G, Xiong P et al. Substantial reduction in risk of breast cancer associated with genetic polymorphisms in the promoters of the matrix metalloproteinase-2 and tissue inhibitor of metalloproteinase-2 genes. *Carcinogenesis* 2004; 25(3):399-404.

(392) Lei H, Hemminki K, Altieri A, Johansson R, Enquist K, Hallmans G et al. Promoter polymorphisms in matrix metalloproteinases and their inhibitors: few associations

with breast cancer susceptibility and progression. *Breast Cancer Res Treat* 2007; 103(1):61-69.

(393) Harendza S, Lovett DH, Panzer U, Lukacs Z, Kuhnl P, Stahl RA. Linked common polymorphisms in the gelatinase a promoter are associated with diminished transcriptional response to estrogen and genetic fitness. *J Biol Chem* 2003; 278(23):20490-20499.

(394) Zhang B, Ye S, Herrmann SM, Eriksson P, de MM, Evans A et al. Functional polymorphism in the regulatory region of gelatinase B gene in relation to severity of coronary atherosclerosis. *Circulation* 1999; 99(14):1788-1794.

(395) Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 2000; 406(6795):532-535.

(396) Eccles D, Gerty S, Simmonds P, Hammond V, Ennis S, Altman DG et al. Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. *BMC Cancer 7:160,* 2007.

(397) Menshikov M, Elizarova E, Plakida K, Timofeeva A, Khaspekov G, Beabealashvilli R et al. Urokinase upregulates matrix metalloproteinase-9 expression in THP-1 monocytes via gene transcription and protein synthesis. *Biochem J* 2002; 367(Pt 3):833-839.

(398) Ricca A, Biroccio A, Del BD, Mackay AR, Santoni A, Cippitelli M. bcl-2 over-expression enhances NF-kappaB activity and induces mmp-9 transcription in human MCF7(ADR) breast-cancer cells. *Int J Cancer* 2000; 86(2):188-196.

(399) Li CI, Daling JR, Malone KE. Incidence of invasive breast cancer by hormone receptor status from 1992 to 1998. *J Clin Oncol* 2003; 21(1):28-34.

(400) Ilunga K, Nishiura R, Inada H, El-Karef A, Imanaka-Yoshida K, Sakakura T et al. Co-stimulation of human breast cancer cells with transforming growth factor-beta and tenascin-C enhances matrix metalloproteinase-9 expression and cancer cell invasion. *Int J Exp Pathol* 2004; 85(6):373-379.

(401) Mackay AR, Ballin M, Pelina MD, Farina AR, Nason AM, Hartzler JL et al. Effect of phorbol ester and cytokines on matrix metalloproteinase and tissue inhibitor of metalloproteinase expression in tumor and normal cell lines. *Invasion Metastasis* 1992; 12(3-4):168-184.

(402) Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008; 40(6):695-701.

(403) Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; 2(8):e124.

(404) Yoo YJ, Bull SB, Paterson AD, Waggott D, Sun L. Were genome-wide linkage studies a waste of time? Exploiting candidate regions within genome-wide association studies. *Genet Epidemiol* 2010; 34(2):107-118.

(405) Manes S, Llorente M, Lacalle RA, Gomez-Mouton C, Kremer L, Mira E et al. The matrix metalloproteinase-9 regulates the insulin-like growth factor-triggered autocrine response in DU-145 carcinoma cells. *J Biol Chem* 1999; 274(11):6935-6945.

(406) Noe V, Fingleton B, Jacobs K, Crawford HC, Vermeulen S, Steelant W et al. Release of an invasion promoter E-cadherin fragment by matrilysin and stromelysin-1. *J Cell Sci* 2001; 114(Pt 1):111-118.

(407) Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer progression. *Nat Rev Cancer* 2002; 2(3):161-174.

(408) Turk V, Kos J, Turk B. Cysteine cathepsins (proteases)--on the main stage of cancer? *Cancer Cell* 2004; 5(5):409-410.

(409) Sternlicht MD, Werb Z. How matrix metalloproteinases regulate cell behavior. *Annu Rev Cell Dev Biol* 2001; 17:463-516.

(410) Stetler-Stevenson WG. Matrix metalloproteinases in angiogenesis: a moving target for therapeutic intervention. *J Clin Invest* 1999; 103(9):1237-1241.

(411) Cornelius LA, Nehring LC, Harding E, Bolanowski M, Welgus HG, Kobayashi DK et al. Matrix metalloproteinases generate angiostatin: effects on neovascularization. *J Immunol* 1998; 161(12):6845-6852.

(412) Wilson CL, Matrisian LM. Matrilysin: an epithelial matrix metalloproteinase with potentially novel functions. *Int J Biochem Cell Biol* 1996; 28(2):123-136.

(413) Sengupta N, MacDonald TT. The role of matrix metalloproteinases in stromal/epithelial interactions in the gut. *Physiology (Bethesda)* 2007; 22:401-409.

(414) Pellikainen JM, Ropponen KM, Kataja VV, Kellokoski JK, Eskelinen MJ, Kosma VM. Expression of matrix metalloproteinase (MMP)-2 and MMP-9 in breast cancer with a special reference to activator protein-2, HER2, and prognosis. *Clin Cancer Res* 2004; 10(22):7621-7628.

(415) Johansson N, Ahonen M, Kahari VM. Matrix metalloproteinases in tumor invasion. *Cell Mol Life Sci* 2000; 57(1):5-15.

(416) Brew K, Dinakarpandian D, Nagase H. Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochim Biophys Acta* 2000; 1477(1-2):267-283.

(417) Yao J, Xiong S, Klos K, Nguyen N, Grijalva R, Li P et al. Multiple signaling pathways involved in activation of matrix metalloproteinase-9 (MMP-9) by heregulin-beta1 in human breast cancer cells. *Oncogene* 2001; 20(56):8066-8074.

(418) Van den Steen PE, Dubois B, Nelissen I, Rudd PM, Dwek RA, Opdenakker G. Biochemistry and molecular biology of gelatinase B or matrix metalloproteinase-9 (MMP-9). *Crit Rev Biochem Mol Biol* 2002; 37(6):375-536.

(419) Yan C, Boyd DD. Regulation of matrix metalloproteinase gene expression. *J Cell Physiol* 2007; 211(1):19-26.

(420) Nair RR, Boyd DD. Expression cloning of novel regulators of 92 kDa type IV collagenase expression. *Biochem Soc Trans* 2005; 33(Pt 5):1135-1136.

(421) Zhao X, Benveniste EN. Transcriptional activation of human matrix metalloproteinase-9 gene expression by multiple co-activators. *J Mol Biol* 2008; 383(5):945-956.

(422) Liotta LA. Tumor invasion and metastases--role of the extracellular matrix: Rhoads Memorial Award lecture. *Cancer Res* 1986; 46(1):1-7.

(423) Curran S, Murray GI. Matrix metalloproteinases: molecular aspects of their roles in tumour invasion and metastasis. *Eur J Cancer* 2000; 36(13 Spec No):1621-1630.

(424) Heppner KJ, Matrisian LM, Jensen RA, Rodgers WH. Expression of most matrix metalloproteinase family members in breast cancer represents a tumor-induced host response. *Am J Pathol* 1996; 149(1):273-282.

(425) Himelstein BP, Canete-Soler R, Bernhard EJ, Muschel RJ. Induction of fibroblast 92 kDa gelatinase/type IV collagenase expression by direct contact with metastatic tumor cells. *J Cell Sci* 1994; 107 (Pt 2):477-486.

(426) Benaud C, Dickson RB, Thompson EW. Roles of the matrix metalloproteinases in mammary gland development and cancer. *Breast Cancer Res Treat* 1998; 50(2):97-116.

(427) Yan S, Sameni M, Sloane BF. Cathepsin B and human tumor progression. *Biol Chem* 1998; 379(2):113-123.

(428) Scorilas A, Talieri M, Ardavanis A, Courtis N, Dimitriadis E, Yotis J et al. Polyadenylate polymerase enzymatic activity in mammary tumor cytosols: A new independent prognostic marker in primary breast cancer. *Cancer Res* 2000; 60(19):5427-5433.

(429) Yousef GM, Scorilas A, Diamandis EP. Genomic organization, mapping, tissue expression, and hormonal regulation of trypsin-like serine protease (TLSP PRSS20), a new member of the human kallikrein gene family. *Genomics* 2000; 63(1):88-96.

(430) Vracko R. Basal lamina scaffold-anatomy and significance for maintenance of orderly tissue structure. *Am J Pathol* 1974; 77(2):314-346.

(431) Nielsen BS, Timshel S, Kjeldsen L, Sehested M, Pyke C, Borregaard N et al. 92 kDa type IV collagenase (MMP-9) is expressed in neutrophils and macrophages but not in malignant epithelial cells in human colon cancer. *Int J Cancer* 1996; 65(1):57-62.

(432) Liotta LA, Stetler-Stevenson WG. Metalloproteinases and cancer invasion. *Semin Cancer Biol* 1990; 1(2):99-106.

(433) Bernhard EJ, Gruber SB, Muschel RJ. Direct evidence linking expression of matrix metalloproteinase 9 (92-kDa gelatinase/collagenase) to the metastatic phenotype in transformed rat embryo cells. *Proc Natl Acad Sci U S A* 1994; 91(10):4293-4297.

(434) Baker EA, Stephenson TJ, Reed MW, Brown NJ. Expression of proteinases and inhibitors in human breast cancer progression and survival. *Mol Pathol* 2002; 55(5):300-304.

(435) Rha SY, Kim JH, Roh JK, Lee KS, Min JS, Kim BS et al. Sequential production and activation of matrix-metalloproteinase-9 (MMP-9) with breast cancer progression. *Breast Cancer Res Treat* 1997; 43(2):175-181.

(436) Duffy MJ, Blaser J, Duggan C, McDermott E, O'Higgins N, Fennelly JJ et al. Assay of matrix metalloproteases types 8 and 9 by ELISA in human breast cancer. *Br J Cancer* 1995; 71(5):1025-1028.

(437) Pacheco MM, Mourao M, Mantovani EB, Nishimoto IN, Brentani MM. Expression of gelatinases A and B, stromelysin-3 and matrilysin genes in breast carcinomas: clinico-pathological correlations. *Clin Exp Metastasis* 1998; 16(7):577-585.

(438) Rha SY, Yang WI, Kim JH, Roh JK, Min JS, Lee KS et al. Different expression patterns of MMP-2 and MMP-9 in breast cancer. *Oncol Rep* 1998; 5(4):875-879.

(439) Kondapaka SB, Fridman R, Reddy KB. Epidermal growth factor and amphiregulin up-regulate matrix metalloproteinase-9 (MMP-9) in human breast cancer cells. *Int J Cancer* 1997; 70(6):722-726.

(440) Nielsen BS, Sehested M, Kjeldsen L, Borregaard N, Rygaard J, Dano K. Expression of matrix metalloprotease-9 in vascular pericytes in human breast cancer. *Lab Invest* 1997; 77(4):345-355.

(441) Wolf C, Rouyer N, Lutz Y, Adida C, Loriot M, Bellocq JP et al. Stromelysin 3 belongs to a subgroup of proteinases expressed in breast carcinoma fibroblastic cells and possibly implicated in tumor progression. *Proc Natl Acad Sci U S A* 1993; 90(5):1843-1847.

(442) Tryggvason K, Hoyhtya M, Pyke C. Type IV collagenases in invasive tumors. *Breast Cancer Res Treat* 1993; 24(3):209-218.

(443) Mira E, Manes S, Lacalle RA, Marquez G, Martinez A. Insulin-like growth factor I-triggered cell migration and invasion are mediated by matrix metalloproteinase-9. *Endocrinology* 1999; 140(4):1657-1664.

(444) Stuelten CH, DaCosta BS, Arany PR, Karpova TS, Stetler-Stevenson WG, Roberts AB. Breast cancer cells induce stromal fibroblasts to express MMP-9 via secretion of TNF-alpha and TGF-beta. *J Cell Sci* 2005; 118(Pt 10):2143-2153.

(445) Atula S, Grenman R, Syrjanen S. Fibroblasts can modulate the phenotype of malignant epithelial cells in vitro. *Exp Cell Res* 1997; 235(1):180-187.

(446) Wang TN, Albo D, Tuszynski GP. Fibroblasts promote breast cancer cell invasion by upregulating tumor matrix metalloproteinase-9 production. *Surgery* 2002; 132(2):220-225.

(447) Lengyel E, Gum R, Juarez J, Clayman G, Seiki M, Sato H et al. Induction of M(r) 92,000 type IV collagenase expression in a squamous cell carcinoma cell line by fibroblasts. *Cancer Res* 1995; 55(4):963-967.

(448) Hanemaaijer R, Verheijen JH, Maguire TM, Visser H, Toet K, McDermott E et al. Increased gelatinase-A and gelatinase-B activities in malignant vs. benign breast tumors. *Int J Cancer* 2000; 86(2):204-207.

(449) Nakajima M, Welch DR, Wynn DM, Tsuruo T, Nicolson GL. Serum and plasma M(r) 92,000 progelatinase levels correlate with spontaneous metastasis of rat 13762NF mammary adenocarcinoma. *Cancer Res* 1993; 53(23):5802-5807.

(450) Kahari VM, Saarialho-Kere U. Matrix metalloproteinases and their inhibitors in tumour growth and invasion. *Ann Med* 1999; 31(1):34-45.

(451) Shapiro SD. Matrix metalloproteinase degradation of extracellular matrix: biological consequences. *Curr Opin Cell Biol* 1998; 10(5):602-608.

(452) Hujanen ES, Vaisanen A, Zheng A, Tryggvason K, Turpeenniemi-Hujanen T. Modulation of M(r) 72,000 and M(r) 92,000 type-IV collagenase (gelatinase A and B) gene expression by interferons alpha and gamma in human melanoma. *Int J Cancer* 1994; 58(4):582-586.

(453) Rao JS, Steck PA, Mohanam S, Stetler-Stevenson WG, Liotta LA, Sawaya R. Elevated levels of M(r) 92,000 type IV collagenase in human brain tumors. *Cancer Res* 1993; 53(10 Suppl):2208-2211.

(454) Giannelli G, Fransvea E, Marinosci F, Bergamini C, Daniele A, Colucci S et al. Gelatinase levels in male and female breast cancer. *Biochem Biophys Res Commun* 2002; 292(1):161-166.

(455) Davies B, Waxman J, Wasan H, Abel P, Williams G, Krausz T et al. Levels of matrix metalloproteases in bladder cancer correlate with tumor grade and invasion. *Cancer Res* 1993; 53(22):5365-5369.

(456) Alvarez OA, Carmichael DF, DeClerck YA. Inhibition of collagenolytic activity and metastasis of tumor cells by a recombinant human tissue inhibitor of metalloproteinases. *J Natl Cancer Inst* 1990; 82(7):589-595.

(457) McGowan PM, Duffy MJ. Matrix metalloproteinase expression and outcome in patients with breast cancer: analysis of a published database. *Ann Oncol* 2008; 19(9):1566-1572.

(458) Remacle A, McCarthy K, Noel A, Maguire T, McDermott E, O'Higgins N et al. High levels of TIMP-2 correlate with adverse prognosis in breast cancer. *Int J Cancer* 2000; 89(2):118-121.

(459) Visscher DW, Hoyhtya M, Ottosen SK, Liang CM, Sarkar FH, Crissman JD et al. Enhanced expression of tissue inhibitor of metalloproteinase-2 (TIMP-2) in the stroma of breast carcinomas correlates with tumor recurrence. *Int J Cancer* 1994; 59(3):339-344.

(460) Daidone MG, Silvestrini R, D'Errico A, Di FG, Benini E, Mancini AM et al. Laminin receptors, collagenase IV and prognosis in node-negative breast cancers. *Int J Cancer* 1991; 48(4):529-532.

(461) Naylor MS, Stamp GW, Davies BD, Balkwill FR. Expression and activity of MMPS and their regulators in ovarian cancer. *Int J Cancer* 1994; 58(1):50-56.

(462) Garbett EA, Reed MW, Brown NJ. Proteolysis in human breast and colorectal cancer. *Br J Cancer* 1999; 81(2):287-293.

(463) Mizumoto H, Saito T, Ashihara K, Nishimura M, Tanaka R, Kudo R. Acceleration of invasive activity via matrix metalloproteinases by transfection of the estrogen receptor-alpha gene in endometrial carcinoma cells. *Int J Cancer* 2002; 100(4):401-406.

(464) Paquette B, Bisson M, Therriault H, Lemay R, Pare M, Banville P et al. Activation of matrix metalloproteinase-2 and -9 by 2- and 4-hydroxyestradiol. *J Steroid Biochem Mol Biol* 2003; 87(1):65-73.

(465) van den Brule FA, Engel J, Stetler-Stevenson WG, Liu FT, Sobel ME, Castronovo V. Genes involved in tumor invasion and metastasis are differentially modulated by estradiol and progestin in human breast-cancer cells. *Int J Cancer* 1992; 52(4):653-657.

(466) Nilsson UW, Garvin S, Dabrosin C. MMP-2 and MMP-9 activity is regulated by estradiol and tamoxifen in cultured human breast cancer cells. *Breast Cancer Res Treat* 2007; 102(3):253-261.

(467) Mott JD, Werb Z. Regulation of matrix biology by matrix metalloproteinases. *Curr Opin Cell Biol* 2004; 16(5):558-564.

(468) Bergers G, Brekken R, McMahon G, Vu TH, Itoh T, Tamaki K et al. Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nat Cell Biol* 2000; 2(10):737-744.

(469) Sadeghi M, Motovali-bashi M, Hojati Z. MMP-9 promoter polymorphism associated with tumor progression of breast cancer in Iranian population. *International Journal of Integrative Biology* 2009; 6(1):33-37.

(470) Roehe AV, Frazzon AP, Agnes G, Damin AP, Hartman AA, Graudenz MS. Detection of polymorphisms in the promoters of matrix metalloproteinases 2 and 9 genes in breast cancer in South Brazil: preliminary results. *Breast Cancer Research & Treatment 102(1):123-4,* 2007.

(471) Przybylowska K, Kluczna A, Zadrozny M, Krawczyk T, Kulig A, Rykala J et al. Polymorphisms of the promoter regions of matrix metalloproteinases genes MMP-1 and MMP-9 in breast cancer. *Breast Cancer Res Treat* 2006; 95(1):65-72.

(472) Hughes S, Agbaje O, Bowen RL, Holliday DL, Shaw JA, Duffy S et al. Matrix metalloproteinase single-nucleotide polymorphisms and haplotypes predict breast cancer progression. *Clin Cancer Res* 2007; 13(22 Pt 1):6673-6680.

(473) Pinto LA, Depner M, Klopp N, Illig T, Vogelberg C, von ME et al. MMP-9 gene variants increase the risk for non-atopic asthma in children. *Respir Res* 2010; 11:23.

(474) Rajagopalan H, Nowak MA, Vogelstein B, Lengauer C. The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer* 2003; 3(9):695-701.

(475) Mantovani A, Porta C, Rubino L, Allavena P, Sica A. Tumor-associated macrophages (TAMs) as new target in anticancer therapy. *Drug Discovery Today: Therapeutic Strategies* 2006; 3(3):361-366.

(476) Sugiura Y, Shimada H, Seeger RC, Laug WE, DeClerck YA. Matrix metalloproteinases-2 and -9 are expressed in human neuroblastoma: contribution of stromal cells to their production and correlation with metastasis. *Cancer Res* 1998; 58(10):2209-2216.

(477) Segain JP, Harb J, Gregoire M, Meflah K, Menanteau J. Induction of fibroblast gelatinase B expression by direct contact with cell lines derived from primary tumor but not from metastases. *Cancer Res* 1996; 56(23):5506-5512.

(478) Bissell MJ, Radisky DC, Rizki A, Weaver VM, Petersen OW. The organizing principle: microenvironmental influences in the normal and malignant breast. *Differentiation* 2002; 70(9-10):537-546.

(479) Itoh T, Tanioka M, Matsuda H, Nishimoto H, Yoshioka T, Suzuki R et al. Experimental metastasis is suppressed in MMP-9-deficient mice. *Clin Exp Metastasis* 1999; 17(2):177-181.

(480) Brown PD. Matrix metalloproteinase inhibitors: a novel class of anticancer agents. *Adv Enzyme Regul* 1995; 35:293-301.

(481) St-Pierre Y, Couillard J, Van TC. Regulation of MMP-9 gene expression for the development of novel molecular targets against cancer and inflammatory diseases. *Expert Opin Ther Targets* 2004; 8(5):473-489.

(482) Koivunen E, Arap W, Valtanen H, Rainisalo A, Medina OP, Heikkila P et al. Tumor targeting with a selective gelatinase inhibitor. *Nat Biotechnol* 1999; 17(8):768-774.

(483) Higgins MJ, Stearns V. Understanding resistance to tamoxifen in hormone receptor-positive breast cancer. *Clin Chem* 2009; 55(8):1453-1455.

(484) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005; 365(9472):1687-1717.

(485) Umar A, Kang H, Timmermans AM, Look MP, Meijer-van Gelder ME, den Bakker MA et al. Identification of a putative protein profile associated with tamoxifen therapy resistance in breast cancer. *Mol Cell Proteomics* 2009; 8(6):1278-1294.

(486) Pontiggia O. Stromal-Epithelial Interactions and Tamoxifen Resistance. *Medicina (Buenos Aires)* 2007; 67 (Supl. II):60-64.

(487) Stahle-Backdahl M, Parks WC. 92-kd gelatinase is actively expressed by eosinophils and stored by neutrophils in squamous cell carcinoma. *Am J Pathol* 1993; 142(4):995-1000.

(488) Martin MD, Matrisian LM. The other side of MMPs: protective roles in tumor progression. *Cancer Metastasis Rev* 2007; 26(3-4):717-724.

(489) Bendrik C, Robertson J, Gauldie J, Dabrosin C. Gene transfer of matrix metalloproteinase-9 induces tumor regression of breast cancer in vivo. *Cancer Res* 2008; 68(9):3405-3412.

(490) Hamano Y, Zeisberg M, Sugimoto H, Lively JC, Maeshima Y, Yang C et al. Physiological levels of tumstatin, a fragment of collagen IV alpha3 chain, are generated by MMP-9 proteolysis and suppress angiogenesis via alphaV beta3 integrin. *Cancer Cell* 2003; 3(6):589-601.

(491) Tapper W, Hammond V, Gerty S, Ennis S, Simmonds P, Collins A et al. The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. *Breast Cancer Res* 2008; 10(6):R108.