Molecular Epidemiology of Breast Cancer

Harlid, Sophia

2011

Molecular Epidemiology of Breast Cancer

Sophia Harlid

LUNDS UNIVERSITET
Medicinska fakulteten

Institutionen för laboratoriemedicin
Skånes universitetssjukhus
Lunds universitet

AKADEMISK AVHANDLING

som med vederbörligt tillstånd av Medicinska Fakulteten vid Lunds universitet för avläggande av doktorsexamen i medicinsk vetenskap kommer att offentligen försvaras i patologiska institutionens föreläsningssal, ingång 78, Skånes universitetssjukhus, Malmö, fredagen den 29 april 2011 kl. 9.00.

Fakultetsopponent:
Professor Per Hall
Karolinska Institutet, Stockholm

DOKUMENTDATABLAD enl SIS 61 41 21

Organization LUND UNIVERSITY Department of Laboratory Medicine Skåne University Hospital Malmö, Sweden	Document name DOCTORAL DISSERTATION
	Date of issue April 29, 2011
Author(s) Sophia Harlid	Sponsoring organization

Title and subtitle

Molecular Epidemiology of Breast Cancer

Abstract

Hereditary breast cancer constitutes a considerable fraction of the total number of breast cancer cases occurring each year. Up until recently very few breast cancer predisposing genes were known, but many new common polymorphisms contributing to increased cancer susceptibility are continuously being identified.

This thesis has focused on familial breast cancer and identifying as well as investigating common low-penetrant polymorphisms contributing to breast cancer risk. We hypothesized that since methylation of the promoter region is a common phenomenon of tumour suppressor genes, turning them off. Inherited methylation potential, in the form of common CpG-SNPs, might affect cancer risk. We conducted a large study in five different independent population cohorts (comprising totally >3000 cases) to test this hypothesis in genes previously implicated either in breast cancer or methylation. In this study we were able to identify one SNP possibly associated with breast cancer in the ESR1 gene.

We also examined previously identified common variants affecting breast cancer risk by replicating them in the same five cohorts and examining how the risk increased with increasing number of risk-alleles. A highly significant increasing trend was seen (p=9.3x10-26). Based on comparisons with other replication studies, using different study designs, we concluded that the addition of SNPs from the two, highly replicated, loci FGFR2 and TOX3 could add information to screening of high risk families.

Possible interactions between common genetic variants and established environmental or phenotypic risk factors for breast cancer were examined in two different studies comprising 2063 and 728 breast cancer cases respectively. The significant findings from these two studies were few and may be contributed to coincidence.

In summary we found that methylation potential might be a factor worth considering when searching for SNPs implicated in cancer risk and that common low-penetrant variants contribute to breast cancer risk with the risk increasing substantially with increasing number of risk alleles. To further evaluate possible interactions between common variants and environmental risk factors very large cohorts will be needed.

Key words: Breast cancer, Heritability, Methylation potential, CpG-SNPs, Common variants, Gene-environment interactions

Classification system and/or index termes (if any):

Supplementary bibliographical information:		Language English
ISSN and key title: 1652-8220		ISBN 978-91-86671-81-5
Recipient's notes	Number of pages 140	Price
	Security classification	

Molecular Epidemiology of Breast Cancer

Sophia Harlid

Doctoral Thesis

Malmö 2011

Cover and Layout: Sophia Harlid
Published and printed by Media Tryck
ISSN: 1652-8220
ISBN: 978-91-86671-81-5
Lund University, Faculty of Medicine Doctoral Dissertation series 2011:33

Till Elvira och Arvid

"The road goes ever on and on

Down from the door where it began.

Now far ahead the road has gone,

And I must follow, if I can,

Pursuing it with eager feet,

Until it joins some larger way

Where many paths and errands meet.

And whither then? I cannot say."

-J.R.R. Tolkien (The Fellowship of the Ring)

TABLE OF CONTENTS

LIST OF PAPERS

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals.

I. **Sophia Harlid**, Malin IL Ivarsson, Salma Butt, Shehnaz Hussain, Ewa Grzybowska, Jorunn Erla Eyfjörd, Per Lenner, Asta Försti, Kari Hemminki, Jonas Manjer, Joakim Dillner, Joyce Carlson

A candidate CpG SNP approach identifies a breast cancer associated ESR1-SNP
International Journal of Cancer, e-pub March 2011

II. **Sophia Harlid**, Malin IL Ivarsson, Salma Butt, Ewa Grzybowska, Jorunn Erla Eyfjörd, Per Lenner, Asta Försti, Kari Hemminki, Jonas Manjer, Joakim Dillner, Joyce Carlson

A Nested Case Control Study to Evaluate Multiple Low-Penetrant Risk Alleles for Breast Cancer and Comparison with other Study Designs
Submitted

III. **Sophia Harlid**, Salma Butt, Malin IL Ivarsson, Per Lenner, Jonas Manjer, Joakim Dillner, Joyce Carlson

Effect of genetic susceptibility and height, BMI, and hormone replacement therapy on the risk of breast cancer
Manuscript

IV. Salma Butt, **Sophia Harlid**, Signe Borgquist, Malin IL Ivarsson, Göran Landberg, Joakim Dillner, Joyce Carlson, Jonas Manjer

Genetic predisposition, parity, age at first childbirth and risk for breast cancer
Submitted

ABSTRACT

Hereditary breast cancer constitutes a considerable fraction of the total number of breast cancer cases occurring each year. Up until recently very few breast cancer predisposing genes were known, but many new common polymorphisms contributing to increased cancer susceptibility are continuously being identified.

This thesis has focused on familial breast cancer and identifying as well as investigating common low-penetrant polymorphisms contributing to breast cancer risk. We hypothesized that since methylation of the promoter region is a common phenomenon of tumour suppressor genes, turning them off. Inherited methylation potential, in the form of common CpG-SNPs, might affect cancer risk. We conducted a large study in five different independent population cohorts (comprising totally >3000 cases) to test this hypothesis in genes previously implicated either in breast cancer or methylation. In this study we were able to identify one SNP possibly associated with breast cancer in the ESR1 gene.

We also examined previously identified common variants affecting breast cancer risk by replicating them in the same five cohorts and examining how the risk increased with increasing number of risk-alleles. A highly significant increasing trend was seen ($p=9.3x10^{-26}$). Based on comparisons with other replication studies using different study designs we concluded that the addition of SNPs from the two, highly replicated, loci FGFR2 and TOX3 could add information to screening of high risk families.

Possible interactions between common genetic variants and established environmental or phenotypic risk factors for breast cancer were examined in two different studies comprising 2063 and 728 breast cancer cases respectively. The significant findings from these two studies were few and may be contributed to coincidence.

In summary we found that methylation potential might be a factor worth considering when searching for SNPs implicated in cancer risk and that common low-penetrant variants contribute to breast cancer risk with the risk increasing substantially with increasing number of risk alleles. To further evaluate possible interactions between common variants and environmental risk factors very large cohorts will be needed.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Bröstcancer är den vanligaste cancerformen i världen bland kvinnor och bara i Sverige insjuknar varje år runt 7000 personer. Vi vet att omkring 10-15% av fallen troligen beror på ärftliga faktorer och det har under senare tid gjorts stora insatser för att klargöra exakt vilka arvsanlag som har betydelse för utveckling av sjukdomen. Under början av 90-talet identifierades två anlag (Bröstcancergen 1 och 2), som båda medför stor risk att drabbas av bröstcancer. Fortfarande står dock majoriteten av all ärftlig bröstcancer utan förklaring. Det man vet är att vissa familjer har en ökad benägenhet att drabbas.

De flesta är nu överens om att orsaken till många fall av ärftlig bröstcancer inte står att finna i några få arvsanlag som ger väldigt hög risk utan istället i flera anlag som bara ökar risken med ett fåtal procent. Det är kombinationen av många små riskökningar som tillsammans gör en person mer benägen att drabbas.

För att finna dessa anlag har forskare världen över använt sig av flera olika metoder, bland annat sökning i s.k. kandidatgener (dvs. anlag som man tror har betydelse för utvecklingen av cancer). Det har även utförts stora studier där man ”sökt igenom” all arvsmassa i cellerna efter skillnader mellan sjuka och friska kvinnor. Båda metoderna har gett resultat, men den senare har varit mest framgångsrik och har hittills identifierat drygt 20 nya platser i arvsmassan som tycks påverka cancerrisken. Nu vill många forskare gå vidare och lära sig mer om vad man funnit och vad de observerade skillnaderna mellan sjuka och friska har för praktiska betydelser.

Studierna i den här avhandlingen har haft som syfte att förbättra kunskaperna kring genetiska riskfaktorer för bröstcancer, dels genom att försöka finna nya faktorer, dels genom att bekräfta andras fynd. Arbetena har även försökt bringa klarhet i om och hur några av de ärftliga faktorerna samverkar med miljöfaktorer som längd, vikt, medicinering och barnafödande.

I arbete 1 försökte vi finna nya genetiska riskfaktorer för bröstcancer genom att leta efter en speciell sorts genetiska variationer (s.k. SNPar), som vi tror skulle kunna påverka på och avstängning av arvsanlagen. Vi använde en metod som bygger på hypotesen att vissa sorters SNPar är mer benägna än andra att förändra normala celler till cancerceller. Genom att fokusera på att leta efter just sådana SNPar i anlag som vi visste var viktiga för bröstkancerutveckling eller prognos identifierade vi en SNP i ett anlag till en receptor som känner igen hormonet östrogen, vilket man vet är starkt länkat till bröstcancerprognos. Vår förhoppning är att denna metod ska kunna användas inte bara på kandidatgener utan på hela arvsmassan.

I arbete 2 har vi undersökt vanliga SNPar med låga riskökningar funna i andra stora studier och testat om vi får samma resultat när dessa SNPar testas i populationer från norra Europa. Vi undersökte även hur risken att få bröstcancer ökar om man har många av dessa SNPar jämfört med om man bara har några få. Vi kom fram till att de flesta anlagen gick att hitta i vår studiepopulation samt att risken för bröstcancer ökar avsevärt ju fler av dessa SNPar man har. När vi jämförde vår studie med andra liknande fann vi att två SNPar hade återkommit i samtliga studier. Vi drog slutsatsen att man skulle kunna testa för dem inom familjer som tycks ha en hög benägenhet för ärftlig bröstcancer.

Inom arbete 3 och 4 analyserade vi möjliga interaktioner mellan de tidigare testade SNParna och flera vanliga etablerade riskfaktorer för bröstcancer. Vi tittade bland annat på om individer med vissa arvsanlag hade specifikt ökad risk att drabbas av bröstcancer när de var under hormonbehandling vid menopaus. Man vet sedan tidigare att denna behandling är starkt förknippad med hög bröstcancerrisk. Vi undersökte även hur ålder vid första graviditet samt antal barn i kombination med de olika SNParna påverkar bröstcancerrisken. I ingen av dessa studier fann vi några interaktioner av hög signifikans. I andra studier där man undersökt liknande interaktioner har man inte heller funnit något som tyder på starka samband mellan genotyp och etablerade riskfaktorer. Detta trots att man tidigare trott att sådana samband borde ligga bakom en stor del av de hittills oförklarade ärftliga bröstcancerfallen. Det är möjligt att interaktioner mellan miljö och arvsmassa är betydligt mer komplexa än man tidigare trott samt att vi och andra använt för små studiebaser för att kunna hitta några samband.

Sammanfattningsvis kan man säga att studierna i denna avhandling bidragit till vår förståelse av hur vanliga genetiska variationer (SNPar) påverkar risken att insjukna i bröstcancer samt belyst svårigheten med att fastställa samband mellan miljörelaterade och ärftliga riskfaktorer.

ABBREVIATIONS

BMI	Body Mass Index
CCPRB	Cancer Control using Population-based Registries and Biobanks
CI	Confidence Interval
CIS	Carcinoma in Situ
CNV	Copy Number Variants
DNA	Deoxyribonucleic Acid
ERα	Estrogen Receptor α
ERβ	Estrogen Receptor β
GWAS	Genome Wide Association Studies
HER2	Human Epidermal Growth Factor Receptor 2
HRT	Hormone Replacement Therapy
LD	Linkage Disequilibrium
LOD	Logarithm of Odds
MAF	Minor Allele Frequency
MBP	Methyl Binding Protein
MDC	Malmö Diet and Cancer Study
MPP	Malmö Preventive Project
MGB	Minor Grove Binder
MSP	Mammography Screening Program
NSHDS	North Sweden Health and Disease Study
OR	Odds Ratio
OC	Oral Contraceptives
PCR	Polymerase Chain Reaction
PR	Progesterone Receptor
RR	Relative Risk
SNP	Single Nucleotide Polymorphism
SAP	Shrimp Alkaline Phosphatase
TNM	Tumour Node Metastases
WHO	World Health Organisation

Gene Names (HUGO Gene Nomenclature Committee)

ATM	Ataxia telangiectasia mutated
BRCA1	Breast cancer 1, early onset
BRCA2	Breast cancer 2, early onset
BRIP1	BRCA1 interacting protein C-terminal helicase 1
CASP8	Caspase 8
CDH1	Cadherin 1
CHEK2	CHK2 checkpoint homolog
DNMT1	DNA (cytosine-5-) -methyltransferase 1
ESR1	Estrogen receptor 1
FGFR2	Fibroblast growth factor receptor 2
HCN1	Hyperpolarization activated cyclic nucleotide-gated potassium channel 1
H19	H19, imprinted materna lly expressed transcript
LSP1	Lymphocyte specific protein 1
MAP3K1	Mitogen activated protein kinase kinase kinase 1
MECP2	Methyl CpG binding Protein
PALB2	Partner and localizer of BRCA2
PGR	Progesterone receptor
PTEN	Phosphatase and tensin homolog
STK11	Serine/Threonine kinase 11
TOX3	TOX high mobility group box family member 3
TP53	Tumour protein 53

INTRODUCTION

Molecular epidemiology is an extension of traditional epidemiology in which molecular methods are used to reach or support epidemiological conclusions (1, 2). The term and concept acquired recognition in the early 1980s when applied to infectious diseases and chronic illnesses.

As methods have become more and more advanced, it has opened a new arena for epidemiologists making it possible to not only to establish an association between an exposure and a specific disease but also to determine how and why the exposure causes the disease (the distinction between traditional epidemiology and molecular epidemiology is depicted in Figure 1). An important milestone in this development has been the discovery of polymerase chain reaction (PCR) that has made it possible to identify different genetic susceptibly to diverse diseases (3). One of the diseases that attract most attention in discussions of genetic susceptibly is cancer.

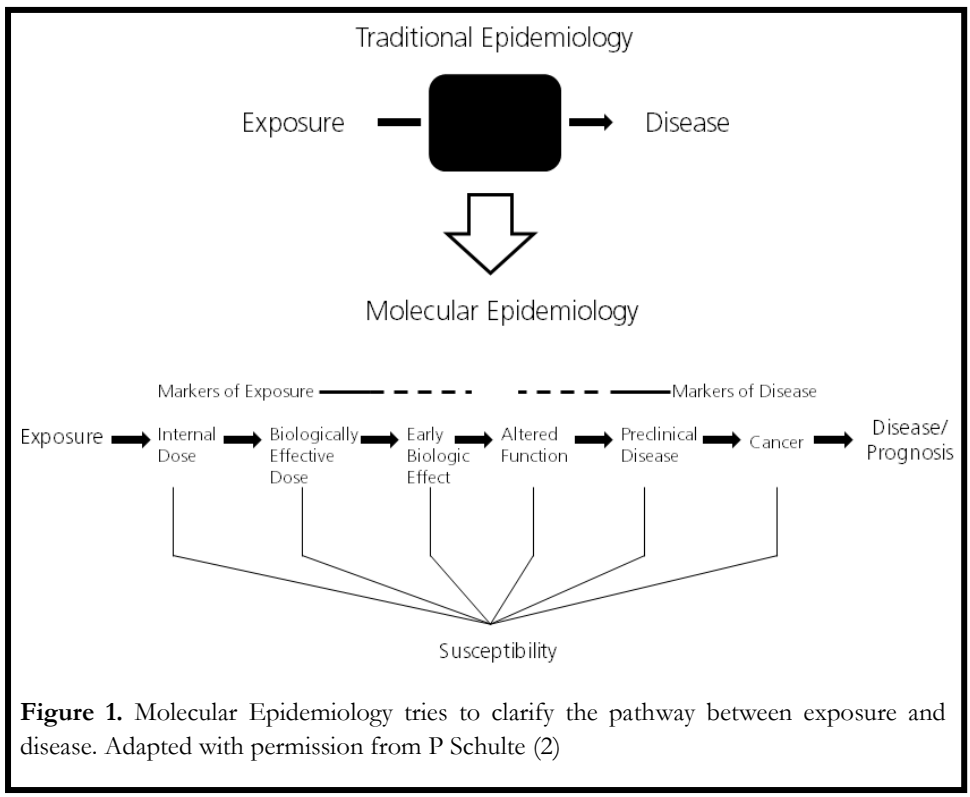


Figure 1. Molecular Epidemiology tries to clarify the pathway between exposure and disease. Adapted with permission from P Schulte (2)

Most cells in an organism have access to the complete genome, which holds far more information than a single cell will ever require. The cells maintain the ability to grow and divide even after development has been completed, something that serves an important function if a tissue is injured. But sometimes the system fails and access to the entire genome can become harmful when individual cells retrieve information that normally would be denied to them. In addition the genome is subjected to mutations and damage that can alter its structure and information content. The consequence can be devastating when the resulted, mutated genes can produce cells with abnormal and sometimes aggressively growing phenotypes (4).

One of the earliest described cancer forms is breast cancer, mentioned in the Edwin Smith Papyrus, an ancient Egyptian medical text on surgical trauma. It describes eight cases of "ulcers" of the breast that were treated by cauterization, the burning of a body part. In the text the writing refers to the condition as being untreatable (5, 6).

Today we have come a long way since that statement and breast cancer is no longer untreatable. Still it remains one of the most feared diseases in the world and the most common cancer form among women worldwide (7). The incidence of breast cancer has been growing, especially in developed countries; something that is being counteracted by the fact that the five-year survival is constantly increasing (8).

Breast cancer has long been known to have a hereditary component, and to cluster in certain families. The relative contributions of pure genetic effects and of lifestyle remain unclear. What is becoming more evident is that genotypic inheritance and lifestyle are probably inseparably intertwined. The combination of genetic factors and lifestyle makes us who we are and also determines our individual risks of attracting disease. The purpose of the investigations in this thesis was to search for new hereditary components of breast cancer that might be specifically susceptible to environmental exposures and to cast some light on how other established genetic factors might interact with recognized environmental and lifestyle factors known to affect breast cancer risk (9).

BACKGROUND

Breast Cancer Definitions and Classification

All tumours arise from normal tissue and breast cancer, as the name implies, is defined as cancer that originates from breast tissue (4) but the progression from normal breast tissue to invasive cancer is poorly understood. Non-invasive breast cancer is called carcinoma in situ (CIS) and can arise from either ductal or lobular hyperplasia of epithelial cells (10). Cancer that has progressed into surrounding tissue is called invasive breast cancer and usually has the ability to metastasise (11). Tumours are categorized according to type and size, histopathology, invasiveness, tumour stage and receptor expression. As our molecular techniques have improved we have gained a deeper understanding of diverse breast cancer types and how they differ (12).

Tumours are classified by the WHO into six main types; ductal, lobular, mucinous, medullary, papillary and tubular carcinoma. Ductal and lobular tumours represent around 90-95% of all cases (13). Histological grade is often classified according to the Nottingham Grade classification which was introduced in the 1990s and includes three different parameters (tubule formation, nuclear pleomorphism and mitotic counts) (14). Tumour stage classification incorporates Tumour size (T), lymph Node status (N) and Metastasis (M) (usually shortened to TNM). The TNM system has been somewhat controversial but remains well used by clinicians (15, 16).

Expression of different receptors, known to affect the prognostic and predictive values of therapy, is also used to characterize the tumours. They are classified according to expression of estrogen receptor α and β (ERα and ERβ), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). The lack of any ER or PR on the tumour cells makes the prognosis worse whereas the lack of HER2 expression does not. If the tumour lacks all three receptors it is called triple negative, this feature often indicates a poor prognosis (17, 18).

Epidemiology

Breast cancer is known to be one of the most common cancer forms among women worldwide (7) and countries with the highest breast cancer incidence rates are Switzerland, US (Caucasians), Italy, and other Western European countries (Figure

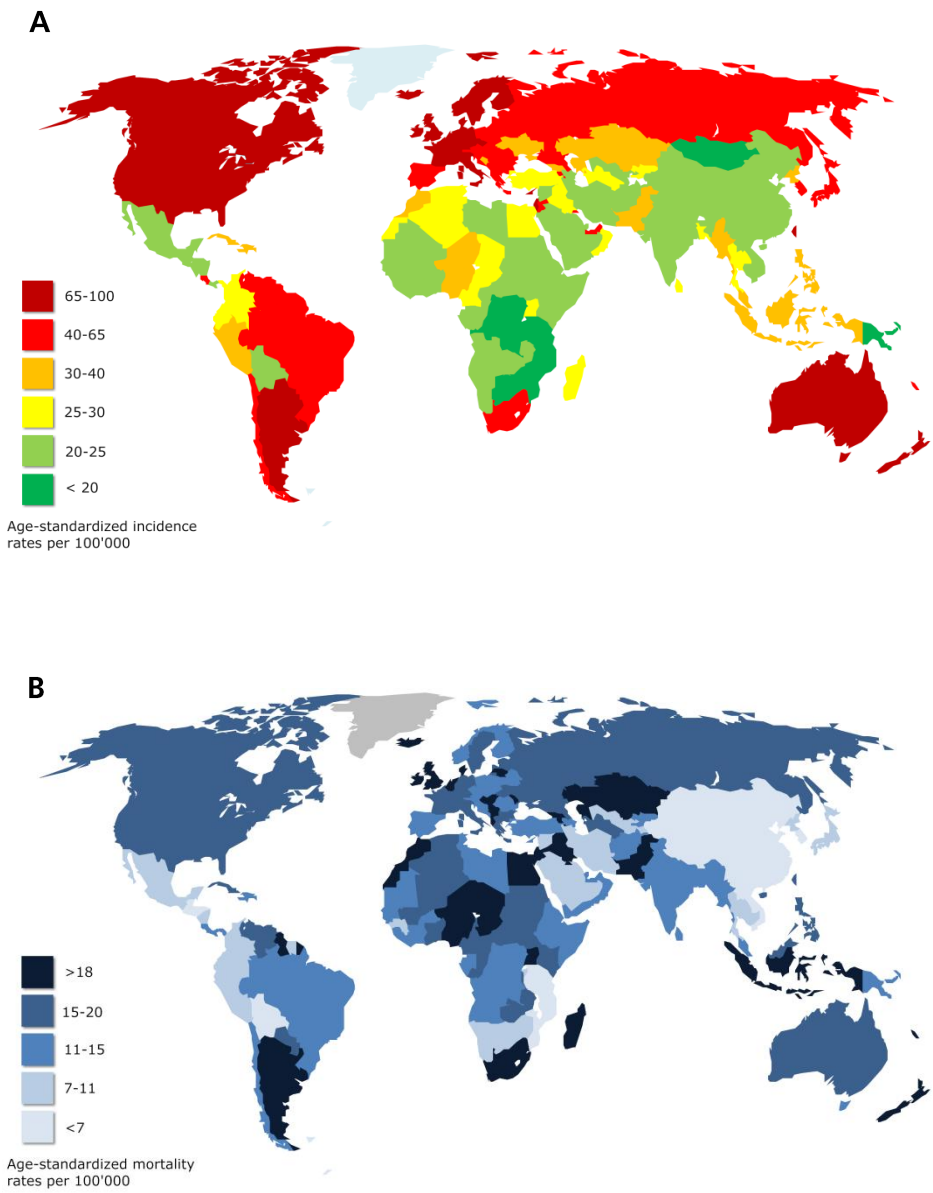


Figure 2. Breast Cancer incidence **(A)** and mortality **(B)** worldwide 2008. Adapted from GLOBOCAN 2008 (19)

2A). The high incidence of breast cancer in the United States and Europe is believed to reflect reproductive choices lifestyles associated with a higher standard of living, including early menarche (20), late child bearing, fewer pregnancies, use of hormone replacement therapy (HRT) as well as increased detection (through mammography) (8).

Throughout the 20th century we saw a rise in breast cancer rates all over the world with the most dramatic increases in developing countries (9, 21). In many developed countries though, breast cancer incidence has started to decrease during the past decade, something that can probably be attributed to a reduction in the use of HRT (8). On the contrary, the increase in breast cancer rates observed in many Eastern European, Asian, Latin American, and African countries have not been declining and could be a sign of their corresponding lifestyle changes as these countries become "westernized".

The incidence of breast cancer is more closely associated to age than to any other risk factor, it increases rapidly during the reproductive years and then more slowly after about 50 years of age (average age of menopause) (9). In many countries, including Sweden, a peak is reached at about 65 years of age after which the incidence reaches a plateau or starts to decline again (19). Breast cancer is often more aggressive in young women, the group most affected by inheritance of high risk mutations in the BRCA1 or BRCA2 genes (22, 23).

Patterns of mortality rates differ somewhat from those for incidence rates (Figure 2A and B). One distinction not visible on the map is that US Caucasians and Australians have relatively low rates and US African-Americans have very high rates (19). The low mortality rate in some western countries is believed to reflect improved screening and treatment in these countries. The opposite is true for many of the poorer countries in the world were an unfavourable mortality trend can be attributed to lack of early detection and limited treatment. In Algeria only 40% of afflicted women survive 5 years after diagnosis compared to 89% of all women in the United States and more than 80% of women in Northern and Central Europe (8, 24).

Hereditability

Hereditability of cancer can be viewed from several angles. Inherited cancer may refer to all cancer cases bearing established causal genetic mutations, and it may refer to cases in families with multiple cancer cases, although no common causal genetic trait has been identified.

Breast cancer has been recognized to cluster in families since the 1860s, when a family suffering from extensive numbers of breast cancer cases was described by the French surgeon Paul Broca (25, 26). Ever since then, explanations for these clusters have been sought. It is estimated that about five to ten percent of all breast cancer cases are caused by mutations in high risk genes, such as BRCA1 and BRCA2 (27-30). Patients who carry mutations in these genes or have a pattern of inheritance that corresponds to the Mendelian model are denoted as having hereditary breast cancer. In many families no such pattern can be found, but the history is still indicative of some kind of genetic predisposition. Women from these families are sometimes said to have familial breast cancer, a classification that is not clearly defined, and the exact percentage that these cases contribute vary (Figure 3) (26, 31-33). One could speculate that some of these clusters might be due to a shared environment rather than genetic factors. Nonetheless twin as well as simulation studies suggest that this is not the case, but that genetic factors are ultimately responsible for the observed familiar clustering (34, 35).

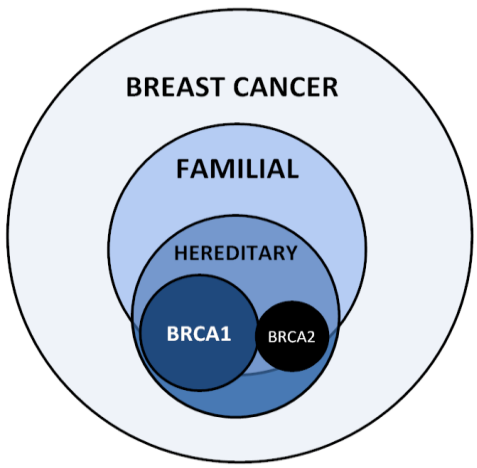


Figure 3. Relationship between hereditary breast cancer (including BRCA1/2), familial breast cancer and sporadic breast cancer. Mutations in BRCA1/2 genes are responsible for about 5% of all breast cancer. Hereditary breast cancer denotes cancer with early onset, several affected family members and often a clear inheritance pattern, and it is suggested to make up about 10% of all breast cancer cases. The definition of familial breast cancer is debated but usually it includes all individuals with one or more first- and or second degree relatives with breast cancer including those where there is no clear inheritance pattern. The familial breast cancer cases are alleged to add as much as an additional 15% to the total number of breast cancer cases in selected populations. *Reprinted with permission from The Journal of Clinical Oncology*(33).

ENVIRONMENTAL RISK FACTORS FOR BREAST CANCER

Migration studies indicate that most of the risk factors for breast cancer are dependent on the environment we live in (36, 37). Even though they do not explain all of the additional risk a substantial part of it can be attributed to lifestyle related factors (9).

Reproductive Factors

Factors such as age at menarche and menopause, age at first birth, number of births and duration of breast feeding are all related to hormonal factors and prolonged exposure to endogenous hormones seems to increase breast cancer risk (38).

Age at menarche and menopause are crucially related to the total hormonal exposure during a woman's lifetime. Late onset of menarche as well as early menopause are associated with significant risk decreases of 5% per year and 3% per year respectively (39).

Nulliparity (no childbirths) is associated with increased breast cancer risk and studies of age at first birth show that a young age is associated with lower overall risk. Women who give birth to their first child after 30 years of age have a high risk of breast cancer in the years immediately following delivery, compared to women with first childbirth before the age of 20 (Figure 4) (40-42).

The connection between parity and breast cancer risk may have several explanations. The breast undergoes drastic changes during pregnancy mostly involving increased proliferation and differentiation of epithelial tissue. In the first years following delivery there is an increased breast cancer risk that eventually declines below the risk estimated for nulliparous women (42, 43). One plausible hypothesis is that the risk increase is due to proliferation of malignant transformations present in the breast before and during pregnancy. An older woman might have accumulated more malignant transformations in the breast tissue prior to pregnancy and will therefore have a higher risk of developing breast cancer. The net effect of pregnancy is fewer epithelial structures vulnerable to malignant transformation ultimately resulting in risk reduction (44, 45). A later study showed that a specific genomic signature is induced in breast tissue following the first pregnancy, something that might have a specific protective effect (46).

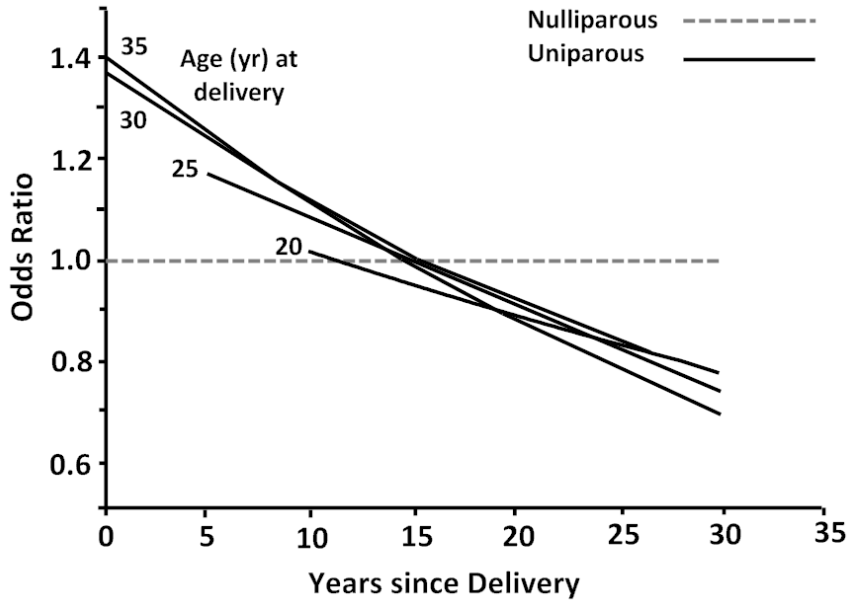


Figure 4. Odds Ratios for the risk of breast cancer in uniparous women of various ages at delivery, according to the numbers of years since delivery. Nulliparous women were used as the reference group. *Reprinted with permission from The New England Journal of Medicine (42).*

The effect of breastfeeding on breast cancer risk has been debated and if there is a true risk reduction it is probably quite small. A large reanalysis of data from more than 45 studies in 30 different countries found a 4% risk reduction for every 12 months of breast feeding (47), subsequent studies have however failed to find a connection (48).

Hormone Replacement Therapy

Several studies have shown that most risk factors for breast cancer probably act through hormone-related pathways, with oestrogens playing a major role. Women with post-menopausal breast cancer often exhibit increased concentrations of circulating oestrogens in the bloodstream and reproductive factors affecting breast cancer risk (described above) are also often related to oestrogen exposure (45).

The use of hormone replacement therapy (HRT) to alleviate symptoms of menopause was introduced in the 1960s. During the 1990s HRT was often prescribed

to reduce heart disease risks as well. The connection between HRT and breast cancer was not clear at the time but became the focus of two large, randomized, placebo-control trials funded by the National Institutes of Health in the United States. The first study was halted prematurely because of a 26% risk increase observed in the women receiving HRT, this number was later changed to 49% when correcting for women who dropped out (49, 50). The conclusions were that breast cancer risk is increased mainly for women who recently entered menopause. The risk increase starts after 3-5 years of HRT use and then rises progressively; it returns to normal within 3-5 years of cessation (51). These risks are mainly attributed to use of combined hormonal therapy, and seem to favour tumours with low proliferation rates and a better prognostic outcome (52).

Several independent studies have also confirmed the connection between use of HRT and increased breast cancer risk. One Swedish study (53) examining risk of HRT in women with previous breast cancer was ended prematurely in 2003 because of high risks for the women involved and the conclusion from the million women study in the UK was that that all types of HRT significantly increased the risk of breast cancer (54).

The decrease in breast cancer incidence observed during recent years in the United States could be attributed significantly reduced HRT use, even though other factors might contribute (49).

Oral Contraceptives

Many studies evaluating the contribution of oral contraceptive (OC) use to breast cancer risk have been undertaken, some with conflicting results. It is reasonable to conclude that the use of OC leads to a minor increase in breast cancer risk for current users or women who have used OC during the last ten years. No evidence of lasting effects ten or more years after cessation of use has been reported (55, 56).

Anthropometrics

Anthropometrics refers to measurements and include height, weight, body mass index (BMI) and other proportions of the human body. Some of these factors have been connected to breast cancer risk.

Height has been modestly associated with breast cancer risk and within populations a 10 cm increase in height corresponds to a 10 % risk increase. There is as yet no clear explanation for this connection but it has been suggested to depend on energy intake

during early life and adolescence (9, 57). An interesting fact when it comes to height and breast cancer risk is the fact that oestrogen, as outlined above, play key roles both in breast cancer development and human growth regulation. Oestrogen stimulates the pubertal growth spurt and mutations in the ESR1 gene (coding for ERα) have been reported to delay fusion of the epiphyseal plates at puberty (58, 59). An association between body height and mutations in ESR1 has also been found (60) , and might point towards a more hormone related link.

BMI has been connected to increased breast cancer risk in postmenopausal women (57, 61). This could be attributed to increased concentrations of oestrogen in the bloodstream that is associated with a higher weight (62). In premenopausal women however, this connection is unclear (63).

Other Factors Related to Environment or Lifestyle

Many other environmental factors potentially contribute to increased breast cancer risk of varying degrees, including; ionizing radiation, exposure to chemicals, dietary factors, alcohol, passive smoking and socioeconomic status amongst others (9, 64, 65). Some of these are firmly established, such as radiation and some chemical exposures (49, 66), while others remain more controversial, such as smoking and diet (65, 67). Physical activity has been inversely associated with breast cancer risk and the connection is considered well recognised (68, 69).

EPIGENETICS AND BREAST CANCER

All cells in the body share the same DNA; still they exhibit diversely different functions. How is this possible? And by what means do they regulate their gene-expression patterns? At least some of the answers to this question can be found in the field of epigenetics which can be viewed as the middle ground between environment and genetics.

The term epigenetics refers to factors that can be inherited through mitotic cell-division but which are not themselves directly part of the DNA sequence. Mainly these factors consist of methylation patterns or histone modifications (70, 71).

Methylation in Cancer

DNA-methylation was one of the first aspects of epigenetics to gain large interest from cancer researchers. It refers to the binding of a methyl-group to the 5-carbon position on the nucleotide cytosine, a modification normally performed by the actions of a DNA-Metyltransferase protein (DNMT). In mammals, DNA-methylation occurs almost exclusively at cytosines followed by a guanine, so called CpG dinucleotides. About 75% of CpG dinucleotides in the human genome are methylated in normal cells and more than 90% of them reside in transposons and repetitive elements (72, 73). A methylated cytosine can easily be transformed to a thymine by deamination explaining why this is the most common form of mutation in the genome (74).

The first connection to cancer was suggested in 1983 when it was discovered that cancer tissue differed from normal tissue when comparing methylation patterns (75). Normal cells are usually mostly methylated, the exception being CpG islands which are short stretches of DNA located in the promoter region of about 60% of all genes. In the CpG-islands methylation is rare and has been shown to mediate repression of transcription both directly (by inhibiting binding of transcription factors (TF)) and indirectly (by attracting methyl-CpG binding proteins that interact with co-repressor molecules and chromatin to silence transcription) (Figure 5) (76). In cancer cells the pattern is reversed with large hypomethylated regions on a genome wide scale and hypermethylation occurring at CpG islands (77, 78).

Breast cancer tissue also exhibits this distinct methylation profile, and tumours are often methylated on a genome-wide scale, even though distinct genes frequently lack hypomethylation. On the other hand more than 100 individual genes have been reported to be hypermethylated in breast cancer tumour tissue or cell lines (79). One

of these genes is BRCA1 (described in detail below) that is sometimes found to be turned off by hypermethylation of the promoter region in sporadic breast cancer (80). Another important protein is $p16^{ink4A}$ that functions in the cyclinD-Rb pathway and is found hypermethylated in many forms of cancers including breast cancer (79, 81).

Receptor status is an important prognostic marker in breast cancer and tumours lacking expression of ERα, ERβ or PR often have a worse prognosis than those expressing these receptors. Methylation of the genes ESR1 (coding for ERα) and PGR (coding for PR) have been proposed as mechanisms for the development of receptor negative tumours, something that still needs confirmation (77, 82).

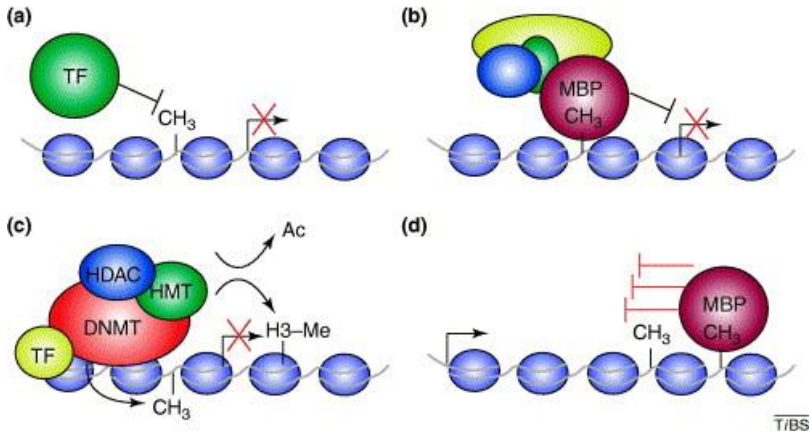


Figure 5. Mechanisms of DNA-methylation-mediated repression. **(a)** DNA methylation in the cognate DNA-binding sequences of some transcription factors (TF) can result in inhibition of DNA binding. By blocking activators from binding targets sites, DNA methylation directly inhibits transcriptional activation. **(b)** Methyl-CpG-binding proteins (MBPs) directly recognize methylated DNA and recruit co-repressor molecules to silence transcription and to modify surrounding chromatin. **(c)** In addition to their DNA methyltransferase activities, DNMT enzymes are also physically linked to histone deacetylase (HDAC) and histone methyltransferase (HMT) activities. In this case, the addition of methyl groups to DNA is coupled to transcriptional repression and chromatin modification. **(d)** DNA methylation within the body of genes can also have a dampening effect on transcriptional elongation. MBPs might be involved in inhibiting elongation, either directly or by their effects on the surrounding chromatin structure. *Reprinted with permission from Elsevier* (76).

IDENTIFYING GENETIC RISK FACTORS

Family Based Studies

One of the first approaches used to identify the position of disease loci, harbouring the genes responsible for genetic disorders, was to examine families demonstrating a specific syndrome or trait. This led to the development of familial linkage analysis. The simplest form is called parametric linkage analysis and takes advantage of the fact that loci situated close together on the same chromosome are inherited together more often than loci that are far apart or on different chromosomes. The further apart two loci are the more likely it is that a recombination will occur during meiosis and break up the co-segregation. The number of times that this occurs in a subset of familial offspring can be used to calculate the recombination fraction θ (the probability of recombination between two loci at meiosis). The two loci of interest are usually represented by a marker (with a known genetic position) and a disease associated loci (with unknown position) (Figure 6) (71, 83).

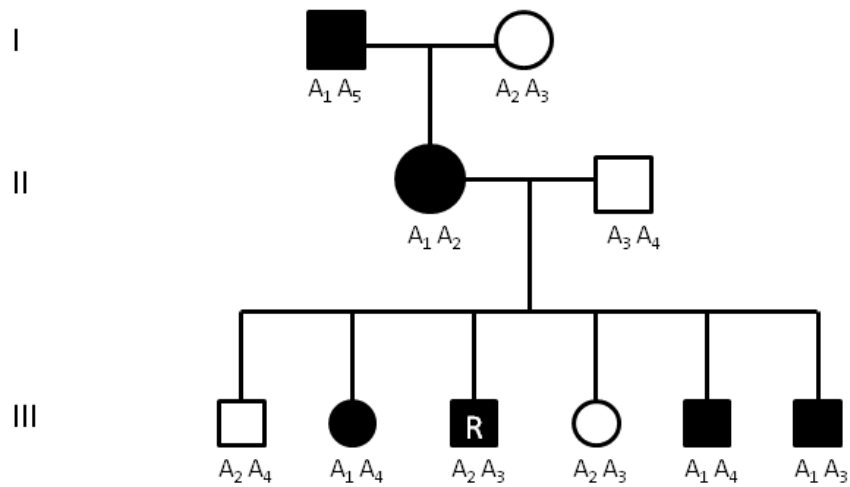


Figure 6. Pedigree used to identify the position of a disease associated loci in linkage with a known marker (A). Individuals suffering from the disease and therefore carrying a mutation in the disease locus are blackened. All individual variations at marker position A are known. Individual III_3 can be identified as a recombinant (marked with R), helping to evaluate the distance between the disease loci mutation and marker A.

In these studies large subsets of markers spanning the entire genome are used and the logarithm of odds (LOD) score is calculated between the different markers and the disease loci. The LOD score is a function of the recombination fraction θ and the chromosomal position and denotes the likelihood that the two loci are linked. A high LOD score equals increased probability of linkage and vice versa. Mathematical functions are used to calculate the maximum LOD score depending on θ or position. For parametric linkage analysis to work satisfactorily the genetic model must also be specified (e.g. mode of inheritance and frequency of disease alleles) (71, 83).

Problems that render linkage analysis more difficult are genetic heterogeneity (when a disease can be caused by mutations in different genes), incomplete penetrance (when a carrier of the disease trait does not always exhibit the disease phenotype) and multifactorial diseases (where several genes and environmental factors contribute to disease risk). To cope with these factors non-parametric linkage analysis can be used, that for example takes advantage of sibling pairs (83).

When it comes to breast cancer, parametric linkage analysis was successful in identifying the two major breast cancer genes (BRCA1 and 2) despite heterogeneity (83). The mapping of BRCA1 to 17p21 used data from 23 extended families with 146 cases of breast cancer (84) and BRCA2 was subsequently mapped to 13q12 using data from 15 families known to lack mutations in BRCA1(85). Regions surrounding the mapped loci of interest were then scrutinized for likely causal genes, a process denoted as positional cloning (86, 87).

Candidate Gene Studies

After the discoveries of BRCA1 and BRCA2 and the disclosure of the fact that they were involved in DNA-repair the insight into the molecular pathology of breast cancer widened considerably. This led to the possibility of targeting other genes, coding for proteins involved in the same or similar pathways as BRCA1 and 2 (88).

One method of examining these so called "candidate genes" has been by complete resequencing of interesting loci in large numbers of cases and controls to compare the total number of possibly pathogenic mutations. This approach is both tedious and costly and has yielded few results (28, 88, 89).

The utilisation of association studies to determine if a gene or locus could be connected to breast cancer risk has instead become more and more common. The difference between linkage and association is that linkage describes a specific genetic relationship between loci (not alleles) and association is simply a statistical statement

about the occurrence of specific alleles in cases vs. controls (71). The primary focus for these studies was to centre on candidate genes, analyse them for the occurrence of common low-penetrance susceptibility variants and compare frequencies in cases and controls. This provided a cheaper and simpler approach than the resequencing described above (89). Unfortunately these studies have not been very successful and only one variant (rs1045485) in the gene CASP8 has persisted in large replication studies (90-92).

Some of the disadvantages of these studies have been their small size, inconsistency and lack of replication resulting in false positives (93, 94). Current collaborations providing large numbers of samples could compensate for this and in the future additional studies of candidate genes are still likely to play a part in elucidating genetic risk factors for breast cancer (91).

Genome Wide Association Studies

The lack of success in identifying new risk genes and loci using the candidate gene approach led to suggestions of unbiased scanning of the whole genome for possible cancer associated variants. This approach presented many obstacles that must be overcome (94). To begin with, all common variants in the genome needed to be identified, a task that was undertaken by the HapMap project initiated in October 2002 (95-97). This large collaboration had as a goal to identify and characterise genetic similarities and differences between humans, something that included mapping the patterns of linkage disequilibrium (LD) among SNPs (98) (Figure 7).

The emergence of these high-density maps of SNPs as well as the development of more affordable genotyping platforms made Genome Wide Association Studies (GWAS) possible. When taking advantage of LD, it was possible to generate panels of a few hundred thousand SNPs in order to "tag" the whole genome (88, 98).

Still, a major problem with this kind of approach is the emergence of false positives. When several hundred thousand SNPs are genotyped chance alone will give rise to many apparently significant associations. These problems are dealt with primarily by setting a very stringent significance threshold (a p-value of $<10^{-7}$) and by replicating initial findings in very large materials comprising tens of thousands of independent cases and controls. These measures appear to have been successful in filtering out findings that are true significant associations (described below) (89, 99). In contrast to candidate gene studies, GWAS has been successfully identifying new potential risk loci for a number of common inherited diseases such as breast cancer, diabetes,

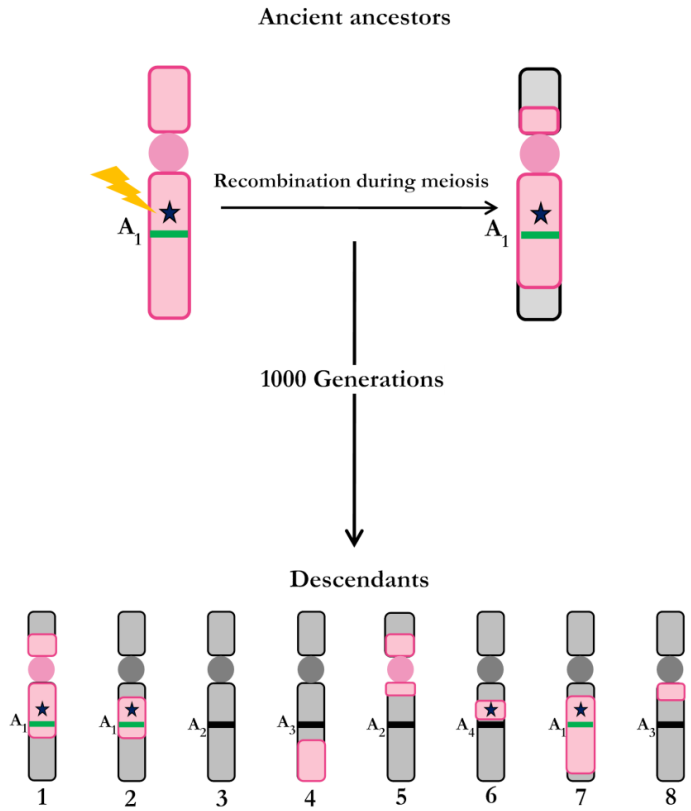


Figure 7. Linkage disequilibrium between a marker (A_1) and a mutation (marked with a star). A common ancestor suffered a mutation close to marker A_1. Since they are so close together they are inherited together most of the time and can therefore be followed on population level. Pink parts represent the ancestral chromosome and grey parts represent chromosomes inherited from others. After 1000 generations only descendant 6 has the mutation without marker A_1. LD differs from normal linkage used in family based studies where marker and mutation tend to be further apart and the chance of recombination at any single meiosis is greater (98, 105, 106).

Alzheimer disease, schizophrenia, colon cancer, etc (100-104). This unbiased approach has revealed many loci that were previously not suspected of being involved in disease development. With the help of these new findings it may become possible to uncover new potential pathways in, e.g. breast cancer pathogenesis(89).

The GWAS approach has met with a fair amount of criticism, mainly concerning the stringent cut-off values that will make it likely to miss associations and the lack of power when it comes to identifying sequence variants with a minor allele frequency

(MAF) of <5%. Also SNPs are the preferred markers used for this type of studies, as they are easily genotyped and have been extensively mapped throughout the genome. This can pose a problem as they are not always in LD with other potentially cancer associated variants like insertions, deletions, inversions or copy number variants (CNV). CNVs especially could prove interesting to examine and have recently gained attention because of their influence on gene expression (98, 107, 108).

Illuminating the "Dark Matter"

Despite the implementation of all the above strategies and the fact that many both rare and common variants have been discovered only a portion of the familial breast cancer risk can be explained by currently identified genes and loci. The remaining "dark matter" may be SNPs missed in GWAS due to lack of power to detect them or structural variants not detectable by SNP genotyping (109). The possibility of complex interactions between the variants themselves or between the variants and the environment is a field that has just started to be explored (110). Finally, uncommon variants or variants lacking haplotype linkage that have been excluded from GWAS panels may require genome-wide resequencing of large regions to be detected. The current rapid development of high-throughput sequencing techniques (111-113) may soon replace SNP-arrays in the search for the complete set of genetic factors explaining familial breast cancer.

GENETIC RISK FACTORS FOR BREAST CANCER

It is known that a familial history of breast cancer constitutes a major risk factor. In women with a first degree relative the risk is doubled and if the affected is a mother or sister it is even higher (114). The quest for responsible factors behind increased genetic predisposition to breast cancer has gained a lot of interest in recent years. Much of this is probably due to the fact that we have seen the discovery of many new low-penetrance loci, and are beginning to unravel the causal mechanisms behind them (115). The different approaches, described in previous sections, used to discover genes involved in breast cancer risk and have led to identification of three main types of predisposing alleles; rare high-penetrant, rare intermediate-penetrant, and common low-penetrant alleles (Figure 8) (88, 89). Despite key breakthroughs in this area we are still a long way from explaining the majority of familial breast cancer cases occurring.

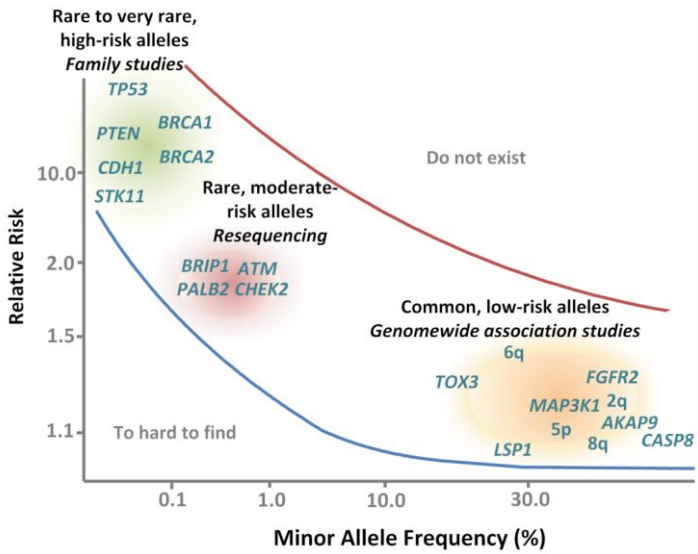


Figure 8. All known breast-cancer susceptibility genes are shown between the red and blue lines. No genes are believed to exist above the red line, and no genes have been identified below the blue line. High-risk syndromic genes are highlighted in green. The moderate-penetrance genes (highlighted in red) have an approximate relative risk of 2.0. There are probably many more genes in this class, but they can be identified only by resequencing candidate genes in affected persons in families with breast cancer. The common, low-risk genes are shown in orange. *Reprinted with permission from The New England Journal of Medicine* (28).

High-Penetrance Genes

In the early nineties the two genes, BRCA1 and BRCA2, were discovered and linked to hereditary breast cancer (85-87). The two corresponding proteins have been shown to have important functions in maintaining genomic stability by promoting repair of double strand breaks in the DNA (116). Other high-penetrance genes seem to be rare and very few have been identified in relation to breast cancer. One is the TP53 gene which is frequently mutated in most malignant tumours. TP53 codes for the P53 protein that functions as a guardian of the genome (117). Mutations in any of these genes usually confer a relative risk (RR) for breast cancer of ten or more (Figure 8).

BRCA1 was the first identified breast cancer gene. It spans >80kb on chromosome 17q21 and cancer associated variants of this gene usually contain loss of function mutations, commonly deletions or insertions, resulting in a truncated or dysfunctional protein (118). Mutations in this gene have been shown to increase the risk of predominantly breast and ovarian cancer, the lifetime breast cancer risk in an individual suffering from a mutation in BRCA1 differs depending on the origin of the cases studied (familial or population based) but seems to be somewhere between 45-80% and the highest risk is seen in carriers under 40 years of age (119, 120). BRCA1 tumours more often have a medullary histopathology, are of higher histologic grade and are frequently triple negative for receptors. TP53 mutations are also common in these tumours compared to sporadic breast cancer (121, 122).

BRCA2 is located on chromosome 13q12 and codes for a 380-kd protein. Abnormalities in this gene include frameshift deletions, insertions or loss of function mutations (123). The BRCA2 protein is not apparently related in structure to BRCA1 but they seem to share many functional similarities explaining why dysfunctional versions of both increase the risk of breast and ovarian cancer (124). The pathological features of BRCA2 induced tumours are less well-defined than BRCA1 tumours and they appear more similar to sporadic tumours even though they are generally of a higher overall grade (121). The lifetime risk of an individual with inherited mutation in the BRCA2 gene is somewhere between 30-90% and just as for BRCA1 mutation carriers it depends on the category of affected individuals studied. Also, biallelic mutation of BRCA2 causes a subtype of Fanconi-anemia designated D1 (88, 119). Fanconi-anemia is a rare syndrome giving rise to bone marrow failure and increased risk of malignancies (125).

TP53 codes for the tumour suppressor protein P53 that is known to play an important part in protecting the cells from malignant transformation (126). Individuals with germ-line mutations in TP53 suffer from Li-Fraumeni syndrome, a

rare autosomal dominantly inherited disorder that increases the risk of several forms of cancer particularly in children and young adults. The high penetrance of TP53 mutations and associated early mortality has prevented acquisition of sufficient numbers of family members for classical familial linkage analysis. Instead it was knowledge of P53s primary functions in maintaining cellular control combined with symptoms of the disease that eventually led to the proposition of TP53 as a plausible candidate gene and its subsequent sequencing in five families with Li-Fraumeni syndrome. Mutations were confirmed in all affected individuals (127-129). The occurrence of germline TP53 mutations in breast cancer families that do not suffer from the classical Li-Fraumeni syndrome is extremely uncommon resulting in a very low contribution to the total number of hereditary breast cancer cases (88, 130, 131).

Three additional putative high-penetrance genes have been identified through familial linkage analysis: **STK11**, **PTEN**, and **CDH1** (88, 132). Peutz-Jegher syndrome, linked to mutations in STK11 and Cowden syndrome, linked with PTEN both result in increased risk of many different malignancies including breast cancer (133, 134). CDH1 was discovered through linkage analysis of a single family in New Zealand and results in increased risk of breast and gastric cancer (135, 136). The exact penetrance of these three genes is debated and might be overestimated due to the bias of studying families with strong phenotypes (88).

The characterisation of these highly predisposing genes marked the start of a new era in cancer genetics and many new findings seemed to be right around the corner. However, this turned out not to be the case. Instead the search for the new BRCA3 gene has remained futile and it appears that the majority of familiar breast cancer cases are not explained by single mutations in high risk genes.

Intermediate-Penetrance Genes

Candidate gene association analysis has identified four genes conferring RR of between two and four: CHEK2, ATM, BRIP1, and PALB2. Mutations in these genes explain a very small part of familial breast cancer cases since they are all very rare in the population (132, 137).

CHEK2 was discovered in families harbouring the Li-Fraumeni syndrome. Activation of the corresponding protein kinase occurs in response to DNA damage and leads to phosphorylation of, amongst others, the p53 and BRCA1 proteins (138, 139).

ATM has multiple functions involving DNA repair and was originally discovered as the gene behind ataxia-telangiectasia, a condition that predisposes the bearer to multiple forms of cancer in addition to immune deficiency and lack of muscle control (140). Mutations in ATM were early suggested to increase breast cancer risk since women related to individuals suffering from ataxia-telangiectasia seemed to have increased susceptibility (141). Many studies were undertaken to establish the connection between ATM and breast cancer risk, although most were inconclusive. Eventually large studies proved that the same mutations that caused ataxia-telangiectasia also moderately increased breast cancer susceptibility (142, 143).

BRIP1 codes for another protein involved in DNA repair that interacts with, amongst others, BRCA1 (144). Truncating mutations in the BRIP1 gene have been linked to breast cancer by familial studies (145). It is also connected to subtype J of Fanconi-anemia called that is phenotypically distinct from subtype D1 (caused by BRCA2 mutations) (146).

The last confirmed gene to give rise to moderate penetrance is **PALB2**, coding for a protein that was discovered in protein precipitation complexes involving BRCA2. Like both BRCA2 and BRIP1 it has been connected to Fanconi-anemia, subtype FA-N, which is very similar to the subtype caused by biallelic BRCA2 mutations (147). This fact suggested that PALB2 was also a breast cancer susceptibly gene, something that was confirmed in subsequent studies (137).

Two common traits of all of these genes with intermediate penetrance are their involvement in DNA repair and their function in the same pathways as BRCA1 and 2. Nonetheless, they confer a much lower risk for breast cancer than those genes. The mechanisms by which these intermediate-penetrance genes increase breast cancer risk is currently unknown (88).

It has been suggested that resequencing of large numbers of cases and controls will be needed in order to identify and confirm additional, presently unknown, intermediate penetrance genes (132).

Low-Penetrance Loci

The largest fraction of familial breast cancer still remains unaccounted for, but current facts seems to favour a polygenic model, where the combination of many different low-penetrance variants together are responsible for increased familial risk (148, 149). The last five years have seen a tremendous increase in the number of common low-penetrant genetic variants uncovered by GWAS, and it seems likely that

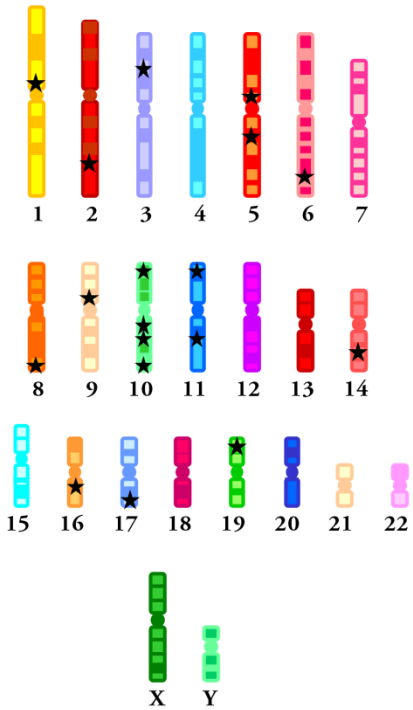
many more exist. These variants usually have a MAF of >0.10 and a RR of <1.5 (Figure 4) (28, 89). Table 1 shows an overview of 18 low-penetrant loci currently known (150), the underlying biology remains mostly unidentified but it seems clear that they are all involved in multiple complex signalling pathways(132). Descriptions of the most prominent loci follow below.

10q26 was revealed as a breast cancer susceptibility locus in two of the first large-scale breast cancer GWAS (115, 151) and has remained positively associated with breast cancer in many subsequent studies (152-156). In the primary studies two different variants (rs2981582 and rs1219648) were reported, both in intron two of the gene FGFR2. The FGFR2-protein is a tyrosine kinase receptor that is important in breast tissue development. It is overexpressed in 5-10% of human breast cancers especially in ER-positive tumours and somatic mutations that induce overactivity of the protein have been identified in tumour tissue (89, 157, 158). Fine scale mapping of the region surrounding the discovered SNPs in populations from the UK and Asia eventually led to the emergence of one SNP (rs2981578) that appears to be the causal variant behind the increased breast cancer risk observed at this locus (115, 159). It lies in a region of open chromatin conformation that is easily accessible to regulatory factors and has been found to alter the binding affinity of several transcription factors leading to increased overexpression of FGFR2 (159, 160).

16q12, harbours the second most associated low-penetrance locus. It was primarily identified by Easton et al and Stacey et al (115, 161) and the SNP showing greatest degree of association in both studies was rs3803662. A fine scale mapping of the region has been done in European, Asian, and African-American cases and controls, but failed to reveal a causative variant (162). Interestingly the variant giving the highest risk increase in European and Asian women was protective in the African-American subjects. This has been found before (161), but remains unexplained. The locus contains two genes, TOX3 (suggested to act as a transcription factor) and LOC643714 (a hypothetical gene with no known function). TOX3 was recently suggested to act as a neuronal survival factor by regulating Ca^{2+}dependent transcription and thereby enhancing cell survival, something that could be important in cancer development (163).

The locus on **5q11** has been shown to associate with breast cancer and contains several genes (MAP3K1, MGC33648, and MIER3) (164). Out of these MAP3K1 appears to be the most likely to be connected to breast cancer risk since it is part of the MAPK pathway, known to be implicated in apoptosis and cell growth (165). The

Table 1. Low-penetrance susceptibility Loci

Reference	Locus	Karyotype	Genes in region	Variant	MAF	Per allele OR (95% CI)
Easton et al 2007	10q26		FGFR2	rs2981582	0.38	1.26 (1.23-1.30)
Easton et al 2007	16q12		TOX3	rs3803662	0.25	1.20(1.16-1.24)
Easton et al 2007	5q11		MAP3K1	rs889312	0.28	1.13 (1.10-1.16)
Easton et al 2007	8q24		FAM84B/c-MYC	rs13281615	0.40	1.08 (1.05-1.11)
Easton et al 2007	11p15		LSP1	rs3817198	0.30	1.07 (1.04-1.11)
Cox et al 2007	2q33		CASP8	rs1045485	0.13	0.88 (0.84-0.92)
Stacey et al 2007	2q35		TNP1/IGFBP5…	rs13387042	0.44	1.20 (1.14,1.26)
Stacey et al 2008	5p12		MRPS30/FGFR10	rs10941617	0.26	1.19 (1.13-1.26)
Ahmed et al 2009	3p24		NEK10/SLC4A7	rs4973768	0.46	1.11 (1.08–1.13)
Ahmed et al 2009	17q23		COX11	rs6504950	0.27	0.95 (0.92–0.97)
Thomas et al 2009	1p11		NOTCH2/FCGR1B	rs11249433	0.40	1.16(1.09-1.24)
Thomas et al 2009	14q24		RAD51L1	rs999737	0.24	1.06 (1.01-1.14)
Zheng et al 2009	6q25		ESR1…	rs2046210	0.35	1.29 (1.21-1.37)
Turnbull et al 2010	9p21		CDKN2A/CDKN2B	rs1011970	0.17	1.09 (1.11-1.09)
Turnbull et al 2010	10p15		ANKRD16/FBXO18	rs2380205	0.43	0.94 (0.91-0.98)
Turnbull et al 2010	10q21		ZNF365	rs10995190	0.15	0.86 (0.82-0.91)
Turnbull et al 2010	10q22		ZMIZ1	rs704010	0.39	1.07 (1.03-1.11)
Turnbull et al 2010	11q13		MYEOV/CCND1…	rs614367	0.15	1.15 (1.10-1.20)
Antoniou et al 2010	19p13		ABHD8/ANKLE1…	rs8170	0.17	1.26 (1.17-1.35)

identified SNP (rs889312) tags an LD-block spanning 298kb and no attempt has yet been made to identify the true causal variant.

11p15 contains two genes with variants associated with breast cancer, rs3817198 in LSP1 and rs2107425 in H19. LSP1 is an F-actin binding cytoskeletal protein expressed in hematopoietic cells and H19 codes for an imprinted maternally expressed untranslated mRNA involved in regulation of the insulin growth factor gene IGF2 (115). This locus has failed to be replicated in some subsequent studies and no fine mapping of the region has been done (89, 166).

8q24 is a locus that contains no known genes, although it has been implicated as a "hot-spot" for associations with several different forms of cancer in addition to breast cancer including prostate, colorectal, ovarian, and bladder cancer (89, 167-171). The SNP associated with breast cancer (rs13281615) is located between two genes that could both be important for cancer development. FAM84B is a breast cancer membrane protein and c-MYC is a known pro-oncogene recognized for being overexpressed in different cancers (172). Even so the identified SNP does not appear to effect expression levels of either of these proteins and neither the true causal variant nor it's mode of action is known (89).

2q35 was initially identified as a breast cancer susceptibility locus by Stacey et al in a predominantly Icelandic population (161). Just as 8q24 it lies in a region with no known genes, the nearest being TNP1 (181 kb upstream), IGFBP5 (345 kb upstream), IGFBP2 (376 kb upstream), and TNS1 (761 kb downstream). In 2009 a major effort was made to evaluate this variant in larger sample sets from other European countries and it was genotyped in >30,000 cases and >35,000 controls, the difference between ER+ and ER- breast cancer was also addressed. The conclusions from this study were that 2q35 is significantly associated with breast cancer in the European population with no significant differences between cohorts or estrogen receptor status (173). The causal variant remains to be found.

The three initial GWAS all independently reported associated variants on the **5p** locus (115, 151, 161), but none of these associations reached significance levels in replication studies. Stacey et al then performed a follow up study of the region and were able to identify two variants (rs4415084 and rs10941679) that were in LD with primary detected variants and passed more stringent significance tests, both located on 5p12(174). The nearest gene is MRPS30, which has previously been implicated in pro-apoptotic events (175).

6q25 was not among the initial loci identified as associated to breast cancer in women of European descent. However, a GWAS performed on Chinese women did identify

two SNPs in this region as potentially cancer associated (rs2046210 and rs10872676). As the signal was significantly stronger for rs2046210 they continued to evaluate this variant and found it to confer susceptibility in European women as well although the association was not as strong as in China (166). A follow-up study replicated this finding in Chinese, Japanese and European women and was able to identify a putative causal variant (rs6913578) by functional genomic studies (176). The locus 6q25 contains several genes the most interesting one in relation to breast cancer being ESR1 coding for estrogen receptor α.

Only one low-penetrant locus (**2q33**) has been discovered using other methods than GWAS, and has reached significance levels comparable to other established loci with the same magnitude of association. 2q33 was revealed using a candidate gene approach evaluating CASP8, which codes for Caspase-8, a protein involved in apoptosis (90). The minor allele frequency of rs1045485, coding for a D302H substitution, is only 0.13 i.e. relatively low compared to other GWAS-identified SNPs. This fact may make it harder to detect its signal at a genome wide level (88). The contribution of this locus to genetic risk has been difficult to replicate but a meta analysis concluded that rs1045485 does reduce the risk of breast cancer in carriers of the A-allele, at least in Caucasian populations (177-179).

Many more variants have been or are being discovered (Table 1), most with unknown biological properties. The question as to how all of these findings could be put to clinical use has, however, been debated. It has been suggested that a combination of low-penetrant risk alleles could be used to distinguish women at risk on a population level (180). This has however also been refuted as adding far too little information to already well used risk prediction models based on family history, mammography screening, etc (181). The task of elucidating the context by which these different alleles affect breast cancer risk as well as how they connect to each other, lifestyle and environmental factors remains.

PRESENT INVESTIGATIONS

Aims

The overall aim of this thesis was to evaluate a novel hypothesis about the characteristics of SNPs prone to associate with cancer risk, and to assess the relationship between known common low-penetrance SNPs both on the genetic and environmental level.

Specific aims of each paper are listed below:

Explore the hypothesis that SNPs that disrupt or create a CpG site in the promoter region of a candidate gene are related to breast cancer risk (**paper I**).

Replicate findings from GWAS and evaluate the importance of multiple low-penetrant risk alleles for breast cancer risk (**paper II**).

Investigate the joint effect of, and potential interactions between, certain known genetic and phenotypic risk factors, specifically eight previously established low-penetrant variants, and height, BMI and hormone replacement therapy (HRT) (**paper III**).

Investigate possible interactions between previously identified low-penetrant variants and parity/age at first childbirth with regard to breast cancer risk (**paper IV**).

Subjects and Methods

Study Cohorts

Malmö Diet and Cancer Study

This prospective cohort study, initiated in 1991, had as a primary objective to clarify if diet was associated with certain forms of cancer and to provide a resource for future epidemiological studies. A secondary objective was analysis of the influence of oxidative stress on the impact of diet on cancer development (182).

All residents of Malmö born 1926-1945 were included in the first group invited to participate in the Malmö Diet and Cancer Study (MDCS), recruitment was by personal invitation or public advertisement. In 1995 the study was extended to include men born 1923-1945 and women born 1923-1950. Women of a younger age were included in order to be able to study breast cancer among pre-menopausal women. Participation included a first visit where measurements of height, weight, blood pressure, and body composition were performed. Blood samples were drawn and stored at -80ºC or -140ºC according to blood component. The subjects also received a questionnaire to fill out at home and were asked to keep a menu-book for seven consecutive days. At a follow-up visit, about two weeks later, the questionnaire and menu-book were collected. An interview regarding dietary habits was also conducted. All baseline examinations were performed between 1991 and 1996. The total number of participants included 28 098 individuals of whom 11063 were men and 17035 were women (182, 183).

By linkage to the national cancer registry until the 31st of December 2004, 544 prospective cases (diagnosed after enrolment and free from known breast cancer at enrolment) of invasive breast cancer were identified among participants in the MDCS. They were consecutively matched to 1088 controls according to sex, age (+/- 6 months), time of sampling at baseline (+/- 2 months), and vital status at time of diagnosis. In 2008 a new linkage to the Regional Tumour Registry ascertained 186 unique new cases diagnosed before 31st of December 2007 and these were matched to 372 new controls. All individuals with prevalent cancers were excluded (cervical cancer in situ was not defined as cancer).

Totally 730 cases and 1460 controls were selected and consecutively used in all four papers (I-IV).

Malmö Preventive Project

The Malmö Preventive Project (MPP) is a preventive case-finding programme that started in 1974 to try to identify individuals susceptible to cardiovascular risk factors, alcohol abuse and breast cancer (184). The project aimed to examine large strata of middle-aged individuals born in pre-specified years and residing in Malmö.

Subjects were invited to participate in a broad health screening programme (including a physical examination, blood tests and a self administered questionnaire on lifestyle and medical history) between 1974 and 1992. During this time 22 444 men and 10902 women were recruited and more than 40% of them have also attended a rescreening that started in 2002 and is ongoing (185, 186). The rescreening included a new questionnaire and DNA sample that is stored. Some of the same individuals that attended MDCS were also included in MPP; the overlap is estimated to be about 30%.

Among those attending the re-examination (and were non-participants in MDCS) we identified 215 prospective invasive breast cancer cases by cancer registry linkage up until 31st of December 2007 and 430 controls (Matching criteria were: sex, age (+/- 6 months), time of sampling at baseline (+/- 2 months), and vital status at time of diagnosis. All individuals with any prevalent cancer were excluded before selection (cervical cancer in situ was not defined as cancer). These cases and controls were used in papers I and II.

North Sweden Health and Disease Study

The North Sweden Health and Disease Study (NSHDS) is comprised of individuals residing in the counties of Västerbotten and Norrbotten in northern Sweden. It includes three subcohorts; the Västerbotten Intervention Program (VIP), the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) and the Mammography Screening Project (MSP) (186).

The VIP is an ongoing preventive program for cardiovascular disease (CVD) and type-2 diabetes that was initiated in 1985. All individuals residing in the county of Västerbotten are invited for a primary health screening at 40, 50 and 60 years of age. They are also asked to donate a blood sample for research purposes. After 10 years a follow up sample is collected. Currently, there are more than 83,000 blood samples from 70,000 participants in the VIP (186-189).

The North Sweden MONICA project was initiated in 1985 as a part of the WHO MONICA (started in 1982). Its main purpose was to assess if mortality in different

countries could be explained by differences in the population load of conventional cardiovascular risk factors (190, 191). It has been conducted as cross-sectional surveys in 1986, 1990, 1994, 1999, 2004, and 2009. The participants were randomly selected, stratified for sex and divided into 10-year age-groups. The visit included an examination for cardiovascular risk factors and a questionnaire. Around 90% of participants also agreed to donate a blood sample for research purposes and the cohort contains blood samples from more than 9000 individuals (192).

The MSP has been ongoing since 1995 and invites all women in Västerbotten between 50 and 70 years of age to undergo regular mammography screening every 2-3 years. At the same time they are asked to donate a blood sample. This subcohort now contains 48,000 samples from 27,500 women (186).

From the NHSDS 1680 prospectively occurring invasive breast cancer cases were identified through linkage to the national cancer registry up until 31st December 2008 and matched to 2369 controls for sex, age (+/- 6 months), time of sampling at baseline (+/- 2 months), and controls were alive and free from all malignant cancers prior to their respective cases diagnosis date. All individuals with malignant prevalent cancers were excluded prior to selection. Samples from the NSHDS were used in papers I-III.

Icelandic Cohort

The Icelandic samples used in this thesis were collected between 1998 and 2006 and represents 45-77% of all Icelandic women with invasive breast cancer diagnosed between 1957 and 2007. The rate of participation varied somewhat depending on the year of diagnosis and was highest between 1999 and 2003 (77%). Unmatched controls were collected between 2000 and 2004, either from women who participated in the population-based cervical or breast cancer screening program and were found free of breast cancer or from older women in retirement homes who had not been diagnosed with breast cancer, to generally reflect the ages of the cases. By linkage to the Icelandic cancer registry in 2008 we identified cases diagnosed before 31st of December 2007. Totally 866 cases (collected controls that had developed breast cancer before registry linkage was included as cases in this group) and 948 controls had DNA available and were eligible to us. These samples were used in papers I and II.

Polish Cohort

Cases with early onset or familial breast cancer, free from BRCA1/2 mutations, were recruited at the genetic counselling clinic in Silesia between 1997 and 2006. Samples

from unmatched controls were collected between 2003 and 2009 from healthy women attending the same clinic, but who had no or sparse family history of breast cancer. Totally 391 cases and 306 controls had DNA available for use in our studies. These samples were included in papers I and II.

Genotyping

All samples were genotyped using either the Sequenom MassARRAY® system (MALDI-TOF Mass Spectrometer) or TaqMan® SNP genotyping assays. Both systems are described below:

The Sequenom MassARRAY® system is based on a mass spectrometry platform. It measures the mass of individual DNA-amplicons and can discriminate between masses differing by only a few Da. First the sequence containing the SNP is amplified in a mutiplex PCR reaction to generate abundant specific amplicons. After this all unincorporated dNTPs are neutralised using Shrimp Alkaline Phosphatase (SAP). A locus specific primer extension reaction then follows in which an oligonucleotide primer anneals immediately upstream of the SNP site being genotyped. The primer and target are incubated with mass-modified dideoxynucleotide terminators that extend the sequence by a single base after which the extension is terminated. The mass of the extended primer is then determined using MALDI.-TOF mass spectrometry. The results are depicted in a spectrum (Figure 9) (193). This method was used for samples in all four papers.

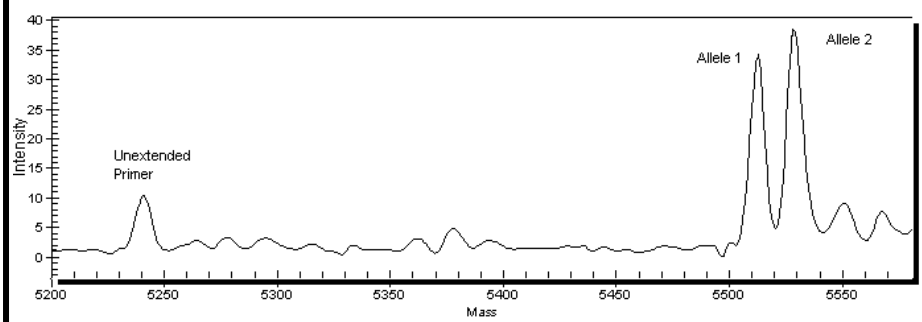


Figure 9. Spectrum of A and G terminated extension products (alleles 1 and 2). The figure also shows the peak of the unextended primer. *Reproduced with permission from Sequenom®.*

The TaqMan® genotyping assays take advantage of the intrinsic exonuclease activity of the Taq-polymerase. The assay contains one forward and reverse primer as well as two probes (specific for either SNP variant) labelled with different fluorophores and incorporating the minor grove binder (MGB) technology on the 3' end. Each specific probe attaches to its corresponding DNA sequence between the primers. When the Taq-polymerase reaches the start of the probe the exonuclease destroys the probe and releases the fluorophore from the quencher hereby making the fluorescence detectable by the instrument. The ratio between the different light emissions is used to determine the correct genotype of the sample (Figure 10) (194, 195). This method was used for samples not included in the Sequenom MassARRAY® multiplexes. It was used in papers II, III and IV.

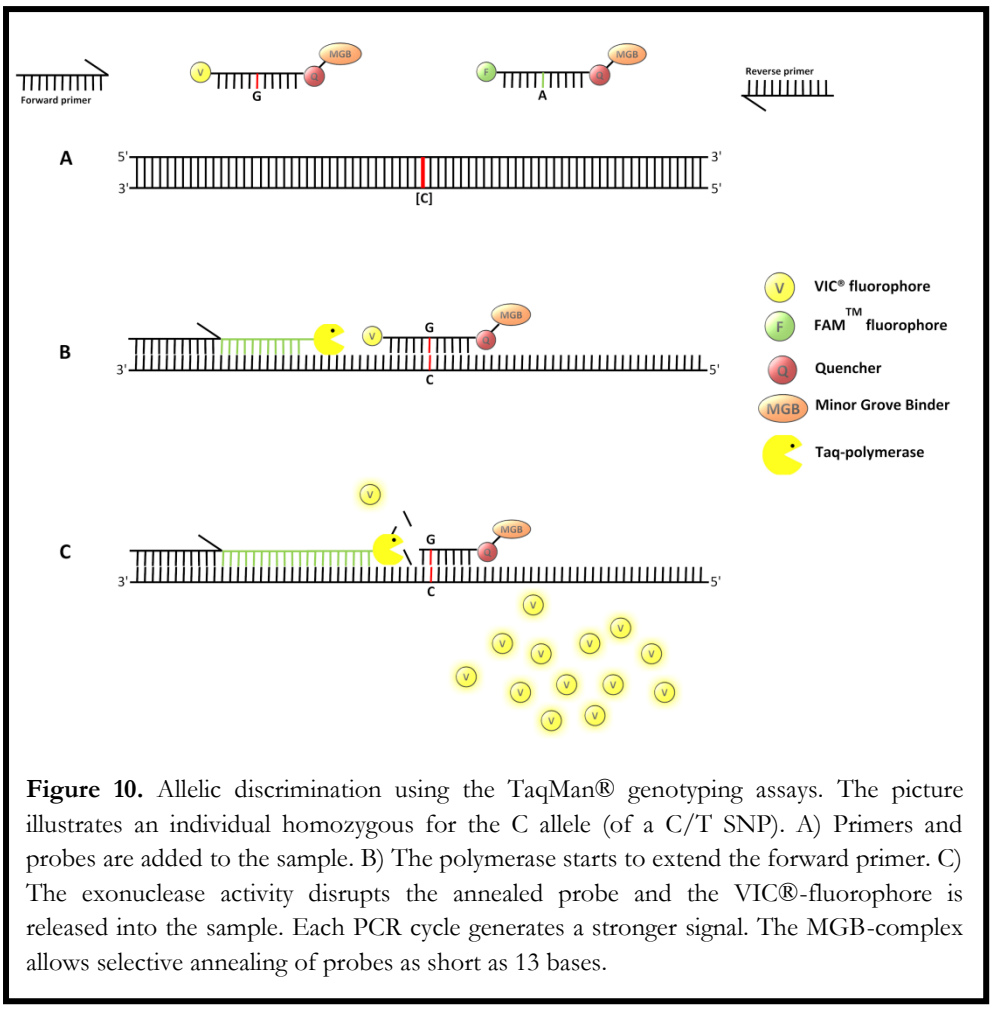


Figure 10. Allelic discrimination using the TaqMan® genotyping assays. The picture illustrates an individual homozygous for the C allele (of a C/T SNP). A) Primers and probes are added to the sample. B) The polymerase starts to extend the forward primer. C) The exonuclease activity disrupts the annealed probe and the VIC®-fluorophore is released into the sample. Each PCR cycle generates a stronger signal. The MGB-complex allows selective annealing of probes as short as 13 bases.

Statistics

All calculations were performed using SPSS statistics version 16.0 or 17.0. In all papers control samples were tested for consistency with Hardy-Weinberg equilibrium (HWE) for each SNP. Unconditional logistic regression models were used to measure the association between SNPs and the risk for breast cancer using homozygotes for the common allele (AA) as reference. For each SNP, ORs and 95% CIs were calculated for each genotype and unconditional per-allele ORs (p-trend) were calculated using 0, 1 or 2 copies of the minor allele (a) as a continuous variable. P-values of <0.05 were considered statistically significant. When appropriate the significance threshold was adjusted using the Bonferroni correction method by dividing the p-value with number of comparisons.

In paper I ORs were adjusted for age and individuals were stratified into two age-groups corresponding to menopausal age ($\leq$ 50 vs > 50). In order to integrate the different backgrounds of the cohorts all results were also adjusted using the Mantel-Haenszel method.

In paper II samples were stratified according to age ($\leq$ 50 vs > 50 years) and cohort. As two of the SNPs presented with very different OR point estimates in the respective age-groups heterogeneity between the groups was examined using adjusted case-case models in unconditional logistic regression analysis. The p-value for inhomogeneity (p_{ih}) of OR between cohorts was calculated using the Breslow-Day test.

In paper III OR and 95% CI were calculated between each phenotypic variable (height, BMI and HRT) and risk for breast cancer, these results were also age adjusted. Data were stratified into tertiles according to height (<162 cm, 162-166 cm and >166 cm), and into subcategories of BMI according to the WHO guidelines (Normal weight: 18.5-25, Overweight: 25-30 and Obese >30). For HRT the data were stratified according to reported "non use" and "current use". The current users were further divided into users of only estrogen or combined hormones. OR and 95% CI were calculated for each variable (height, BMI and HRT) and risk for breast cancer. An interaction term was introduced in the logistic regression model in order to assess any potential interactions between each pair of genotype/phenotypic factor.

In paper IV the potential relationship between previously identified SNPs and parity/age at first childbirth were assessed. Cases and controls from the MDCS cohort were compared with regard to established and potential risk factors for breast cancer in order to identify possible confounders. All analyses were subsequently adjusted for matching criteria, age, year of inclusion into the study, and potential

confounders. Only two factors, HRT and type of occupation, were identified as confounders i.e. factors with >5% difference in prevalence between cases and controls. Analyses were stratified for parity and age at first childbirth. The material was also stratified according to genotype and the breast cancer risk associated with parity and increasing age at first birth was calculated. These associations were reported using continuous analysis. As in paper III an interaction term was introduced in the logistic regression model in order to access any potential interactions between each pair of genotype and parity/age at first childbirth.

Results and Discussion

Main findings and discussions of each paper are summarized below:

Paper I

In paper I we examined the potential association of SNPs that destroyed or created a CpG site (preferably in the promoter region of selected genes) with breast cancer risk. First we conducted a screening study (testing 173 SNPs) on samples from the MDCS that were diagnosed before December 31st 2004. Based on results from the screening we selected 19 out of the 173 SNPs to be further analysed in a verification panel on the basis of at least borderline significance and/or OR point estimates higher than 2.5.

The verification panel consisted of samples from: MDCS (diagnosed between 31st December 2004 and 31st December 2007), MPP, NSHDS, Iceland, and Poland. Totally 7434 samples (3211 cases and 4223 controls) were included. When the samples were analyzed for consistency with HWE four SNPs failed this test and were excluded. The 15 remaining SNPs were tested for potential associations with breast cancer.

One SNP (rs7766585), situated in an intronic region of the ESR1 gene, was associated with increased breast cancer risk in heterozygote carriers compared to homozygote major allele carriers. With an unadjusted point estimate OR (95% CI) of 1.30 (1.17-1.45) ($p = 2.1x10^{-6}$), the significance approached levels comparable to cut-offs used in GWAS. The point estimates varied somewhat between the cohorts with significant associations for northern Sweden, Iceland and Poland but non-significant results for southern Sweden (MDCS, MPP). The significance was somewhat increased after we applied Mantel-Haenszel statistics to control for cohort. Another SNP in ESR1 (rs851987) was weakly associated with breast cancer risk in homozygote minor allele carriers vs. homozygote major allele carriers (p-trend = 0.03).

Using this candidate CpG SNP approach to search for SNPs affecting breast cancer risk we were able to identify at least one interesting association between rs7766585 and breast cancer risk. The SNP is situated in intron six of ESR1 and belongs to a linkage group consisting of more than 27 SNPs in the same vicinity. The loci comprising ESR1 (6q25) has previously been implicated as breast cancer associated (166) and a large SNP-tagging study focusing on ESR1 has been performed (196), that discovered one SNP in intron four (rs3020314) weakly enhancing cancer risk

(OR 1.05 95% CI 1.02-1.09). They tagged SNPs in linkage with rs7766585 but none of those SNPs passed the initial screening and were not examined further.

One other ESR1 SNP (rs851987) among the 15 stood out as interesting. The T allele tended towards protection in the screening phase (unadjusted per allele OR 0.92 95% CI 0.79-1.08), and the results were similar in the verification phase (unadjusted per allele OR 0.92 95% CI 0.85-0.99). The P-value (0.03) did not pass the Bonferroni-correction threshold when we adjusted for multiple comparisons (15 SNPs), but rs851987 nonetheless remained interesting because of its location 3.7kb 5' of the ESR1 transcriptional initiation site and the fact that it is a CpG altering SNP and a potential binding site for the MeCP2 protein. The potential association of this SNP with breast cancer should be verified in a much larger cohort.

Our study design incorporates cases and controls from five different cohorts and involves both strengths and weaknesses. The Swedish study-bases (MDCS and NSHDS) have matched controls and are prospective population based cohorts. They all involve women recruited mainly at quite a late age (>40 years), even though there are exceptions, and the exclusion of prevalent cases removes early breast cancer from the majority of these populations. In the Iceland cohort mainly prevalent cases of breast cancer were included since sample collection occurred long after initiation of case acquisition. This may have resulted in an exclusion of lethal cases, and older women with other causes of death from the earliest recruitment period. A similar bias is present in the MPP cohort despite prospective population-based design, as DNA samples were acquired at a delayed follow-up. It is therefore possible that these two study-populations are biased towards breast cancer cases with a more favourable outcome. The Polish cases were recruited from families with multiple breast cancer cases, or because of early onset of breast cancer. In the Polish cohort we might therefore expect the presence of other highly penetrant risk factors to overshadow the mild effect anticipated within our methylation hypothesis.

Through this approach we were able to identify a new possible risk SNP, in the ESR1 gene, which seems to be associated with increased breast cancer risk. The results indicate that our model, specifically designed to identify genetic risk that may interact with lifestyle, may provide a useful complement in the search for new risk alleles both in breast cancer and other forms of cancer.

Paper II

In paper II we tested the effect of having several low-penetrant alleles on breast cancer risk and aimed to replicate twelve previous GWAS findings (Table 2) in our

cohorts (same as in paper I). We also compared differences in results with reference to study population and design with three other published studies (154, 181, 197).

Associations between eight of the twelve previously reported SNPs and breast cancer were confirmed in our material, with age adjusted ORs for these SNPs in close proximity to ORs previously described (90, 115, 161). P-trend values for five of the SNPs (rs2981582, rs3803662, rs889312, rs12443621 and, rs13281615) were <0.001 and for the remaining three SNPs (rs13387042, rs3817198 and, rs981782) <0.01.

Two of the SNPs (rs8051542 and rs30099) exhibited age adjusted ORs near those reported (115) but did not pass the significance threshold of 0.05. Associations of the two last remaining SNPs with breast cancer could not be reproduced (rs1045485 and rs4666451).

Stratification of participants into age groups ($\leq$50 vs. >50 years), discovered age specific associations for two of the SNPs (rs30099 and rs981782), one of which turned out to be statistically significant (rs9817183, p-value = 6.2×10^{-4}). Stratification of results according to cohort revealed similar effects for most SNPs, although rs13387042 was most strongly associated with risk in the Icelandic cohort, correlating with the fact that it was originally discovered in an Icelandic population setting (161). Inhomogeneity between cohorts was also greatest for this SNP (p_{ih} = 0.02).

Both cases and controls were stratified according to individual burden of the eight alleles found significantly associated with breast cancer risk in this study (rs2981582, rs3803662, rs889312, rs12443621, rs13387042, rs13281615, rs3817198 and rs981782). A successive increase in point estimate from an OR of 1 for $\leq$ 3 alleles to an OR of 2.33 (95% CI 1.88-2.90) for carriers of $\geq$ 10 risk alleles was detected (overall p for trend = 9.3×10^{-26}).

The total effect of having multiple low risk alleles has been investigated before by Wacholder et al (181) and Reeves et al (197). Wacholder utilizes a simple additive model like our own whereas Reeves makes use of a more complex approach involving grouping cases according to quintile of a polygenic risk score and according to number of risk alleles. Both come to the conclusion that even though the total risk is significantly affected by having multiple low risk alleles this is not useful as a predictive approach in a clinical setting. A third study of Hemminki et al (154) has investigated the effect of some of the same GWAS SNPs in a cohort consisting of familial cases lacking BRCA1/2 mutations. They found that the impact of at least two of the SNPs, located in FGFR2 and TOX3 respectively, seem to yield higher OR

Table 2. Low-penetrance SNPs investigated in paper II

Reference	Locus	Gene	Variant	Per allele OR (95% CI) Reference study	Per allele OR (95% CI) Our study*
Easton et al 2007	10q26	FGFR2	rs2981582	1.26 (1.23-1.30)	1.25 (1.17-1.33)
Easton et al 2007	16q12	TOX3	rs3803662	1.20 (1.16-1.24)	1.18 (1.11-1.27)
Easton et al 2007	16q12	TOX3	rs12443621	1.11 (1.08-1.14)	1.13 (1.06-1.20)
Easton et al 2007	16q12	TOX3	rs8051542	1.09 (1.06-1.13)	1.06 (1.00-1.13)
Easton et al 2007	5q11	MAP3K1	rs889312	1.13 (1.10-1.16)	1.13 (1.06-1.21)
Easton et al 2007	8q24	-	rs13281615	1.08 (1.05-1.11)	1.13 (1.07-1.21)
Easton et al 2007	11p15	LSP1	rs3817198	1.07 (1.04-1.11)	1.10 (1.03-1.17)
Easton et al 2007	5p12	HCN1	rs981782	0.96 (0.93-0.99)	0.90 (0.84-0.96)
Easton et al 2007	2p24	-	rs4666451	0.97 (0.94-1.00)	1.01 (0.95-1.08)
Easton et al 2007	5q11	-	rs30099	1.05 (1.01-1.10)	1.05 (0.94-1.18)
Stacey et al 2007	2q35	-	rs13387042	1.20 (1.14-1.26)	1.10 (1.04-1.17)
Cox et al 2007	2q33	CASP8	rs1045485	0.88 (0.84-0.92)	0.97 (0.88-1.06)

* Consisting of totally 8647 samples (3584 cases and 5063 controls)

point estimates in familial breast cancer cases contributing even more to the Population Attributable Risk (PAR) than moderate-penetrance variants in CHEK2, ATM, BRIP1 or PALB2.

FGFR2 and TOX3 SNPs have consistently been replicated in all published reports and efforts have been made to uncover the causal variants behind the risk increases. Fine mapping of the region surrounding rs2981582 in FGFR2 using women of African-American descent has identified a possible causal variant, rs2981578, that is situated in an open chromatin configuration and has been found to alter the binding affinity of two different transcription factors (Oct1/Runx2 and C/EBPβ) leading to increased expression of FGFR2 (159, 160). Further mapping of the TOX3 locus in European, Asian, and African-American cases and controls has however failed to reveal the causative variant. A recent report of Hazard Ratios for FGFR2, TOX3, MAPK13 and LSP1 SNPs in BRCA1/2 mutation carriers (198) suggests little additional risk in BRCA1 mutation carriers, but independent additional risk for BRCA2 mutation carriers, comparable to reported risks in non-selected populations.

The intergenic SNP rs981782 on 5p12, a region previously yielding significant SNPs for breast and other cancers (89), was one of the 3 SNPs of secondary significance in the Easton study. We found that the protective effect of the minor allele was notably more pronounced in premenopausal breast cancer (women $\leq$50 years). The p-value (6.2×10^{-4}) for heterogeneity was significant indicating that there is a genuine

difference between age groups, although previous analysis performed by other methods found no significant difference (115, 161). In a fine mapping of the region, Stacey et al (174) identified two SNPs in the same region (rs4415084 and rs10941679) as the possible causal variants behind this association.

SNP rs13387042, originally reported by Stacey et al (161) was identified in a screening panel containing 1600 Icelandic women and verified in a large, predominantly Icelandic panel of 4554 cases and 17577 controls. Our results for the Swedish and Polish cohorts clearly differ from the Icelandic population, perhaps included in both studies, whose carriers of the rs13387042 A-allele demonstrate an increased risk, although this SNP has been significant in other populations (173). It is possible that this "Icelandic SNP" represents a common ancestral haplotype.

The strengths and weaknesses of our five population cohorts have been discussed in paper I (above). Methodological strengths with this specific study included the exclusion of samples with <80% successful genotypes and by 100% concordant genotypes in 270 duplicate samples. Although the use of $p<0.05$ as significance limit is appropriate for this replication study verifying reported associations, the occurrence of false negative findings cannot be excluded. Lack of significance, in particular of the CASP8 (rs1045485) association, might be attributable to insufficient statistical power.

Our findings, including total risk score, are similar but not identical to other replication studies. While we agree with other authors that routine clinical use of all the low penetrant risk alleles thus far identified by GWA studies adds little to routine risk assessment based on phenotypic information, the consistent significant findings for FGFR2 and TOX3 SNPs could motivate their inclusion in screening panels for members of multiple-case families.

Paper III

In paper III we focused on investigating potential interactions between initially thirteen (later reduced to eight) common variants and height, BMI, and HRT (Table 3). We used samples (2063 cases and 3613 controls) from MDCS and NSHDS, described above.

Nine of the thirteen SNPs were independently significant ($p <0.05$) in our material and were primarily selected to be analysed further. One SNP (rs851987 in ESR1) exhibited borderline significance ($p = 0.07$) which was also deemed low enough for further analysis. The three SNPs in TOX3 (rs3803662, rs12443621 and rs8051542)

Table 3. Low-penetrance SNPs originally included in paper III.

Reference	Locus	Gene	Variant	Per allele OR (95% CI) Participating cohorts*	Included for interactions analysis
Easton et al 2007	10q26	FGFR2	rs2981582	1.27 (1.17.1.37)	Yes
Easton et al 2007	16q12	TOX3	rs3803662	1.19 (1.09-1.29)	Yes
Easton et al 2007	16q12	TOX3	rs12443621	1.15 (1.07-1.24)	No (Linked to rs3803662)
Easton et al 2007	16q12	TOX3	rs8051542	1.09 (1.01-1.18)	No (Linked to rs3803662)
Easton et al 2007	5q11	MAP3K1	rs889312	1.16 (1.07-1.27)	Yes
Easton et al 2007	8q24	-	rs13281615	1.13 (1.05-1.22)	Yes
Easton et al 2007	11p15	LSP1	rs3817198	1.12 (1.03-1.21)	Yes
Easton et al 2007	5p12	HCN1	rs981782	0.92 (0.85-0.99)	Yes
Easton et al 2007	5q11	-	rs30099	0.95 (0.82-1.10)	No (Low significance)
Stacey et al 2007	2q35	-	rs13387042	1.06 (0.98-1.14)	No (Low significance)
Cox et al 2007	2q33	CASP8	rs1045485	0.91 (0.80-1.03)	No (Low significance)
Paper I	6q25	ESR1	rs7766585	1.14 (1.02-1.27)	Yes
Paper I	6q25	ESR1	rs851987	0.93 (0.85-1.01)	Yes (Borderline significance)

* MDCS and NSHDS

exhibited linkage when adjusted for each other and significance remained for only rs3803662 something that has been previously reported (115, 197). Hence rs12443621 and rs80515442 were excluded from further analysis. After these initial exclusions eight SNPs remained for which interactions with above phenotypes were examined (Table 3).

Each phenotypic variable was also analysed independently within the study population. These results indicated an increased risk for individuals ≥162 cm, something that was weakened following age adjustment. No correlation between BMI and risk was found in this population. For current use vs. non-use of HRT, a significantly increased risk was seen for users, OR (95% CI) 1.23 (1.07-1.42), which remained after adjustment for age.

After stratification for height, one SNP (rs851987) in ESR1 had a p-interaction = 0.01 with height, with a seemingly increasing protective effect of the major allele in taller women, but it did not pass the threshold for multiple comparisons.

None of the SNPs showed any tendency towards significant interactions after stratification according to BMI. Following stratification of genotypes according to reported current use or non-use of hormone replacement therapy, rs13281615 (8q24) was significant only in non users and rs3817198 (LSP1) was significant only in current

users of HRT with a p for interaction of 0.05 in both cases, neither interaction passed the threshold for multiple comparisons ($p = 0.002$).

The most significant finding was that between height and one of the SNPs discovered by us in paper I (rs851987). Taller women homozygous for the T-allele appeared to have reduced breast cancer risk. Rs851987 is situated in the far end of the extended promoter region of ESR1, about 3.7kb 5' of exon F. Exon F and its promoter were originally described by Thompson et al (199) and have later been shown to affect the level of expression in osteoblastic cells (200, 201). A potential association between ESR1 and height has been described in another study comprising adult males from two Swedish population cohorts (60). Mutations in the estrogen-receptor alpha gene have been reported to delay fusion of the epiphyseal plates at puberty (58), and one may speculate that our SNP either participates in this biological effect or is linked to other causal variants.

Since the first GWAS on breast cancer became available in 2007 several replication and interaction studies of varying sizes have been published (153-155, 202). Recently before we finished this study another a large interactions study comprising 7610 breast cancer cases from the Million Women Study in UK was released in which potential interactions between 12 different SNPs and 10 different variables (including height, BMI and HRT) had been tested (110). That study did not find any significant gene-environment interactions. Our study originally included ten of the same polymorphisms as in the Million Women Study (excluding rs1982073 in TGFB1 and rs1800054 in ATM), but also included one additional SNP from Easton et al (115) and two additional SNPs from our own candidate CpG study (paper I) (rs7766585 and rs851987 both in ESR1). Although our material was not as large, our study was comprised of two well described cohorts that were prospectively followed for breast cancer incidence using the comprehensive, population-based Swedish Cancer Registry (203). A limitation of our study was the fact that HRT was reported only once (at recruitment) without information about duration. We also lacked information about other risk factors (aside from age, height, BMI, HRT) and therefore could not adjust our results for potential confounders.

Totally, no significant interactions could be demonstrated after Bonferroni correction for multiple testing although some of our findings could be worthy of further investigation.

Paper IV

In paper IV we examined possible interactions between 14 common variants (Table 4) and the two reproductive factors parity and age at first childbirth. As data on these environmental risk factors were available only for participants in the MDCS cohort, this was the only study-base investigated in the paper.

Seven of the 14 tested SNPs were statistically significantly associated with the risk of breast cancer in the per allele analysis.

Women with one child and heterozygous for rs8051542 (in TOX3) had a statistically significantly increased risk for breast cancer compared to women homozygous for the major allele.

For women with two children, both hetero- and homozygote carriers of the minor allele for rs2981582 (FGFR2) had an increased risk of breast cancer compared to homozygous carries of the major allele. There was an inverse association for breast cancer in homozygote minor allele carriers of rs981782.

Table 4. Low-penetrance SNPs included in paper IV.

Reference	Locus	Gene	Variant	Per allele OR (95% CI) MDCS	Included in previous papers
Easton et al 2007	10q26	FGFR2	rs2981582	1.28 (1.12-1.46)	Yes
Easton et al 2007	16q12	TOX3	rs3803662	1.21 (1.05-1.40)	Yes
Easton et al 2007	16q12	TOX3	rs12443621	1.18 (1.04-1.34)	Yes
Easton et al 2007	16q12	TOX3	rs8051542	1.13 (1.00-1.29)	Yes
Easton et al 2007	5q11	MAP3K1	rs889312	1.19 (1.04-1.37)	Yes
Easton et al 2007	8q24	-	rs13281615	1.12 (0.98-1.28)	Yes
Easton et al 2007	11p15	LSP1	rs3817198	1.18 (1.02-1.36)	Yes
Easton et al 2007	11p15	H19	rs2107425	0.86 (0.74-0.99)	**No**
Easton et al 2007	5p12	HCN1	rs981782	0.89 (0.77-1.02)	Yes
Easton et al 2007	2p24	-	rs4666451	1.03 (0.91-1.18)	Yes
Easton et al 2007	5q11	-	rs30099	0.98 (0.79-1.21)	Yes
Stacey et al 2007	2q35	-	rs13387042	0.90 (0.79-1.03)	Yes
Cox et al 2007	2q33	CASP8	rs1045485	1.10 (0.79-1.51)	Yes
Paper I	6q25	ESR1	rs7766585	1.03 (0.86-1.23)	Yes

For women with three or more children, there were some statistically significant interactions observed in TOX3, LSP1 and H19 SNPs. In the interactions analysis however, no statistically significant interactions between parity and the different

SNPs could be confirmed. There was also no statistically significant trend over parity when stratifying on different alleles.

When the women were stratified according to age at first childbirth some statistical significant interactions were observed in the younger age groups (<20 years and > 20 <25 years). Notably in these groups risk was increased for minor allele carriers compared to major allele carriers of rs3817198 in LSP1 and rs889212 in MAP3K1.

There was also a statistically significant interaction ($p = 0.04$) between age at first childbirth and rs2107425 (H19), where the relative risk associated with both heterozygote and homozygote minor allele carriers was increased compared to major allele carriers among women in the older age groups, and decreased in the younger age groups. This result did not pass the threshold for multiple comparisons ($p = 0.0025$).

In this paper samples with <80% genotyping data were initially included in all analysis. A sensitivity analysis was later performed excluding theses samples and came to roughly the same conclusions, though some analysis with borderline significant ORs became significant when only individuals with information on $\geq$ 80% of all SNPs were analysed.

About 40% of the women invited to participate in the MDCS actually participated, and they have previously been shown to be of a higher socioeconomic status with a higher incidence of breast cancer than the general population (183). However, as this study used internal comparisons, yielding relative risks rather than incidence rates, the impact of a potential selection bias was probably limited.

Parity and age at first childbirth were the main exposures of this study and were obtained from questionnaires answered at baseline. No reliable information was available on number of miscarriages or abortions. However, a previous large meta-analysis has found no association between breast cancer risk and previous abortion (204). All women were 44 years or older at baseline, hence unlikely to have given birth to more children thereafter.

As many comparisons were made there is a potential risk of false positives. As these analyses are made with an *a priori* hypothesises, the Bonferroni correction for multiple testing was not considered relevant. Concerning interaction analyses and the stratified analyses, theses analyses were exploratory and hypothesis generating, and statistically significant findings will have to be confirmed in repeated analyses. Moreover, using corrections for multiple comparisons, there were no statistically significant interactions.

This study is small in comparison to the other three papers and the risk of false negatives is high due to very few individuals in some analyses generating wide confidence intervals.

At least two previous studies have evaluated the possible connection between breast cancer risk and parity/age at first childbirth (110, 205). One very large study comprising women from the UK and described in paper III did not find any interactions between reproductive factors and common variants (110). A small Japanese study by Kawase et al (205) found a statistically significant interaction between parity and rs2981582 (FGR2), the size of this study is more comparable to our own (456 cases and 912 controls) but only one SNP was examined.

In this study we found one potential interaction between rs2107425 (H19) and age at first childbirth, but this result would need to be confirmed in independent cohorts.

CONCLUSIONS AND FUTURE PERSPECTIVES

The present knowledge about familial breast cancer and the causative genes and loci is growing with incredible speed. Recent advances have led to the identification of numerous common variants, most without any clear biological function. The stage is set for success in elucidating much of the previously unexplained familial clustering and for understanding more about breast cancer pathology by examining the functions of these new discoveries.

In this thesis we tested a previously unexplored hypothesis concerning CpG altering SNPs and breast cancer risk, we came to the conclusion that even though our results were not extensive we were able to identify at least one new potential risk SNP by this method and if it were to be used in a wider context, e.g. by scanning the whole genome for CpG SNPs, the results might prove rewarding.

We also examined the effect of having multiple established low-penetrant risk alleles on breast cancer risk and found that the effect was substantial, with a statistically significant increasing risk-trend with increasing number of risk alleles. A reliable clinical use of this information remains unclear. Other studies have also examined this relationship and come to similar conclusions. The general opinion seems to be that care should be taken before incorporating any of these results into clinical practise and that not much would be gained by using these variants in population based screening programs. While we agree with this concept, the consistent replication of the FGFR2 and TOX3 loci, with increasing significance in multiple case families suggests that inclusion of variants from these two loci could very well add valuable information to risk assessments in such high risk families.

Potential interactions between low-penetrant loci and environmental factors have previously been thought to clarify a large part of the unexplained familial risk observed. Unfortunately many recent studies have been unable to confirm any strong interactions between the most well-known low-penetrance loci and environmental factors connected to breast cancer risk. We examined potential gene environment interactions for height, BMI and HRT (paper III) and for reproductive factors (parity/age at first childbirth) (paper IV), but also failed to find any strong and consistent results. Perhaps the unexplained familial cases cannot be attributed to interactions between genotype at these loci and the environment. The explanations may be much more complex than previously thought. The incorporation of epigenetic factors to explain familial breast cancer risk is an area that remains unexplored and could possibly help us increase our understanding.

Taken together the work in this thesis contributed to our knowledge of common genetic variants associated with low-penetrant breast cancer risk. We will hopefully see many more studies in the near future that will help us integrate knowledge from different fields in order to eventually clarify all of the factors behind familial breast cancer.

ACKNOWLEDGEMENTS

Many people have contributed to this thesis and deserve my sincere gratitude. I would especially like to thank:

My supervisors:

Joyce Carlson, thank you so much for taking me on and sharing your enthusiasm and project plans with me, for (almost) always answering your phone even at odd hours and for spending your vacations revising my papers.

Joakim Dillner, thank you for introducing me to your group, for always being positive despite drawbacks, for believing in my projects and for letting me chose my own path.

Jonas Manjer, thank you for saving me from being swept away by a statistical tangle, for helping me structure up my projects to actually fit into a thesis and for answering all my questions on epidemiology.

My colleagues and co-workers:

Malin Ivarsson, my mentor, co-author, instructor, and friend! Thank you for always listening patiently to all my problems, for teaching me everything about DNA extraction and genotyping and for becoming one of my best friends!

Helena Persson and **Anna Olofsson Franzoia,** for providing invaluable help with practical details and administration.

Maria Sterner and Liselotte Hall, for helping out with everything surrounding the Masspek, and for always being friendly and helpful.

Anders Dahlin, thank you for selecting my cases and controls and for being so helpful with everything from scanning documents to answering questions.

Johanna Ekström, my roommate and "sparring partner", constantly available to answer questions or contribute with interesting discussions and thoughts. Thank you for making my final time here so enjoyable and fun.

Maria Anderberg, thank you for all our great (and unhealthy) lunches.

Salma Butt, thank you for trying to figure out the answers to all my questions regarding SPSS and statistics and for always seeming so happy about doing it.

Everyone in the Dillner group who has helped create the nice, friendly and warm atmosphere that makes it fun to go to work every day. Especially: **Anna, Aline, Agustín, Carina, Cecilia, Christina, Davit, Hanna, Helena, Herman, Janka, Kia, Kristin, Natasa, Tina, Vesna and Zoltan**

My paper co-authors: **Jorunn E Eyfjörd, Asta Försti, Kari Hemminki, Per Lenner, Ewa Grzybowska**, for valuable comments and contributions to my papers.

My family:

My brother-in-law, **Björn**, for always helping out when chaos strikes, and being only 10 minutes away. You have saved my life more times than I can count!

My Sisters, **Jenny** and **Emmy**, what would I do without you? Thank you for being the best sisters in the world, always supporting me and for being great travel-partners!

Mamma och **Mormor,** kvinnorna av stål. Tack för att ni alltid ställt upp när det behövts. Tack för alla jular och midsomrar, semestrar och helger som ni ordnat. Tack för att ni lärt mig att kvinnorna i vår släkt är fantastiska.

My Grandmother **Ina** and my Father **Ralph**, thank you for passing on the "never satisfied" gene to me. Most likely the reason behind why I am here today is because I have always aimed for something higher and probably always will.

My Children, **Elvira** and **Arvid,** for making me understand what's important in life, for making it fun to come home from work every day (and fun returning in the morning). For making me feel like the best mom in the world despite all my shortcomings, I love you both so much.

My husband and best friend, **Peter**, thank you for all the cleaning, vacuuming, washing, folding, mopping and child caring you've done, I appreciate it more than you know. Thank you for the "last minute" revision on this thesis and most of all, thank you for being part of my life and for making "old age" seem like something to look forward to.

REFERENCES

1. Chen, Y.C. and D.J. Hunter, *Molecular epidemiology of cancer.* CA Cancer J Clin, 2005. **55**(1): p. 45-54; quiz 57.
2. Schulte, P. and F.P. Perera, eds. *Molecular Epidemiology: Principles and Practices*. 1993, Academic Press.
3. Ambrosone, C.B. and C.C. Harris, *The development of molecular epidemiology to elucidate cancer risk and prognosis: a historical perspective.* Int J Mol Epidemiol Genet, 2010. **1**(2): p. 84-91.
4. Weinberg, R., A., *The biology of cancer*. 2007: Garland Science, Taylor and Francis group.
5. Diamandopoulos, G.T., *Cancer: an historical perspective.* Anticancer Res, 1996. **16**(4A): p. 1595-602.
6. Rayter, Z. and J. Mansi, eds. *Medical therapy of breast cancer*. 2003, Cambridge University Press.
7. WHO. *Breast Cancer: prevention and control.* 2011; Available from: http://www.who.int/cancer/detection/breastcancer/en/index.html.
8. Jemal, A., et al., *Global patterns of cancer incidence and mortality rates and trends.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(8): p. 1893-907.
9. Key, T.J., P.K. Verkasalo, and E. Banks, *Epidemiology of breast cancer.* Lancet Oncol, 2001. **2**(3): p. 133-40.
10. Wellings, S.R., *Development of human breast cancer.* Adv Cancer Res, 1980. **31**: p. 287-314.
11. Svenska-Bröstcancergruppen. *Nationella riktlinjer för behandling av bröstcancer*. 2011 [2011-03-19]; Available from: http://www.swebcg.se.
12. Simpson, P.T., et al., *Molecular evolution of breast cancer.* J Pathol, 2005. **205**(2): p. 248-54.
13. Tavassoli, F., A. and P. Devilee, eds. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Breast and Female Organs*. 2003, IARC Press: Lyon.
14. Elston, C.W. and I.O. Ellis, *Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.* Histopathology, 1991. **19**(5): p. 403-10.
15. Benson, J.R., et al., *The TNM staging system and breast cancer.* Lancet Oncol, 2003. **4**(1): p. 56-60.

16. Park, Y.H., et al., *Clinical relevance of TNM staging system according to breast cancer subtypes.* Ann Oncol, 2010.

17. Bauer, K.R., et al., *Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry.* Cancer, 2007. **109**(9): p. 1721-8.

18. Parise, C.A., et al., *Breast cancer subtypes as defined by the estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2) among women with invasive breast cancer in California, 1999-2004.* Breast J, 2009. **15**(6): p. 593-602.

19. Ferlay, J., et al. *GLOBOCAN 2008, Cancer incidence and Mortality Worlwide: IARC CancerBase No. 10 [Internet].* 2010; Available from: http://globocan.iarc.fr.

20. Parent, A.S., et al., *The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration.* Endocr Rev, 2003. **24**(5): p. 668-93.

21. Hery, C., et al., *Changes in breast cancer incidence and mortality in middle-aged and elderly women in 28 countries with Caucasian majority populations.* Ann Oncol, 2008. **19**(5): p. 1009-18.

22. Anders, C.K., et al., *Breast cancer before age 40 years.* Semin Oncol, 2009. **36**(3): p. 237-49.

23. Shannon, C. and I.E. Smith, *Breast cancer in adolescents and young women.* Eur J Cancer, 2003. **39**(18): p. 2632-42.

24. Coleman, M.P., et al., *Cancer survival in five continents: a worldwide population-based study (CONCORD).* Lancet Oncol, 2008. **9**(8): p. 730-56.

25. Broca, P., *Traité des tumeurs.* Paris: Asselin, 1866.

26. Lynch, H.T., et al., *Hereditary breast cancer: part I. Diagnosing hereditary breast cancer syndromes.* Breast J, 2008. **14**(1): p. 3-13.

27. Easton, D.F., *Familial risks of breast cancer.* Breast Cancer Res, 2002. **4**(5): p. 179-81.

28. Foulkes, W.D., *Inherited susceptibility to common cancers.* N Engl J Med, 2008. **359**(20): p. 2143-53.

29. Lux, M.P., P.A. Fasching, and M.W. Beckmann, *Hereditary breast and ovarian cancer: review and future perspectives.* J Mol Med, 2006. **84**(1): p. 16-28.

30. Olopade, O.I., et al., *Advances in breast cancer: pathways to personalized medicine.* Clin Cancer Res, 2008. **14**(24): p. 7988-99.

31. Bradbury, A.R. and O.I. Olopade, *Genetic susceptibility to breast cancer.* Rev Endocr Metab Disord, 2007. **8**(3): p. 255-67.

32. Margolin, S., et al., *Family history, and impact on clinical presentation and prognosis, in a population-based breast cancer cohort from the Stockholm County.* Fam Cancer, 2006. **5**(4): p. 309-21.

33. Verhoog, L.C., et al., *Prognostic significance of germline BRCA2 mutations in hereditary breast cancer patients.* J Clin Oncol, 2000. **18**(21 Suppl): p. 119S-24S.

34. Hopper, J.L. and J.B. Carlin, *Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale.* Am J Epidemiol, 1992. **136**(9): p. 1138-47.

35. Lichtenstein, P., et al., *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland.* N Engl J Med, 2000. **343**(2): p. 78-85.

36. John, E.M., et al., *Migration history, acculturation, and breast cancer risk in Hispanic women.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(12): p. 2905-13.

37. Stanford, J.L., et al., *Breast cancer incidence in Asian migrants to the United States and their descendants.* Epidemiology, 1995. **6**(2): p. 181-3.

38. Key, T., et al., *Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies.* J Natl Cancer Inst, 2002. **94**(8): p. 606-16.

39. *Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer.* Lancet, 1997. **350**(9084): p. 1047-59.

40. Ewertz, M., et al., *Age at first birth, parity and risk of breast cancer: a meta-analysis of 8 studies from the Nordic countries.* Int J Cancer, 1990. **46**(4): p. 597-603.

41. Hinkula, M., et al., *Grand multiparity and the risk of breast cancer: population-based study in Finland.* Cancer Causes Control, 2001. **12**(6): p. 491-500.

42. Lambe, M., et al., *Transient increase in the risk of breast cancer after giving birth.* N Engl J Med, 1994. **331**(1): p. 5-9.

43. Liu, Q., et al., *Transient increase in breast cancer risk after giving birth: postpartum period with the highest risk (Sweden).* Cancer Causes Control, 2002. **13**(4): p. 299-305.

44. Russo, J., G. Wilgus, and I.H. Russo, *Susceptibility of the mammary gland to carcinogenesis: I Differentiation of the mammary gland as determinant of tumor incidence and type of lesion.* Am J Pathol, 1979. **96**(3): p. 721-36.

45. Travis, R.C. and T.J. Key, *Oestrogen exposure and breast cancer risk.* Breast Cancer Res, 2003. **5**(5): p. 239-47.

46. Russo, J., G.A. Balogh, and I.H. Russo, *Full-term pregnancy induces a specific genomic signature in the human breast.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(1): p. 51-66.

47. *Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease.* Lancet, 2002. **360**(9328): p. 187-95.

48. Yang, L. and K.H. Jacobsen, *A systematic review of the association between breastfeeding and breast cancer.* J Womens Health (Larchmt), 2008. **17**(10): p. 1635-45.

49. Gray, J., et al., *State of the evidence: the connection between breast cancer and the environment.* Int J Occup Environ Health, 2009. **15**(1): p. 43-78.

50. Rossouw, J.E., et al., *Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial.* JAMA, 2002. **288**(3): p. 321-33.

51. Santen, R.J., et al., *Postmenopausal hormone therapy: an Endocrine Society scientific statement.* J Clin Endocrinol Metab, 2010. **95**(7 Suppl 1): p. s1-s66.

52. Borgquist, S., et al., *Breast tumours following combined hormone replacement therapy express favourable prognostic factors.* Int J Cancer, 2007. **120**(10): p. 2202-7.

53. Holmberg, L. and H. Anderson, *HABITS (hormonal replacement therapy after breast cancer--is it safe?), a randomised comparison: trial stopped.* Lancet, 2004. **363**(9407): p. 453-5.

54. Beral, V., *Breast cancer and hormone-replacement therapy in the Million Women Study.* Lancet, 2003. **362**(9382): p. 419-27.

55. *Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Collaborative Group on Hormonal Factors in Breast Cancer.* Lancet, 1996. **347**(9017): p. 1713-27.

56. Deligeoroglou, E., E. Michailidis, and G. Creatsas, *Oral contraceptives and reproductive system cancer.* Ann N Y Acad Sci, 2003. **997**: p. 199-208.

57. Hunter, D.J. and W.C. Willett, *Diet, body size, and breast cancer.* Epidemiol Rev, 1993. **15**(1): p. 110-32.

58. Emons, J., et al., *Expression of vascular endothelial growth factor in the growth plate is stimulated by estradiol and increases during pubertal development.* J Endocrinol, 2010. **205**(1): p. 61-8.

59. Simm, P.J., et al., *Estrogens and growth.* Pediatr Endocrinol Rev, 2008. **6**(1): p. 32-41.

60. Dahlgren, A., et al., *Association of the estrogen receptor 1 (ESR1) gene with body height in adult males from two Swedish population cohorts.* PLoS One, 2008. **3**(3): p. e1807.

61. Reeves, G.K., et al., *Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study.* BMJ, 2007. **335**(7630): p. 1134.

62. Key, T.J., et al., *Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women.* J Natl Cancer Inst, 2003. **95**(16): p. 1218-26.

63. Renehan, A.G., et al., *Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies.* Lancet, 2008. **371**(9612): p. 569-78.

64. Pruitt, S.L., et al., *Association of area socioeconomic status and breast, cervical, and colorectal cancer screening: a systematic review.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(10): p. 2579-99.

65. Singletary, S.E., *Rating the risk factors for breast cancer.* Ann Surg, 2003. **237**(4): p. 474-82.

66. Laden, F. and D.J. Hunter, *Environmental risk factors and female breast cancer.* Annu Rev Public Health, 1998. **19**: p. 101-23.

67. Wiseman, M., *The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective.* Proc Nutr Soc, 2008. **67**(3): p. 253-6.

68. Sprague, B.L., et al., *Lifetime recreational and occupational physical activity and risk of in situ and invasive breast cancer.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(2): p. 236-43.

69. West-Wright, C.N., et al., *Long-term and recent recreational physical activity and survival after breast cancer: the California Teachers Study.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(11): p. 2851-9.

70. Roloff, T.C. and U.A. Nuber, *Chromatin, epigenetics and stem cells.* Eur J Cell Biol, 2005. **84**(2-3): p. 123-35.

71. Strachan, T. and A.P. Read, *Human Molecular Genetics 2*. 1999: BIOS Scientific Publishers Ltd.

72. Ballestar, E. and M. Esteller, *Epigenetic gene regulation in cancer.* Adv Genet, 2008. **61**: p. 247-67.

73. Feinberg, A.P. and B. Tycko, *The history of cancer epigenetics.* Nat Rev Cancer, 2004. **4**(2): p. 143-53.

74. Pfeifer, G.P. and A. Besaratinia, *Mutational spectra of human cancer.* Hum Genet, 2009. **125**(5-6): p. 493-506.

75. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts.* Nature, 1983. **301**(5895): p. 89-92.

76. Klose, R.J. and A.P. Bird, *Genomic DNA methylation: the mark and its mediators.* Trends Biochem Sci, 2006. **31**(2): p. 89-97.

77. Jovanovic, J., et al., *The epigenetics of breast cancer.* Mol Oncol, 2010. **4**(3): p. 242-54.

78. Liang, G., et al., *DNA methylation differences associated with tumor tissues identified by genome scanning analysis.* Genomics, 1998. **53**(3): p. 260-8.

79. Hinshelwood, R.A. and S.J. Clark, *Breast cancer epigenetics: normal human mammary epithelial cells as a model system.* J Mol Med, 2008. **86**(12): p. 1315-28.

80. Esteller, M., et al., *Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors.* J Natl Cancer Inst, 2000. **92**(7): p. 564-9.

81. Herman, J.G., et al., *Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers.* Cancer Res, 1995. **55**(20): p. 4525-30.

82. Gaudet, M.M., et al., *DNA hypermethylation of ESR1 and PGR in breast cancer: pathologic and epidemiologic associations.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(11): p. 3036-43.

83. Dawn Teare, M. and J.H. Barrett, *Genetic linkage studies.* Lancet, 2005. **366**(9490): p. 1036-44.

84. Hall, J.M., et al., *Linkage of early-onset familial breast cancer to chromosome 17q21.* Science, 1990. **250**(4988): p. 1684-9.

85. Wooster, R., et al., *Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13.* Science, 1994. **265**(5181): p. 2088-90.

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

87. Wooster, R., et al., *Identification of the breast cancer susceptibility gene BRCA2.* Nature, 1995. **378**(6559): p. 789-92.

88. Turnbull, C. and N. Rahman, *Genetic predisposition to breast cancer: past, present, and future.* Annu Rev Genomics Hum Genet, 2008. **9**: p. 321-45.

89. Ghoussaini, M. and P.D. Pharoah, *Polygenic susceptibility to breast cancer: current state-of-the-art.* Future Oncol, 2009. **5**(5): p. 689-701.

90. Cox, A., et al., *A common coding variant in CASP8 is associated with breast cancer risk.* Nat Genet, 2007. **39**(3): p. 352-8.

91. Pasche, B. and N. Yi, *Candidate gene association studies: successes and failures.* Curr Opin Genet Dev, 2010. **20**(3): p. 257-61.

92. Pharoah, P.D., et al., *Association between common variation in 120 candidate genes and breast cancer risk*. PLoS Genet, 2007. **3**(3): p. e42.

93. Cazier, J.B. and I. Tomlinson, *General lessons from large-scale studies to identify human cancer predisposition genes*. J Pathol, 2009. **220**(2): p. 255-62.

94. Pharoah, P.D., et al., *Association studies for finding cancer-susceptibility genetic variants*. Nat Rev Cancer, 2004. **4**(11): p. 850-60.

95. *International HapMap Project*. [cited 2011; Available from: http://hapmap.ncbi.nlm.nih.gov.

96. *The International HapMap Consortium -The International HapMap Project*. Nature, 2003. **426**(6968): p. 789-96.

97. *The International HapMap Consortium - A haplotype map of the human genome*. Nature, 2005. **437**(7063): p. 1299-320.

98. Manolio, T.A. and F.S. Collins, *The HapMap and genome-wide association studies in diagnosis and therapy*. Annu Rev Med, 2009. **60**: p. 443-56.

99. Barnes, M.R., ed. *Bioinformatics for geneticists*. 2007, Wiley.

100. Bertram, L. and R.E. Tanzi, *Genome-wide association studies in Alzheimer's disease*. Hum Mol Genet, 2009. **18**(R2): p. R137-45.

101. Billings, L.K. and J.C. Florez, *The genetics of type 2 diabetes: what have we learned from GWAS?* Ann N Y Acad Sci, 2010. **1212**: p. 59-77.

102. Carlson, C.S., et al., *Mapping complex disease loci in whole-genome association studies*. Nature, 2004. **429**(6990): p. 446-52.

103. Tenesa, A. and M.G. Dunlop, *New insights into the aetiology of colorectal cancer from genome-wide association studies*. Nat Rev Genet, 2009. **10**(6): p. 353-8.

104. Van Winkel, R., et al., *REVIEW: Genome-wide findings in schizophrenia and the role of gene-environment interplay*. CNS Neurosci Ther, 2010. **16**(5): p. e185-92.

105. Burton, P.R., M.D. Tobin, and J.L. Hopper, *Key concepts in genetic epidemiology*. Lancet, 2005. **366**(9489): p. 941-51.

106. Cordell, H.J. and D.G. Clayton, *Genetic association studies*. Lancet, 2005. **366**(9491): p. 1121-31.

107. Stranger, B.E., et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes*. Science, 2007. **315**(5813): p. 848-53.

108. Terwilliger, J.D. and T. Hiekkalinna, *An utter refutation of the "fundamental theorem of the HapMap"*. Eur J Hum Genet, 2006. **14**(4): p. 426-37.

109. Varghese, J.S. and D.F. Easton, *Genome-wide association studies in common cancers--what have we learnt?* Curr Opin Genet Dev, 2010. **20**(3): p. 201-9.

110. Travis, R.C., et al., *Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study.* Lancet, 2010. **375**(9732): p. 2143-51.

111. *http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems.html, Accessed December 29, 2010.* Applied Biosystems.

112. *http://www.illumina.com/systems/genome_analyzer_iix.ilmn, Accessed December 29th, 2010.* Illumina.

113. *http://454.com/products-solutions/454-sequencing-system-portfolio.asp, accessed December 29th, 2010.* Roche, 454.

114. Pharoah, P.D., et al., *Family history and the risk of breast cancer: a systematic review and meta-analysis.* Int J Cancer, 1997. **71**(5): p. 800-9.

115. Easton, D.F., et al., *Genome-wide association study identifies novel breast cancer susceptibility loci.* Nature, 2007. **447**(7148): p. 1087-93.

116. Gudmundsdottir, K. and A. Ashworth, *The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability.* Oncogene, 2006. **25**(43): p. 5864-74.

117. Lane, D.P., *Cancer. p53, guardian of the genome.* Nature, 1992. **358**(6381): p. 15-6.

118. Walsh, T., et al., *Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer.* JAMA, 2006. **295**(12): p. 1379-88.

119. Antoniou, A., et al., *Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies.* Am J Hum Genet, 2003. **72**(5): p. 1117-30.

120. Easton, D.F., D. Ford, and D.T. Bishop, *Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium.* Am J Hum Genet, 1995. **56**(1): p. 265-71.

121. *Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Breast Cancer Linkage Consortium.* Lancet, 1997. **349**(9064): p. 1505-10.

122. Rakha, E.A., J.S. Reis-Filho, and I.O. Ellis, *Basal-like breast cancer: a critical review.* J Clin Oncol, 2008. **26**(15): p. 2568-81.

123. Venkitaraman, A.R., *Chromosome stability, DNA recombination and the BRCA2 tumour suppressor.* Curr Opin Cell Biol, 2001. **13**(3): p. 338-43.

124. Rajan, J.V., et al., *Brca2 is coordinately regulated with Brca1 during proliferation and differentiation in mammary epithelial cells.* Proc Natl Acad Sci U S A, 1996. **93**(23): p. 13078-83.

125. Alter, B.P. and G. Kupfer, *Fanconi Anemia.* 1993.

126. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

127. Li, F.P. and J.F. Fraumeni, Jr., *Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome?* Ann Intern Med, 1969. **71**(4): p. 747-52.

128. Malkin, D., et al., *Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms.* Science, 1990. **250**(4985): p. 1233-8.

129. Nigro, J.M., et al., *Mutations in the p53 gene occur in diverse human tumour types.* Nature, 1989. **342**(6250): p. 705-8.

130. Evans, D.G., et al., *Low rate of TP53 germline mutations in breast cancer/sarcoma families not fulfilling classical criteria for Li-Fraumeni syndrome.* J Med Genet, 2002. **39**(12): p. 941-4.

131. Lalloo, F., et al., *BRCA1, BRCA2 and TP53 mutations in very early-onset breast cancer with associated risks to relatives.* Eur J Cancer, 2006. **42**(8): p. 1143-50.

132. Mavaddat, N., et al., *Genetic susceptibility to breast cancer.* Mol Oncol, 2010. **4**(3): p. 174-91.

133. Hemminki, A., et al., *A serine/threonine kinase gene defective in Peutz-Jeghers syndrome.* Nature, 1998. **391**(6663): p. 184-7.

134. Nelen, M.R., et al., *Localization of the gene for Cowden disease to chromosome 10q22-23.* Nat Genet, 1996. **13**(1): p. 114-6.

135. Guilford, P., et al., *E-cadherin germline mutations in familial gastric cancer.* Nature, 1998. **392**(6674): p. 402-5.

136. Masciari, S., et al., *Germline E-cadherin mutations in familial lobular breast cancer.* J Med Genet, 2007. **44**(11): p. 726-31.

137. Rahman, N., et al., *PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene.* Nat Genet, 2007. **39**(2): p. 165-7.

138. Bell, D.W., et al., *Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome.* Science, 1999. **286**(5449): p. 2528-31.

139. Nevanlinna, H. and J. Bartek, *The CHEK2 gene and inherited breast cancer susceptibility.* Oncogene, 2006. **25**(43): p. 5912-9.

140. McKinnon, P.J., *ATM and ataxia telangiectasia.* EMBO Rep, 2004. **5**(8): p. 772-6.

141. Swift, M., et al., *Breast and other cancers in families with ataxia-telangiectasia.* N Engl J Med, 1987. **316**(21): p. 1289-94.

142. Renwick, A., et al., *ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles.* Nat Genet, 2006. **38**(8): p. 873-5.

143. Thompson, D., et al., *Cancer risks and mortality in heterozygous ATM mutation carriers.* J Natl Cancer Inst, 2005. **97**(11): p. 813-22.

144. Cantor, S.B., et al., *BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function.* Cell, 2001. **105**(1): p. 149-60.

145. Seal, S., et al., *Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles.* Nat Genet, 2006. **38**(11): p. 1239-41.

146. Levitus, M., et al., *The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J.* Nat Genet, 2005. **37**(9): p. 934-5.

147. Reid, S., et al., *Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer.* Nat Genet, 2007. **39**(2): p. 162-4.

148. Antoniou, A.C. and D.F. Easton, *Models of genetic susceptibility to breast cancer.* Oncogene, 2006. **25**(43): p. 5898-905.

149. Dragani, T.A., F. Canzian, and M.A. Pierotti, *A polygenic model of inherited predisposition to cancer.* FASEB J, 1996. **10**(8): p. 865-70.

150. Hindorff, L.A., et al. *A Catalog of Published Genome-Wide Association Studies.* Acessed: [22-02-2011]; Available from: www.genome.gov/gwastudies.

151. Hunter, D.J., et al., *A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.* Nat Genet, 2007. **39**(7): p. 870-4.

152. Garcia-Closas, M., et al., *Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics.* PLoS Genet, 2008. **4**(4): p. e1000054.

153. Gorodnova, T.V., et al., *Distribution of FGFR2, TNRC9, MAP3K1, LSP1, and 8q24 alleles in genetically enriched breast cancer patients versus elderly tumor-free women.* Cancer Genet Cytogenet, 2010. **199**(1): p. 69-72.

154. Hemminki, K., et al., *Low-risk variants FGFR2, TNRC9 and LSP1 in German familial breast cancer patients.* Int J Cancer, 2010. **126**(12): p. 2858-62.

155. Huijts, P.E., 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): p. R78.

156. Rebbeck, T.R., 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): p. 269-74.

157. Adnane, 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): p. 659-63.

158. Moffa, A.B., S.L. Tannheimer, and S.P. Ethier, *Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells.* Mol Cancer Res, 2004. **2**(11): p. 643-52.

159. Udler, M.S., et al., *FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation.* Hum Mol Genet, 2009. **18**(9): p. 1692-703.

160. Meyer, K.B., et al., *Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer.* PLoS Biol, 2008. **6**(5): p. e108.

161. Stacey, S.N., et al., *Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer.* Nat Genet, 2007. **39**(7): p. 865-9.

162. Udler, M.S., et al., *Fine scale mapping of the breast cancer 16q12 locus.* Hum Mol Genet, 2010. **19**(12): p. 2507-15.

163. Dittmer, S., et al., *TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex.* J Cell Sci, 2011. **124**(Pt 2): p. 252-60.

164. Ericson, U., et al., *High folate intake is associated with lower breast cancer incidence in postmenopausal women in the Malmo Diet and Cancer cohort.* Am J Clin Nutr, 2007. **86**(2): p. 434-43.

165. Dhanasekaran, D.N. and G.L. Johnson, *MAPKs: function, regulation, role in cancer and therapeutic targeting.* Oncogene, 2007. **26**(22): p. 3097-9.

166. Zheng, W., et al., *Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1.* Nat Genet, 2009. **41**(3): p. 324-8.

167. Amundadottir, L.T., et al., *A common variant associated with prostate cancer in European and African populations.* Nat Genet, 2006. **38**(6): p. 652-8.

168. Ghoussaini, M., et al., *Multiple loci with different cancer specificities within the 8q24 gene desert.* J Natl Cancer Inst, 2008. **100**(13): p. 962-6.

169. Gudmundsson, J., et al., *Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24.* Nat Genet, 2007. **39**(5): p. 631-7.

170. Haiman, C.A., et al., *A common genetic risk factor for colorectal and prostate cancer.* Nat Genet, 2007. **39**(8): p. 954-6.

171. Kiemeney, L.A., et al., *Sequence variant on 8q24 confers susceptibility to urinary bladder cancer.* Nat Genet, 2008. **40**(11): p. 1307-12.

172. Sears, R.C., *The life cycle of C-myc: from synthesis to degradation.* Cell Cycle, 2004. **3**(9): p. 1133-7.

173. Milne, R.L., et al., *Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042.* J Natl Cancer Inst, 2009. **101**(14): p. 1012-8.

174. Stacey, S.N., et al., *Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer.* Nat Genet, 2008. **40**(6): p. 703-6.

175. Cavdar Koc, E., et al., *A new face on apoptosis: death-associated protein 3 and PDCD9 are mitochondrial ribosomal proteins.* FEBS Lett, 2001. **492**(1-2): p. 166-70.

176. Cai, Q., et al., *Replication and Functional Genomic Analyses of the Breast Cancer Susceptibility Locus at 6q25.1 Generalize Its Importance in Women of Chinese, Japanese, and European Ancestry.* Cancer Res, 2011. **71**(4): p. 1344-1355.

177. Palanca Suela, S., et al., *CASP8 D302H polymorphism delays the age of onset of breast cancer in BRCA1 and BRCA2 carriers.* Breast Cancer Res Treat, 2009. **119**(1): p. 87-93.

178. Sergentanis, T.N. and K.P. Economopoulos, *Association of two CASP8 polymorphisms with breast cancer risk: a meta-analysis.* Breast Cancer Res Treat, 2009. **120**(1): p. 229-34.

179. Sigurdson, A.J., et al., *Polymorphisms in apoptosis- and proliferation-related genes, ionizing radiation exposure, and risk of breast cancer among U.S. Radiologic Technologists.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(10): p. 2000-7.

180. Pharoah, P.D., et al., *Polygenes, risk prediction, and targeted prevention of breast cancer.* N Engl J Med, 2008. **358**(26): p. 2796-803.

181. Wacholder, S., et al., *Performance of common genetic variants in breast-cancer risk models.* N Engl J Med, 2010. **362**(11): p. 986-93.

182. Berglund, G., et al., *The Malmo Diet and Cancer Study. Design and feasibility.* J Intern Med, 1993. **233**(1): p. 45-51.

183. Manjer, J., et al., *The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants.* Eur J Cancer Prev, 2001. **10**(6): p. 489-99.

184. Berglund, G., et al., *Long-term outcome of the Malmo preventive project: mortality and cardiovascular morbidity.* J Intern Med, 2000. **247**(1): p. 19-29.

185. Nilsson, P.M., J.A. Nilsson, and G. Berglund, *Population-attributable risk of coronary heart disease risk factors during long-term follow-up: the Malmo Preventive Project.* J Intern Med, 2006. **260**(2): p. 134-41.

186. Pukkala, E., et al., *Nordic biological specimen banks as basis for studies of cancer causes and control--more than 2 million sample donors, 25 million person years and 100,000 prospective cancers.* Acta Oncol, 2007. **46**(3): p. 286-307.

187. Lukanova, A., et al., *Body mass index and cancer: results from the Northern Sweden Health and Disease Cohort.* Int J Cancer, 2006. **118**(2): p. 458-66.

188. Norberg, M., et al., *A combination of HbA1c, fasting glucose and BMI is effective in screening for individuals at risk of future type 2 diabetes: OGTT is not needed.* J Intern Med, 2006. **260**(3): p. 263-71.

189. Weinehall, L., et al., *Reduction of selection bias in primary prevention of cardiovascular disease through involvement of primary health care.* Scand J Prim Health Care, 1998. **16**(3): p. 171-6.

190. *The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators.* J Clin Epidemiol, 1988. **41**(2): p. 105-14.

191. Stegmayr, B., V. Lundberg, and K. Asplund, *The events registration and survey procedures in the Northern Sweden MONICA Project.* Scand J Public Health Suppl, 2003. **61**: p. 9-17.

192. *MONICA-undersökninen i norra Sverige.* 2011 Accessed [2011-03-07]; Available from: http://www9.umu.se/phmed/medicin/monica/.

193. Gabriel, S., L. Ziaugra, and D. Tabbaa, *SNP genotyping using the Sequenom MassARRAY iPLEX platform.* Curr Protoc Hum Genet, 2009. **Chapter 2**: p. Unit 2 12.

194. de Kok, J.B., et al., *Rapid genotyping of single nucleotide polymorphisms using novel minor groove binding DNA oligonucleotides (MGB probes).* Hum Mutat, 2002. **19**(5): p. 554-9.

195. Livak, K.J., J. Marmaro, and J.A. Todd, *Towards fully automated genome-wide polymorphism screening.* Nat Genet, 1995. **9**(4): p. 341-2.

196. Dunning, A.M., et al., *Association of ESR1 gene tagging SNPs with breast cancer risk.* Hum Mol Genet, 2009. **18**(6): p. 1131-9.

197. Reeves, G.K., et al., *Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci.* JAMA, 2010. **304**(4): p. 426-34.

198. Antoniou, A.C., et al., *Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction.* Cancer Res, 2010. **70**(23): p. 9742-54.

199. Thompson, D.A., et al., *Identification of two estrogen receptor transcripts with novel 5' exons isolated from a MCF7 cDNA library.* J Steroid Biochem Mol Biol, 1997. **62**(2-3): p. 143-53.

200. Lambertini, E., et al., *Expression of the human oestrogen receptor-alpha gene is regulated by promoter F in MG-63 osteoblastic cells.* Biochem J, 2003. **372**(Pt 3): p. 831-9.

201. Penolazzi, L., et al., *Methylation analysis of the promoter F of estrogen receptor alpha gene: effects on the level of transcription on human osteoblastic cells.* J Steroid Biochem Mol Biol, 2004. **91**(1-2): p. 1-9.

202. Antoniou, A.C., et al., *Common variants in LSP1, 2q35 and 8q24 and breast cancer risk for BRCA1 and BRCA2 mutation carriers.* Hum Mol Genet, 2009. **18**(22): p. 4442-56.

203. Barlow, L., et al., *The completeness of the Swedish Cancer Register: a sample survey for year 1998.* Acta Oncol, 2009. **48**(1): p. 27-33.

204. Beral, V., et al., *Breast cancer and abortion: collaborative reanalysis of data from 53 epidemiological studies, including 83?000 women with breast cancer from 16 countries.* Lancet, 2004. **363**(9414): p. 1007-16.

205. Kawase, T., et al., *FGFR2 intronic polymorphisms interact with reproductive risk factors of breast cancer: results of a case control study in Japan.* Int J Cancer, 2009. **125**(8): p. 1946-52.