Genetic polymorphisms in breast cancer in relation to risk and prognosis

Bågeman, Erika

2008

Link to publication

Citation for published version (APA):

Total number of authors: 1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Genetic polymorphisms in breast cancer in relation to risk and prognosis

Erika Bågeman

Department of Oncology, Clinical Sciences, Lund
Lund University, Lund, Sweden 2008
To 654 patients
# Table of Contents

Abstract 7  
Studies Included in Thesis 9  
Related Publications 11  
Abbreviations 13  
General Introduction 15  
Background 17  
  Normal Breast Development 17  
  Cancer Development 18  
  Breast Cancer 18  
  Risk Factors 19  
    Hormonal Factors 20  
    Lifestyle 22  
    Genetics 22  
    Gene-Environment Interactions 23  
Genetic Polymorphisms 23  
Diagnosis, Prognosis and Treatment Prediction 25  
  Diagnosis 25  
  Prognostic Factors 26  
  Treatment – Local, Adjuvant and Neoadjuvant 27  
  Treatment Response 28  
  Pharmacogenetics and Pharmacoepidemiology 29  
Aims 35  
Breast Cancer (BC) Blood Study 37  
  Study Design 37  
  Study Subjects 38  
Methods 41  
  DNA Extraction 41  
  Polymerase Chain Reaction (PCR) 41  
  Gel Electrophoresis 42  
  High Performance Liquid Chromatography (HPLC) 42
Abstract

Breast cancer is the most common cancer in women living in Sweden and the second most common cancer in the rest of the world. The risk of developing breast cancer is modified by environment, lifestyle, genetics and a combination of these factors. In this thesis work the combination of lifestyle and genetic polymorphisms on tumor characteristics and early recurrence have been studied.

Absence of the common 19 CA repeat allele in the insulin-like growth factor 1 gene (IGF1-19/-19) has been shown to modify the effect of several breast cancer risk factors. In Study I, multiparous patients with IGF1-19/-19 were 5.9 years younger at diagnosis than all other patients (P=0.007). Women with this combination may thus benefit from earlier breast cancer screening. In Study II concomitant tamoxifen and radiation therapy, increasing alcohol intake and moderate to high coffee consumption were all associated with an increased 2-hydroxyestrone (2-OHE) to 16α-hydroxyestrone (16αOHE1) ratio between the pre- and post-operative samples from the same patients. CYP1A2*1F were correlated with a lower ratio at both the pre- and the post-operative visit. Since a high 2-OHE/16αOHE1 ratio has been associated with improved survival the identified factors may modify breast cancer prognosis. In Study III the combined effect of CYP1A2*1F and coffee consumption was evaluated in relation to age at diagnosis and estrogen receptor (ER) status. In patients with CYP1A2*1F A/A who had never used hormone replacement therapy, higher coffee consumption was associated with a later age at diagnosis (57.7 versus 48.0 years; P=0.001) than in patients with lower consumption. Higher coffee consumption was also associated with a higher proportion of ER negative tumors (14.7% versus 0%, P=0.018). In Study IV the frequencies of CYP2C8*3, CYP2C8*4, CYP2C9*2, CYP2C9*3, GSTM1*0 and GSTT1*0 were evaluated and haplotype blocks constructed. CYP2C8/9 *1/*4/*1/*1 was associated with a lower frequency of axillary lymph node involvement as compared with the wild type in tumors larger than 20 mm, OR 0.13 (95% CI 0.04-0.45; P=0.001). CYP2C8*3 was associated with an increased risk of early recurrence, especially in women who had received tamoxifen, HR 2.93 (95% CI 1.25-6.85; P=0.013). In conclusion, both genetic and lifestyle factors are important for breast cancer.
Studies Included in Thesis


IV. Jernström H, Bågeman E, Rose C, Jönsson P-E, Ingvar C. Glutathione S-transferase (GST)M1, GSTT1 deletions and CYP2C8/9 polymorphisms in relation to tumor characteristics and early recurrences among 653 breast cancer patients. (Submitted)

All publications are reprinted by permission of the copyright holders.
Related Publications


Henningson M, Bågeman E, Sandberg T, Borg Å, Olsson H, Jernström H. Absence of the common IGF1 19 CA repeat allele is more common among BRCA1 mutation carriers than among non-carriers from BRCA1 families. Fam Cancer. 2007;6:445-52.
Abbreviations

APS: Adenosine 5’Phosphosulfate
ATP: Adenosine 5´Triphosphate
AR: Androgen Receptor
BC: Breast Cancer
BRCA1: Breast Cancer 1 (gene)
BRCA2: Breast Cancer 2 (gene)
CCD: Charge Coupled Device
cDNA: complementary DeoxyriboNucleic Acid
CI: Confidence Interval
CMF: Cyclophosphamide, Methotrexate and 5-Fluorouracil
CNC: Copy Number Change
COX: CycloOxygenase
CYP: Cytochrome P450
DEA: DihydroxyEicosatetraenoic Acid
DNA: DeoxyriboNucleic Acid
EIA: Enzyme ImmunoAssay
ER: Estrogen Receptor
EET: EpoxyEicosatrienoic Acid
FEC: Fluorouracil, Epirubicin and Cyclophosphamide
GWA: Genome Wide Association
HER2: Human Epidermal Growth Factor Receptor type 2
HETE: HydroxyEicosaTetraEnoic Acid
HPETE: HydroPeroxyEicosaTetraEnoic Acid
HPLC: High Performance Liquid Chromatography
HR: Hazard Ratio
HRT: Hormone Replacement Therapy
IGF1: Insulin-like Growth Factor-1 (gene)
IGF-1: Insulin-like Growth Factor-1 (protein)
IGFBP3: Insulin-like Growth Factor Binding Protein-3 (gene)
IGFBP-3: Insulin-like Growth Factor Binding Protein-3 (protein)
IHC: ImmunoHistoChemistry
IQR: Inter Quartile Range
LD: Linkage Disequilibrium
LHRH: Luteinizing Hormone Releasing Hormone
MAF: Minor Allele Frequency
mRNA: messenger Ribonucleic Acid
miRNA: micro Ribonucleic Acid
OC: Oral Contraceptives
OR: Odds Ratio
PAGE: PolyAcrylamide Gel Electrophoresis
PCR: Polymerase Chain Reaction
PR: Progesterone Receptor
PSQ: Pyrosequencing
RR: Relative Risk
RT: Radiation Therapy
RT-PCR: Reverse Transcriptase-Polymerase Chain Reaction
SNP: Single Nucleotide Polymorphism
TNM: T: Tumor size, N: lymph Node, M: distant Metastases
TEAA: TriEthylAmmonium Acetate
General Introduction

Breast cancer is the most common cancer in women in Sweden (The National Board of Health and Welfare) and the second most common cancer in the rest of the world (Parkin et al, 2005). In Sweden approximately one in ten woman will develop breast cancer before the age of 75 years (The National Board of Health and Welfare).

Estrogens and Insulin-like growth factor-1 (IGF-1) are involved not only in development of normal breasts but also in breast carcinogenesis (Jones et al, 1995; Laban et al, 2003; Pollak, 2000). Factors that modify the metabolism of estrogens and IGF-1 are therefore considered to be breast cancer risk factors. These factors include hormones, lifestyle and genetics. Disease-causing mutations have been found in several genes, most notably in Breast Cancer 1 and 2 (BRCA1 and BRCA2). However, mutations in these genes account for a minor fraction of all breast cancers, the fraction depending on the population studied (Loman et al, 2001; Narod et al, 2004). A substantial proportion of breast cancer risk may be the result of a combination of genetic and lifestyle factors. Genetic polymorphisms in genes coding for growth factors and genes involved in estrogen and drug metabolism may modify not only breast cancer risk but also prognosis and early recurrences.

Besides identification of carriers of high-risk BRCA1 or BRCA2 mutations, our ability to predict who will develop breast cancer is inadequate. In order to detect cancers early, when they are easier to treat, the recommendation is that all women in Sweden aged 40 to 74 years should attend mammography screening (The National Board of Health and Welfare). The treatment of breast cancer after surgical removal of the primary tumor is based on a number of prognostic and treatment predictive factors, including spread to the lymph nodes, age at diagnosis and estrogen receptor (ER) status. However, many patients do not benefit from the adjuvant treatment (Early Breast Cancer Trialists’ Collaborative Group, 2005). They are either cured by surgery, or they do not respond to the chosen adjuvant treatment as intended. It is therefore important to identify new prognostic and/or treatment predictive markers. Pharmacogenetics may be one answer.

By identifying gene-environment interactions we may be able to better predict a woman’s individual risk or prognosis. Women with a certain combination of lifestyle and genotype may represent a subgroup of individuals that should be offered earlier screening.
Background

Normal Breast Development

Breasts are essential to the survival of the species, as their primary task is to provide babies with milk. The breast consists of milk lobules, milk ducts, fat tissue, blood vessels and lymph ducts. As reviewed by Russo and Russo (Russo et al., 2004), development of the breast is initiated during embryonic life when the ducts are formed. Breast development is subject to regulation by the ovaries that produce ovarian steroid hormones, e.g. estrogens and progesterone. Moreover, the growth factor IGF-1 is essential to breast development (Jones et al., 1995; Laban et al., 2003; Pollak, 2000). It is not until the onset of puberty or within one to two years after menarche that the lobules develop.

As reviewed by Russo and Russo, there are four types of lobules, lobules 1 to 4, the most differentiated and proliferated state being lobule 4 (Russo et al., 2004). In the normal breast of an adult woman three types of lobules are present: lobule types 1 to 3. In nulliparous women lobule type 1 is the most predominant lobule, Fig 1A, whereas in parous women the type 3 lobule is the most common structure. During pregnancy and lactation lobule 3 develops into lobule 4, Fig 1B. The proportion of lobule 3 in parous women peaks during the early reproductive years and then decreases with age. Parous postmenopausal women therefore have almost the same breast composition, i.e. lobule type 1, as the nulliparous woman (Russo et al., 2008; Russo et al., 1994). However, lobule 1 in nulliparous and postmenopausal parous women may be biologically different. Since most cancers are initiated in lobules 1 and 2, the time window between menarche and the birth of the first child, when most lobules mature to lobule 3, has been considered to be a period when the breast cells are sensitive to hormonal stimuli. Additional full-term pregnancies further reduce the number of remaining type 1 lobules into more mature lobules.
Cancer Development

Cancer arises as a result of a number of genetic alterations in the dividing cell. As reviewed by Hanahan and Weinberg (Hanahan et al., 2000), the hallmarks of cancer are cells with limitless cell-dividing capacity that are self-sufficient in growth factors and insensitive to anti-growth signals, and with the ability to evade apoptosis, the programmed cell death by which defective cells are usually eliminated. The tumor cells also acquire the ability to invade tissue, and are then referred to as invasive cancer. These cells are dependent on nutrients and oxygen supplied by the blood, and must therefore be able to create new blood vessels (sustained angiogenesis). The most aggressive form of cancer can penetrate the blood vessels and the lymphatic system, giving rise to metastases in other parts of the body.

Breast Cancer

Breast cancer is the most common malignancy in women in Sweden (The National Board of Health and Welfare) and the second most common cancer in the world.
Breast cancer incidence in women has increased by approximately 1.3% annually for the last 20 years (The National Board of Health and Welfare), Fig 2. Mortality rates have decreased in Sweden (Althuis et al, 2005). The majority of breast cancer patients are aged 60 to 64 years at diagnosis, and fewer than 4% are diagnosed before the age of 40 years (The National Board of Health and Welfare). However, a significant reduction in the number of cancers in women aged over 45 was recently reported (The National Board of Health and Welfare).

Breast cancer is a heterogeneous disease, and a number of histological subtypes have been identified. The most common subtype is ductal cancer, followed by lobular cancer, whilst medullary and papillary subtypes are less common (Ellis et al, 1992). Histological grading can distinguish tumors of different growth characteristics, and other important tumor markers used clinically in classification of breast cancer include ER and progesterone receptor (PR) status, as well as Human Epidermal Growth Factor Receptor type 2 (HER2) status. In addition, molecular subtypes based on gene expression profiling have been established (Perou et al, 2000). However, this classification is not yet used in the clinical setting. Moreover, premenopausal and postmenopausal breast cancer may display distinct features and may even be considered two different diseases. Protective lifestyle factors such as a high body mass index (BMI) in younger women may even be risk factors in postmenopausal women.

Risk Factors

Breast cancer incidence varies in different parts of the world. The fact that a woman’s breast cancer risk increases within one or two generations when moving from a low-

risk to a high-risk area (Ziegler et al, 1993) indicates that not only genetic factors but also the environment affects breast cancer risk (Lichtenstein et al, 2000). Genetic variants and lifestyle differ not only between ethnic groups but also between countries. Several breast cancer risk factors are also associated with prognosis. Paradoxically, a factor that confers an increased breast cancer risk may be associated with an improved prognosis if the woman develops breast cancer.

**Hormonal Factors**

Reproductive factors including age at menarche and menopause, parity, age at first full-term pregnancy and breast-feeding are all considered to influence breast cancer risk (Kelsey et al, 1993). An early age at menarche and a late age at menopause comprise prolonged endogenous hormone exposure and have therefore been associated with an increased breast cancer risk. Pregnancy has a complex association with the risk of breast cancer: it transiently increases the risk after childbirth but reduces the risk in later years. The short-term risk may reflect the growth-promoting effect of high hormone levels during pregnancy on already existing transformed tumor cells. This transient increased risk is most pronounced in women who are at least 30 at the time of their first delivery (Lambe et al, 1994). The increased breast cancer risk declines with time after delivery, and after 15 years the risk in uniparous women aged at least 30 years at delivery is similar to that of nulliparous women, Fig 3A. The short-term increased risk also declines with number of pregnancies, Fig 3B, and the risk following birth of a third child is considered to be the same as in a nulliparous woman (McCredie et al, 1998). The risk of breast cancer further decreases with time since delivery, and the overall effect of pregnancy is protective. Parity therefore has a dual effect on breast cancer risk.

![Figure 3](image_url)  
**Figure 3**  A Odds ratios for the risk of breast cancer in uniparous women of various ages at delivery, according to the number of years since delivery (Lambe et al, 1994). B Odds ratios for the risk of breast cancer in biparous women of various ages at second delivery, according to the number of years since delivery (Lambe et al, 1994). Copyright © [2008] Massachusetts Medical Society. All rights reserved.
Although women in the developing world more often have multiple children at an early age, the risk of early-onset breast cancer is substantially higher than in women in the developed world (Hall et al., 2005). The paradox that multiparity is not protective in all women was studied in Study I.

Breast-feeding has been shown to decrease breast cancer risk by approximately 4% for every 12 months of breast-feeding (Collaborative Group on Hormonal Factors in Breast Cancer, 2002).

Endogenous hormones and growth factors including estrogens, progesterone, testosterone and IGF-1 are essential to normal breast development but also increase breast cancer risk. Estrogen is not a single hormone, but is present in several forms. Estrogens are formed in discrete stages during synthesis from steroid precursors, and can be converted into different metabolites with different ER-stimulating efficiency. The 2-hydroxyestrogen 2-OHE to 16α-hydroxyestrone (16αOHE1) ratio reflects the relative activities of the weak (2-OHE) (Schneider et al., 1984), and the procarcinogenic 16αOHE1 (Telang et al., 1992). Most studies have found an association between a higher urinary 2-OHE/16αOHE1 ratio and a decreased breast cancer risk (Ho et al., 1998; Kabat et al., 1997; Meilahn et al., 1998; Muti et al., 2000). Lifestyle factors, including coffee consumption, may modify this ratio (Jernström et al., 2003b) and may therefore influence breast cancer risk and prognosis. Testosterone also affects breast cancer risk, in both pre- and postmenopausal woman (Hankinson et al., 2007; Tamimi et al., 2007). IGF-1 increases the risk of premenopausal breast cancer (Fletcher et al., 2005; Hankinson et al., 1998; Renehan et al., 2004; Schernhammer et al., 2005; Sugumar et al., 2004) as well as relapse (Vadgama et al., 1999). IGF-1 levels vary during the menstrual cycle, both in plasma (Jernström et al., 1994) and locally in the breast (Dabrosin, 2003). Hormone studies thus need to control for cycle day in premenopausal women.

Exogenous hormones, e.g. oral contraceptives (OCs) and hormone replacement therapy (HRT), are also associated with an increased risk of developing breast cancer. The use of combined OCs increases the breast cancer risk during use and up to 10 years after cessation (Collaborative Group on Hormonal Factors in Breast Cancer, 1996). HRT has also been associated with an increased breast cancer risk during use and up to five years after cessation (Collaborative Group on Hormonal Factors in Breast Cancer, 1997; Jernström et al., 2003a; Magnusson et al., 1999; Olsson et al., 2003; Rosenberg et al., 2006; Rossouw et al., 2002).

Anthropometric factors are associated with estrogens levels and thus breast cancer risk. In postmenopausal women the majority of estrogen is produced through peripheral aromatization from androgens in fat tissue. The local estrogen production is proportional to the amount of body fat (McTiernan et al., 2003). Overweight women have a higher body mass index (BMI), i.e. weight/length$^2$ (kg/m$^2$), and may therefore be at higher postmenopausal breast cancer risk (Boyapati et al., 2004; Magnusson et al., 1998). In premenopausal women estrogens are mainly produced in the ovaries and not locally in the fat tissue, and in this subgroup of women obesity is actually protective (Weiderpass et al., 2004).
Breast volume (large cup size) has also been associated with an increased risk of premenopausal breast cancer in lean women (Kusano et al., 2006).

Lifestyle

*Dietary factors* include alcohol and coffee consumption. Daily alcohol consumption of 3 to 4 drinks (35 to 44 g alcohol) has been associated with an approximately 30% higher risk than is the case with non-consumption (Hamajima et al., 2002). The risk increases to almost 50% with consumption of more than 4 drinks/day (≥45 g alcohol).

Coffee is considered to be the world’s most popular drug. Numerous studies of coffee in relation to breast cancer risk have been performed (Baker et al., 2006; McLaughlin et al., 1992; Michels et al., 2002; Nkondjock et al., 2006; Phelps et al., 1988, Rosenberg et al., 1985; Stensvold et al., 1994; Vatten et al., 1990), but the results are inconsistent. This may be the result of not taking genetic factors into consideration. The association between lifestyle factors, including coffee and a genetic normal variant denoted *CYP1A2*1F, on estrogen metabolite levels was evaluated in Study II. The combined effect of coffee consumption and *CYP1A2*1F on age at breast cancer onset and ER status was evaluated in Study III.

Other lifestyle factors associated with breast cancer risk include smoking, physical activities and socioeconomic status. The effect of smoking on breast cancer risk may be difficult to evaluate, since it may be confounded by alcohol (Hamajima et al., 2002) and OC use (Jernström et al., 2005a). Smoking has been associated with an increased risk in premenopausal women in some (Johnson et al., 2000) but not all studies (Jernström et al., 2005a). In a Swedish cohort of postmenopausal women no association between smoking and breast cancer risk was found (Magnusson et al., 2007). Physical activities or socioeconomics were not evaluated in this thesis.

Even though some women have a certain ‘high-risk’ lifestyle they may not develop breast cancer. This could in part be explained by the fact that genetic factors modify the effect of lifestyle on breast cancer risk (Lichtenstein et al., 2000).

Genetics

Hereditary breast cancer accounts for only 5-10% of all breast cancers and germline mutations with the two major breast cancer susceptibility genes *BRCA1* and *BRCA2*, being responsible for a small fraction (-2-3%) of all breast cancers (Loman et al., 2001; Narod et al., 2004). In addition to *BRCA1* and *BRCA2*, *TP53* and *PTEN* are considered to be high-penetrance breast cancer susceptibility genes, whereas *ATM*, *BRIP1*, *CHEK2*, and *PALB2* are considered to be moderate-penetrance susceptibility genes (Liaw et al., 1997; Stratton et al., 2008). A large proportion of familial aggregation of breast cancer, and possibly non-familial disease, is considered to be due to the effect of low-risk alleles, some being very common and possibly acting via polygenic
mechanisms and in interaction with environmental and lifestyle factors. Although powerful techniques are also available for evaluation of the millions of genetic markers in each patient, and large multicentre cohorts have been analyzed for potential risk factors, the results of these studies have so far been limited. Only a handful of common low-penetrance risk alleles have been replicated, including FGFR2, 2q, CASP8_D302H, MAP3KI, TNRC9, 8q and LSP1 (Stratton et al, 2008). To further identify the numerous suspected less common low-risk alleles, very large case-control cohorts would need to be analyzed in order to statistically secure a risk association.

New approaches for identification of risk and prognostic markers are warranted. A combination of genetic and lifestyle factors, as well as environmental factors that modify the risk of developing breast cancer, probably account for the majority of malignancies (Le Marchand et al, 2008).

**Gene-Environment Interactions**

Gene-environment studies may yield new insights with respect to breast cancer. A given exposure, e.g. lifestyle, may have different or even opposite effects on breast cancer risk or prognosis in women, depending on their genetic variants. When not taking both genetic and environmental factors into account the potential effect on breast cancer may not be detected, since both women with an increased breast cancer risk and those with a decreased risk are combined in one group of cases. In the event of a gene-environment interaction, the combined effect is often greater than that of the genetic variant or lifestyle factor itself (Le Marchand et al, 2008).

In this thesis, a set of candidate genes and specific genetic polymorphisms have been selected, as they are known modifiers of IGF-1 levels, estrogen metabolism and the metabolism of drugs used clinically to treat breast cancer. If these genetic variants are also associated with tumor characteristics, they may be even more relevant. The genetic polymorphisms studied here have been selected based on a minor allele frequency (MAF) in European populations of >5%, thus increasing the chance of the results being of general interest and clinically useful.

**Genetic Polymorphisms**

Each individual is unique, though comparison of the genomes of any two individuals only shows a ~0.1% difference. Single nucleotide polymorphisms (SNPs) explain up to 95% of all variant DNA sites (Meyer, 2004). According to data obtained from the dbsNP on July 16 2008 there are 14,110,048 registered SNPs in the database (Shersey et al, 1999), which can be compared with the approximately 3.2 billion bases in the human haploid genome.
The four different nucleotides (Adenine; A, Cytosine; C, Guanine; G, Thymine; T) that constitute the building blocks of our DNA can be altered in different ways: they can be exchanged, duplicated, deleted or rearranged. Larger genetic (chromosomal) regions can also be gained or lost (copy number changes; CNCs), or rearranged (e.g. in translocations or inversions). These alterations arise somatically at a high rate, particularly in cancer cells where they might become enriched, but they may also occur in germ cells, and can thus be transmitted as constitutional variants to coming generations. Various DNA repair mechanisms normally act to preserve high genome integrity, but never with complete fidelity. From an evolutionary perspective a certain level of continuous germline mutagenesis may be deemed necessary to allow individuals to cope with environmental or lifestyle shifts and challenges. Sequence alterations can take place anywhere in the genome, and the vast majority end up in non-coding sequences and have no or little effect on cell function. Repetitive DNA sequences (e.g. microsatellites) are susceptible to alterations. Other changes occur in coding or regulatory sequences and may alter gene function or expression and confer a selective advantage or disadvantage for the cell or organism.

Single nucleotide polymorphism is by definition a nucleotide exchange that occurs in at least one percent of a population (reviewed in (Risch, 2000)). A genetic variant that occurs in less than 1% of a given population is referred to as a mutation or rare variant. The definition is not univocal, and some people refer to a SNP as a non-disease-causing variant, whereas a mutation is a disease-causing one.

SNPs can be either transitions or transversions. A transition is an exchange of a purine for a purine (A→G or G→A) or a pyrimidine for a pyrimidine (C→T or T→C), whereas a transversion is a replacement of a purine by a pyrimidine (A→C, A→T, G→C, G→T) or vice versa (C→A, C→G, T→A, T→G). Each registered SNP has an rs number, to facilitate nomenclature. A SNP may have either a dominant or a co-dominant effect (Minelli et al, 2005).

When the combination of two or more SNPs occurs in a population more or less frequently than is expected by chance, they are considered to be in linkage disequilibrium (LD). Haplotypes can be defined as SNPs that are located close together on the same chromosome, that are less likely to be disrupted by meiotic crossing-over and that are thus inherited together. Certain regions in the genome are protected against such recombinations and are referred to as haplotype blocks. Therefore a number of SNPs (tagging SNPs) may capture most of the genetic diversity across that specific region (Johnson et al, 2001).

SNPs can be divided into subgroups based on the effect of the genetic variation. SNPs that result in an amino acid substitution or premature stop codon are referred to as nonsynonymous SNPs, whereas SNPs that do not result in an amino acid change are referred to as synonymous SNPs. SNPs can also be divided into silent, harmless, harmful, and latent SNPs (Greenhut et al, 2004). Silent SNPs are variants in non-coding or coding regions and are predominantly thought of as being non-functional. However, silent SNPs may indirectly change the transcription, structure and stability of the mRNA, transcript splicing and the kinetics of translation, and thereby the amount of protein, its structure and its function (Sauna et al, 2007).
Moreover, silent SNPs may not be functional, but are rather markers, i.e. in LD with the functional SNP. Harmless SNPs are located in coding or regulatory regions, but mostly have a subtle impact on genetic and cellular function. Some harmless SNPs may change your phenotype and appearance without causing disease. Harmful SNPs are responsible for the increased risk of diseases such as cancer. Latent SNPs may be harmless unless a certain lifestyle factor or exposure is present, e.g. hormones and breast cancer medications. SNPs that regulate the rate of absorbance, binding, metabolism or excretion of toxic substances may affect a woman’s breast cancer risk or response to a given treatment.

*Copy number changes; duplications/deletions* are the gains or losses of genetic material, from single exons or genes up to whole chromosome regions. Loss of function of a gene copy, for instance through a deletion, can be tolerated by the cell, since the remaining copy sustains gene function, but this may result in haploinsufficiency and be harmful. Loss of both gene copies often has more severe effects, but might appear as latent. Increased copy number, for instance via tandem duplication, can increase gene dosage and be harmful.

*Microsatellites* are tandem repeats of mono-, di-, tri- or tetra-nucleotide units or more that form clusters <10 to >100 base pairs in length. These sequences create problems for the DNA replication machinery, and polymerase slippage that can result in unrepaired deletions or duplications of single or multiple repeat units. This may occur in germ cells, and over time it has consequently resulted in a high level of variability in the number of repeats in several populations, seen as a difference between alleles in and between individuals. The most common repeat is the CA dinucleotide repeat that constitutes approximately 0.25% of the human genome (Lander *et al.*, 2001). The length of these repetitive fragments may affect transcription of genes (Lundin *et al.*, 2007) and thereby protein levels (Tae *et al.*, 1994), but may also occur in coding sequences and alter protein function. Even though these microsatellites vary in size, specific repeat lengths are more common than others and may have a different effect on the gene compared with both longer and shorter repeat sizes (Lundin *et al.*, 2007).

**Diagnosis, Prognosis and Treatment Prediction**

**Diagnosis**

Breast cancer is diagnosed through a triple diagnostic procedure including clinical examination, mammography, and fine needle aspirations or tissue biopsy. The national guidelines state that mammography screening should be offered to all women aged 40 to 74 years (The National Board of Health and Welfare). Young women may therefore have tumors that are detected at a later and more advanced stage.
Breast cancer in young women may in general also have a higher proliferation rate than breast cancer in postmenopausal women, and may thus advance faster.

**Prognostic Factors**

In order to predict outcome after the primary operation prognostic factors are used (Goldhirsch et al, 2007), including:

- Size of the invasive component of the tumor
- Lymph node involvement
- Histological grade
- Age at diagnosis
- Human Epidermal Growth Factor Receptor type 2 (HER2)
- Estrogen Receptor (ER)
- Progesterone Receptor (PR)
- Extensive peritumoral vascular invasion

Treatment predictive factors (ER, PR, and HER2) are used for the choice of therapy.

**TNM classification**

Tumors are classified according to invasive size of the tumor (pT), lymph node involvement (pN) and distant metastases (M), where ‘p’ refers to a pathological examination. pT0 represents no sign of primary tumor. pTis stands for carcinoma *in situ*, which is a pre-invasive cancer where the cancer cells are proliferating in an uncontrolled manner, but have not invaded through the basal membrane into the surrounding normal tissue. pT1-3 represents different sizes of the tumor and T4 represents a tumor that has grown into the chest wall or involves the skin, independently of its size. Increased tumor size is associated with an increased risk of lymph involvement and a decreased chance of survival (Carter et al, 1989). pN0 represents no spread to the lymph nodes. pN1 tumors have spread to the axillary lymph nodes. Increasing lymph node involvement has been associated with decreased survival irrespective of tumor size (Carter et al, 1989). Distant metastases are denoted by M, where M0 represents no distant metastases and M1 distant metastases. Breast cancer can be classified in four main stages based on the TNM classification, but the use of this classification varies from country to country and the cancers are sometimes referred to as early- and later-stage breast cancer, or simply node-negative or node-positive. Owing to increased use of screening, a growing proportion of patients are diagnosed with early-stage disease, emphasising the need for another way of distinguishing between breast cancer with a good prognosis and breast cancer with a poor prognosis.
**Histological grade**

The histological grade of the tumor is based on the evaluation of tubular differentiation, nuclear pleomorphism (grade), and mitotic count (Elston *et al.*, 1991). Each of these three morphologic features is given a score of between 1 and 3. The overall histological grade is obtained by adding the score of each characteristic, giving a possible total score of between 3 and 9. A score of between 3 and 5 denotes a Grade 1 tumor, a score of between 6 and 7 a Grade 2 tumor and a score of between 8 and 9 a Grade 3 tumor.

**Age at diagnosis**

Women diagnosed prior to the age of 35 years are considered to be at a higher risk of recurrence than older patients (Goldhirsch *et al.*, 2007).

**HER2**

HER2 (HER2/neu, c-erbB-2) is a tyrosine kinase receptor that is overexpressed and/or amplified in approximately 15-25% of all breast cancers (Owens *et al.*, 2004; Paik *et al.*, 1990; Press *et al.*, 2005; Rasmussen *et al.*, 2008; Slamon *et al.*, 1987). HER2 is not only a prognostic but also a treatment predictive factor for the response to the monoclonal antibody trastuzumab (McNeil, 1998) and the tyrosine kinase inhibitor lapatinib (Xia *et al.*, 2002).

**ER and PR status**

ER and PR are weak prognostic factors. They are mainly treatment predictive factors. Receptor-positive tumors are more sensitive to endocrine treatment ([EBCTCG], 2005). ER and PR expression is related to the degree of tumor differentiation. ER is expressed in approximately 80% of all newly diagnosed breast cancers in Sweden. In sporadic breast cancer, postmenopausal women tend to have higher ER concentrations than premenopausal women (Yasui *et al.*, 1999). Since the discovery of a second ER, ERβ, it is important to specify that it is the expression of ERα that is currently being analyzed.

**Extensive peritumoral vascular invasion**

According to the St Gallen guidelines extensive peritumoral vascular invasion is recommended as a prognostic factor (Goldhirsch *et al.*, 2007). However, it is not yet included in the Swedish recommendations (SweBCG).

**Treatment – Local, Adjuvant and Neoadjuvant**

Breast cancer is primarily treated with surgery, either by means of modified radical mastectomy, whereby the complete breast is removed, or by means of breast-conserv-
ative surgery, whereby only part of the tissue is removed. During surgery axillary lymph nodes are removed – in many cases only the sentinel node(s).

Adjuvant therapy improves the prognosis for many patients, but this advantage should be considered in relation to side effects. Adjuvant therapies for breast cancer patients include radiation therapy (RT), chemotherapy, endocrine therapy and antibodies, and are chosen based on prognostic and treatment predictive factors. Combinations of these adjuvant regimes are often used. Neoadjuvant treatment is administered in order to decrease tumor size prior to surgery, and to facilitate the evaluation of treatment response.

Polychemotherapy regimes have changed over the years from CMF (cyclophosphamide, methotrexate, and 5-fluorouracil) to anthracycline-based regimes such as FEC (5-fluorouracil, epirubicin, and cyclophosphamide), and taxanes are currently included in the cytostatic treatment for patients with a higher risk of recurrence (SweBCG).

Adjuvant endocrine therapy includes the anti-estrogen tamoxifen, aromatase inhibitors, luteinizing hormone releasing hormone (LHRH) analogues, and oophorectomy or radiation of the ovaries in premenopausal women (Early Breast Cancer Trialists’ Collaborative Group, 2005).

Trastuzumab is an antibody-based therapy directed against the growth factor receptor HER2 (Romond et al, 2005; Smith et al, 2007).

Approximately two thirds of all breast cancer patients are cured by surgical treatment as single treatment modality, and only one third of the patients need to be selected for additional therapy. However, adjuvant therapy is delivered to 80-90% of patients, and a considerable proportion is over-treated. Our selection criteria for low- and high-risk patients need to be improved. Moreover, patients who suffer recurrence after adjuvant therapy need other treatment regimes. New techniques for subclassification of tumors in order to better predict treatment response are currently being evaluated. The most promising results have been obtained by means of gene-expression analyses (cDNA microarray, oligonucleotide array, reverse-transcriptase polymerase chain reaction (RT-PCR)). With these methods 70- and 21-gene profiles have been identified (Paik et al, 2004; van ’t Veer et al, 2002; van de Vijver et al, 2002). Other array-based techniques (array Comparative Genomic Hybridization) and techniques within the field of proteomics are currently also being evaluated.

**Treatment Response**

In addition to tumor characteristics a patient’s response to a certain treatment is dependent on several factors, including absorption of the drug, transportation, metabolism and excretion. Factors that may interfere with the drug response can be viewed as confounding factors. These may be concomitant drug treatments that are metabolized by the same enzyme system, complementary alternative medicines and dietary factors. These factors may regulate the amount of enzyme produced or alter the enzyme function. Other well-known factors that may affect the response to a giv-
en treatment include the age and gender of the patient and his or her general health status.

Metabolism of drugs as well as endogenous hormones is carried out by Cytochrome P450 (CYP) enzymes expressed in the liver and in the small intestine. This family of enzymes includes Phase I enzymes that convert not only exogenous compounds such as drugs but also endogenous hormones. The CYP enzymes have family, subfamily and isozyme names. The CYP2C8 and the CYP2C9 enzymes thus belong to the same family and subfamily of CYP enzymes, i.e. 2C, but they are different isozymes, i.e. 8 and 9, respectively. Phase II enzymes are detoxifying enzymes that transform drugs into more water-soluble molecules that can be excreted. Phase II enzymes include the glutathione-S transferases (GSTs). Phase I and II enzymes are also responsible for the metabolism of estrogens (Rebbeck et al, 2007).

Pharmacogenetics and Pharmacoepidemiology

It is estimated that approximately 20% of drug therapies are influenced by genetic polymorphisms in drug-metabolizing genes (Ingelman-Sundberg, 2004; Kalow et al, 1998).

Pharmacogenomics refers to the general study of all of the many different genes that determine drug behaviour, whereas pharmacogenetics refers to the study of inherited differences (variations) in drug metabolism and response (National Center for Biotechnology Information). These definitions are not absolute. Some refer to the study of selected genes or polymorphisms to “genetics”, and “genomics” is then used when referring to whole genome-wide scans. Pharmacoepidemiology is the study of the use of drugs and their effects and side effects in real life, without the exclusion criteria commonly used in clinical trials.

An example of pharmacogenetics is the attempt to individualize treatment with the anti-estrogen tamoxifen (Goetz et al, 2008). Tamoxifen is a prodrug that is activated by the enzyme CYP2D6 to 4-hydroxytamoxifen, which is then converted into endoxifen – the most potent metabolite (Jin et al, 2005). The CYP2D6 gene is polymorphic, and patients with the CYP2D6*4 allele have been shown to be poor metabolizers of tamoxifen. Poor metabolizers cannot activate tamoxifen and should therefore be offered another type of treatment, e.g. aromatase inhibitors, in postmenopausal women. Antidepressant agents are considered confounders in tamoxifen treatment, since they interfere with the CYP2D6 enzyme (Jin et al, 2005). Patients on antidepressant drugs may therefore not benefit from tamoxifen treatment. Even though this is a very interesting polymorphism, the ethics committee only recently gave us permission to evaluate this SNP, and it has not yet been genotyped.

As many patients do suffer recurrence in spite of various adjuvant cancer therapies (Early Breast Cancer Trialists’ Collaborative Group, 2005), the benefit of customizing treatment based on the individual patient’s pharmacogenetic profile may be substantial. Optimum tailored treatment would not only decrease the number of patients treated but would also increase the response rate. In addition, pharmaco-
genetics may also limit the number of adverse drug reactions, since patients who cannot deactivate the drug administered may be offered a lower dose or alternative treatment. Moreover, the development of new drugs in clinical trials would be more efficient when evaluating the response only in patients who can respond to the treatment administered. In terms of drug therapy, one size does not fit all. Pharmacogenetics may therefore have a major impact, not only on the quality of life of individual patients but also socioeconomically. New drugs are expensive and should be used only for those who will benefit.

BOX 1
Study 1

Insulin-like growth factor-1 (IGF-1) acts as an effector molecule for growth hormone and stimulates proliferation of the epithelial cells of the breast through interaction with the IGF-1 receptor. IGF-1 is mainly produced in the liver, and circulating IGF-1 levels are known to interact with the IGF-1 receptor of breast epithelial cells, predominantly inducing intracellular signaling through the mitogen activated protein kinase and phosphatidylinositol 3-kinase pathways. IGF-1 also acts in an autocrine and a paracrine fashion in the breast (Pollak, 2004). When IGF-1 interacts with its receptor, IGF-1 protects cancer cells from apoptosis and promotes survival (reviewed in (Jones et al., 1995; Rubin et al., 1995)). Plasma concentrations of IGF-1 are higher in breast cancer patients than in healthy controls (Peyrat et al., 1993). IGF-1 is positively correlated with risk of breast cancer in premenopausal women (Fletcher et al., 2005; Hankinson et al., 1998; Renehan et al., 2004; Schernhammer et al., 2005; Sugumar et al., 2004), as well as increased risk of relapse (Vadgama et al., 1999).

The IGF1 gene contains a microsatellite repeat located approximately 1 kb upstream of the transcription start site (Rotwein et al., 1986). The tandem dinucleotide CA repeat ranges in size from 11 to 24 repeats (Jernström et al., 2001a; Jernström et al., 2001b; Schildkraut et al., 2005), the 19 repeat allele is the most common repeat among white women (Jernström et al., 2001a; Jernström et al., 2001b), and over 60% of white women carry at least one 19 CA repeat allele. Only 6-13% of white women have no copy of this repeat length, IGF1-19/-19 (DeLellis et al., 2003; Jernström et al., 2001a; Jernström et al., 2001b; Jernström et al., 2005b; Vaessen et al., 2001).

The IGF1-19/-19 genotype modifies a number of breast cancer risk factors including OCs and breast volume after pregnancy (Cleveland et al., 2006; Jernström et al., 2005b). However, no study has investigated the potential association between IGF1-19/-19 and multiparity on age at breast cancer diagnosis.
A high 2-OHE/16αOHE1 ratio in urine has been associated with a decreased breast cancer risk in most studies (Ho et al., 1998; Kabat et al., 1997; Meilahn et al., 1998; Muti et al., 2000), but not all studies (Ursin et al., 1999). Factors that modify this ratio are therefore of interest with respect to breast cancer.

**CYP1A2** is a key enzyme in the estrogen-metabolism hydroxylating estrone into 2-OHE, Fig 4.

CYP1A2 is also involved in the metabolism of coffee, and coffee is used as a probe when evaluating the activity of this enzyme *in vivo* (Vistisen et al., 1991). The inducibility of the enzyme is affected not only by coffee (Djordjevic et al., 2008) but also by smoking, when carrying a certain *CYP1A2* genotype denoted *CYP1A2*1F A/A (Sachse et al., 1999). *CYP1A2*1F is located in intron 1 of the *CYP1A2* gene. This genetic region is highly conserved between species (Ikeya et al., 1989), and may therefore be of importance for survival.

The *CYP1A2*1FC/C genotype has been associated with a lower urinary 2-OHE/16αOHE1 ratio (Lurie et al., 2005). Coffee has also been associated with increasing 2-OHE/16αOHE1 in healthy young women, but not in women using exogenous hormones (Jernström et al., 2003b). Increased 16αOHE1 levels have also been associated with an increased risk of postmenopausal breast cancer in women not using HRT (Modugno et al., 2006).

Study II evaluated whether coffee consumption in combination with the *CYP1A2*1F genotype affects the estrogen metabolite profile in breast cancer patients. This study was also the first to explore the 2-OHE/16αOHE1 ratio in both pre- and post-operative samples from the same woman. Study III elucidated whether the combination of coffee and *CYP1A2*1F is also associated with prognostic factors, including age at diagnosis and ER status.
Study IV

CYP2C8 contains a number of SNPs including CYP2C8*3 (Dai et al, 2001) and CYP2C8*4 (Bahadur et al, 2002). CYP2C8*3 includes both the amino acid substitutions, Arg139Lys and Lys399Arg, and has been associated with defective metabolism of paclitaxel and of arachidonic acid in vitro (Dai et al, 2001). These genetic polymorphisms are highly linked. We first analyzed 300 samples for both CYP2C8*3 alleles, but since the concordance rate was 100% we decided to continue with only one of these polymorphisms, i.e. Arg139Lys. In addition, CYP2C8*4 has been associated with a non-significantly decreased paclitaxel turnover as compared with wild type in human liver microsomes (Bahadur et al, 2002).

CYP2C8 and CYP2C9 are responsible for the metabolism of approximately 20% of clinically used drugs, including tamoxifen (Jin et al, 2005). The two major SNPs in the CYP2C9 gene are CYP2C9*2 (King et al, 2004) and CYP2C9*3 (Sullivan-Klose et al, 1996). CYP2C9*2 and CYP2C9*3 have both been associated with impaired enzyme activity in vivo (King et al, 2004). CYP2C9*3 is located in the substrate binding site of the enzyme (Gotoh, 1992).

A linkage between the CYP2C8*3 and the CYP2C9*2 allele has been reported in the Swedish population, where ~96% of CYP2C8*3 allele carriers also had a CYP2C9*2. In addition, 85% of CYP2C9*2 carriers also had a CYP2C8*3 (Yasar et al, 2002). The linkage between these two polymorphisms has also been reported in a Finnish population (Hilli et al, 2007) and a Spanish population (Dorado et al, 2008).

Glutathione S-transferases (GSTs) are Phase II enzymes that conjugate glutathione to endogenous and exogenous electrophiles, making them more water-soluble and easier to excrete (Ketterer, 1988). GSTM1*0 and GSTT1*0, or the null allele, refers to the lack of enzyme function as a result of deletion of the entire gene, Fig 5A and B, respectively. The GSTM1*0 is present in approximately 50% of Caucasians and GSTT1*0 in almost 20% of Caucasians (Garte et al, 2001). Approximately 10% of Caucasians have the homozygous double deletion (GSTM1*0/GSTT1*0) (Garte et al, 2001). Breast cancer patients with GSTM1*0 and GSTT1*0 deletions have a reduced risk of disease recurrence and death (Ambrosone et al, 2001). The GSTT1*0 has been associated with an increased risk of early-onset breast cancer and the common normal variant (wild type) allele with poor response to chemotherapy. GSTT1*0 was also associated with increased disease-free survival, especially in carriers of the double deletion (Khedhaier et al, 2003).

**Figure 5A** GSTM1 is embedded in a region with extensive homologies and flanked by two almost identical 4.2-kb regions (gray boxes). The GSTM1 null allele arises by homologous recombination of the left and right 4.2-kb repeats, which results in a 16-kb deletion containing the entire GSTM1 gene (bottom of diagram) (Parl, 2005). Reprinted from Cancer Letters, Vol 221, Parl FF, Glutathione S-transferase genotypes and cancer risk, 123-129 (2005), with permission from Elsevier.
Information on the genetic polymorphisms being studied and their characteristics are summarized in Table I.

<table>
<thead>
<tr>
<th>Genetic polymorphisms</th>
<th>rs number</th>
<th>Base</th>
<th>Amino acid</th>
<th>Repeat length</th>
<th>Location</th>
<th>USCS*</th>
<th>MAF**</th>
<th>SNP EU</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF1</td>
<td>---</td>
<td>(CA)n</td>
<td>---</td>
<td>11-24</td>
<td>12q23.2</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP1A*1F</td>
<td>rs762551</td>
<td>C-163A</td>
<td>---</td>
<td>15q24.1</td>
<td>A 0.708</td>
<td>C 0.292</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2C8*3</td>
<td>rs11572080</td>
<td>G416A</td>
<td>Arg139Lys</td>
<td>---</td>
<td>10q24</td>
<td>A 0.108</td>
<td>G 0.892</td>
<td></td>
</tr>
<tr>
<td>CYP2C8*4</td>
<td>rs1058930</td>
<td>C792G</td>
<td>Ile264Met</td>
<td>---</td>
<td>10q24</td>
<td>C 0.942</td>
<td>G 0.058</td>
<td></td>
</tr>
<tr>
<td>CYP2C9*2</td>
<td>rs1799853</td>
<td>C416T</td>
<td>Arg144Cys</td>
<td>---</td>
<td>10q24</td>
<td>C 0.896</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2C9*3</td>
<td>rs1057910</td>
<td>A1061C</td>
<td>Ile359Leu</td>
<td>---</td>
<td>10q24</td>
<td>A 0.942</td>
<td>C 0.058</td>
<td></td>
</tr>
<tr>
<td>GSTM1*0</td>
<td>---</td>
<td>---</td>
<td>1p13.3</td>
<td>50%***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSTT1*0</td>
<td>---</td>
<td>---</td>
<td>22q11.23</td>
<td>20%***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*USCS Genome Bioinformatics website: http://genome.ucsc.edu/
***Garte et al, 2001

Genetic polymorphisms that were genotyped but not included

CYP3A4 is involved in the metabolism of multiple clinically used drugs (Läkemedelsindustriföreningen Service AB, 2007) including the drugs tamoxifen (Jin et al, 2005) and anastrozole (Grimm et al, 1997). The metabolic polymorphism CYP3A4*1B is an A-to-G transition at position -290, (rs2740574) (Rebbeck et al, 1998). We have genotyped 653 patients for this polymorphism. Since the activity of this enzyme is modified by several factors large sample sizes are required, and we have therefore not yet included this polymorphism in any studies.

The Insulin-like Growth Factor Binding Protein-3 (IGFBP-3) is the main binding protein of IGF-1. The IGFBP3 gene contains a polymorphism denoted (A-202C), (rs2854744). The A allele has not only been associated with increasing IGFBP-3 levels (Al-Zahrani et al, 2006; Jernström et al, 2001b) but also with decreased breast cancer risk (Al-Zahrani et al,
This polymorphism was also evaluated in a subset of 508 patients. However, it was not statistically significantly associated with tumor characteristics.

SULT1A1 is a Phase II enzyme that is involved in the elimination of active metabolites of tamoxifen (Nowell et al, 2002). The genetic variation, SULT1A1*2 (rs9282861), has been associated with half the sulphation of the anti-estogenic metabolite 4-hydroxy-tamoxifen as compared with the SULT1A1*1 common wild type allele. The risk of death has also been reported as being higher in tamoxifen treated patients homozygous for the SULT1A1*2 allele, whereas no difference in Hazard Ratio (HR) was seen between patients not treated with tamoxifen, regardless of SULT1A1 status (Nowell et al, 2002). However, others reported opposite results (Wegman et al, 2005) or no effect (Wegman et al, 2007). To date, we have genotyped 400 patients for SULT1A1*2.

The GSTPI gene contains several polymorphisms, including Ile105Val, denoted GSTPI*B (rs1138272), and Ala114Val, denoted GSTPI*C (rs1799811). Since the expression of the GSTPI gene is also regulated by methylation (Jhaveri et al, 1998) we have not yet included these polymorphisms in our studies.
**Aims**

To study:

**Study I**
- whether absence of the common CA repeat in the *IGF1* gene (*IGF1*-19/-19) is more common in younger patients than in older patients or is associated with an earlier age at diagnosis after multiparity

**Study II**
- factors that may influence the estrogen metabolite profile, 2-OHE/16αOHE1, in blood samples taken pre- and post-operatively from the same woman

**Study III**
- whether coffee consumption or *CYP1A2*1F are associated with ER status or age at diagnosis
- whether coffee consumption differs according to *CYP1A2*1F genotype in breast cancer patients

**Study IV**
- the frequency of the genetic polymorphisms *CYP2C8*3, *CYP2C8*4, *CYP2C9*2, *CYP2C9*3, *GSTM1*0 and *GSTT1*0
- any linkage, and to construct *CYP2C8/CYP2C9* haplotypes
- whether these genetic polymorphisms (mainly in Phase I and Phase II metabolizing genes) are associated with tumor characteristics and early recurrences
Breast Cancer (BC) Blood Study

All papers in this thesis are based on the Breast Cancer (BC) blood study. Since this is a blood study the genetic polymorphisms that are analyzed are all germline variations that are present in all normal cells containing DNA, and these normal genetic variations are inherited.

The BC blood study was designed as a prospective cohort study to enable studies of the joint effect of lifestyle and genetic polymorphisms on breast cancer treatment response. Data including blood samples and clinical information are collected pre-operatively and on multiple occasions post-operatively. This design has enabled us to also use this material to study gene-environment interactions in a case-only manner. Similarly, we can study the effects of genetic polymorphisms and lifestyle on tumor characteristics.

Study Design

The case-only study design is a non-traditional epidemiological approach to evaluate gene-environment interactions (Khoury et al., 1996). There are a number of methodological issues that must be taken into consideration with this study design. First of all, patients should be selected carefully, as with any study design. Optimally, the cases are population-based, which increases the generalizability of findings. Moreover, genotype and exposure should not be interdependent. When data extracted from a case-only study is evaluated it is important to realize that the results are the combined effect of gene and exposure, and not the effect of these factors alone. In addition, the genetic polymorphism being studied may not be the functional variant. The polymorphism may thus be a marker, in LD with the functional variant. Finally, the measures should be interpreted as departures from multiplicative effects.

In this case-only cohort a candidate gene/polymorphism approach has been applied. The function of the given gene or polymorphism must first be evaluated when
using the candidate gene or candidate genetic polymorphism approach (Rebbeck et al., 2004). This evaluation includes not only the effect of a given polymorphism on the nucleotide sequence but also the importance of the genetic region being studied. A polymorphism in an evolutionary conserved region may be of major interest, since these genetic regions could be important for the survival of the species. In addition, experimental studies must be consistent. The effect in humans must be evaluated and the effect at the target site must be known. The genetic polymorphisms must be proven to effect metabolism in the tissue, e.g. the breast. Moreover, the frequency of the genetic polymorphisms being studied must be evaluated, since deviations from the expected genotype frequencies in healthy controls may imply that the genetic polymorphism may be associated with an increased breast cancer risk (Rebbeck et al., 2004). However, since all studies in this thesis have been performed in a case-only cohort, risk evaluation should be carefully interpreted (Khoury et al., 1996). Finally, the associations from epidemiological studies must be reproducible. Results from a candidate gene approach study are interpreted based on current knowledge – which may both be a strength and a limitation.

Another study design is the whole genome wide association (GWA) study. This study design generates a huge amount of information. In order to obtain reliable results when performing multiple statistical comparisons, a large number of patients needs to be included (Wang et al., 2005). The major limitation when evaluating the results of these GWA studies is the restricted amount of available information on exposure or lifestyle (Le Marchand et al., 2008).

Study Subjects

Women assessed pre-operatively at Lund University Hospital and Helsingborg Hospital, Sweden for a first breast cancer were invited to take part in an ongoing study regarding genetic and non-genetic factors that could be associated with breast cancer prognosis. Data included in this thesis is from patients who were enrolled between October 2002 and October 2007 in Lund and between April 2006 and October 2007 in Helsingborg. Women were invited to participate regardless of ethnic background, age and stage. Patients who had been diagnosed and treated for another type of cancer during the past ten years were not eligible for participation, and women who underwent diagnostic breast surgeries and were found to have breast cancer were not included either. The study was approved by the Ethics Committee of Lund University. Written informed consents were collected during the pre-operative visit at the Department of Surgery at Lund University Hospital and Helsingborg Hospital. At the same visit the research nurse collected blood samples (EDTA plasma and serum) and recorded the time and date of sampling. Serum, EDTA plasma and blood
cells were stored at –70°C. All samples were labeled with serial codes to enable blind-
ed analyses.

Body measurements and breast volumes were measured at the pre-operative visit. All patients filled out a pre-operative questionnaire, including questions on date of birth, coffee consumption, smoking, alcohol intake, use of exogenous hormones and concomitant medications, reproductive history and family history of cancer. There were no questions on ethnicity, though the vast majority of women included were ethnic Swedes.

Patients who had not had a menstrual period during the last year were defined as postmenopausal, though postmenopausal patients who used HRT may have experienced HRT-induced bleedings and may therefore have been misclassified as pre-menopausal. Patients who had had their uterus removed prior to menopause but not their ovaries may also have been misclassified as postmenopausal. We therefore classified patients according to age (<50 years or ≥50 years) instead of reported menopausal status.

Additional preoperative information, including type of surgery, sentinel node biopsy and axillary node dissection, was obtained from each patient’s chart. Tumor size, histological type and grade, axillary node involvement, signs of distant metastases, ER and PR status were obtained from each patient’s pathology report. HER2 status was routinely analyzed as of November, 2005.

As of January 2008 the median follow-up time was 14 months (range 0-62 months). Forty four patients experienced breast cancer-related events. Of these 44 patients, 18 had received tamoxifen, 13 polychemotherapy, six aromatase inhibitors, five neo-adjuvant therapy and three patients interstitial laser thermotherapy prior to recurrence.

Lund University Hospital is one of nine hospitals in the South Swedish Health Care Region performing breast cancer surgery. Helsingborg Hospital is approximately 55km north of Lund. The Lund University Hospital catchment area serves almost 300,000 inhabitants, and Helsingborg Hospital serves another 250,000 inhabitants. Breast cancer patients are not referred to other hospitals for surgery. We consider this study to be population-based, and patients between ages 21 and 99 years have been included. According to data from the Regional Tumor Registry, as of 25 June 2008 a total of 6,765 primary female breast cancers were registered between 1 October 2002 and 31 October 2007 in the South Swedish Health Care Region, Table II. During the same period 893 breast tumors were registered in Lund, of which 866 were primary and received surgery. Five hundred and seventeen of them (60%) were included in our study. Between 1 April 2006 and 31 October 2007 330 breast tumors were registered in Helsingborg, of which 298 were primary and received surgery. One hundred and thirty six of them (46%) were included in our study. We only included women who attended the pre-operative visit and who had not been diagnosed with other cancers during the past 10 years.
Table II. Age at breast cancer diagnosis for the patients and ER and PR status for the tumors registered in the South Swedish Health Care Region, Lund and Helsingborg, and for the patients included in Study IV from Lund and Helsingborg as of 25 June 2008.

<table>
<thead>
<tr>
<th>Date (MM/DD/YYYY)</th>
<th>South Swedish Health Care Region</th>
<th>Lund</th>
<th>BC Study in Lund</th>
<th>Helsingborg</th>
<th>BC Study in Helsingborg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10/01/2002-10/31/2007</td>
<td>6,765</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/01/2002-10/31/2007</td>
<td>866</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/01/2002-10/31/2007</td>
<td>517</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>04/01/2006-10/31/2007</td>
<td>330</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>04/01/2006-10/31/2007</td>
<td>136</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of patients included</td>
<td>6,765</td>
<td>866</td>
<td>517</td>
<td>330</td>
<td>136</td>
</tr>
<tr>
<td>Mean age at diagnosis, years</td>
<td>63.5</td>
<td>59.9</td>
<td>59.0</td>
<td>64.4</td>
<td>58.5</td>
</tr>
<tr>
<td>Registered ER status, %</td>
<td>79</td>
<td>91</td>
<td>98</td>
<td>89</td>
<td>96</td>
</tr>
<tr>
<td>ER+ tumors, %</td>
<td>84</td>
<td>85</td>
<td>86</td>
<td>85</td>
<td>89</td>
</tr>
<tr>
<td>Registered PR status, %</td>
<td>78</td>
<td>89</td>
<td>98</td>
<td>89</td>
<td>96</td>
</tr>
<tr>
<td>PR+ tumors, %</td>
<td>67</td>
<td>68</td>
<td>68</td>
<td>59</td>
<td>68</td>
</tr>
</tbody>
</table>
Methods

The method used to determine a genetic polymorphism depends on the number of samples and the number of genetic polymorphisms to be analyzed. The BC blood study is an ongoing study, with new patients being included every week. This far over 750 patients have been included, and a dozen polymorphisms have been genotyped. All genetic analyses have been performed using the techniques currently practised in the national oncogenetic laboratory for BRCA1 and BRCA2 mutation screening at the Department of Oncology, Lund University, Sweden.

DNA Extraction

DNA is extracted from peripheral blood using Wizard, Genomic DNA Purification Kit (Promega, Madison, WI, USA), in accordance with the manufacturer’s instruction.

Polymerase Chain Reaction (PCR)

The genetic region of interest is amplified through PCR. The primers are modified to enable further analysis: biotinylated, fluorescently labeled or PIG-tailed (Brownstein et al., 1996). The amplified region varies in size from approximately 100 bases up to ~500 bases, depending on the technique used in the following analysis.

Absence of the GSTM1*0 and GSTT1*0 deletions, i.e. the normal variants (wild types), was evaluated in a multiplex reaction where β-globulin was used as a positive PCR control (Chacko et al., 2005). Briefly, the PCR primers were designed to amplify the wild type allele. β-globulin was included in the multiplex reaction to facilitate the conclusion that absence of a PCR product is the result of a deletion and not a non-functional PCR.

The limitation of this assay is that it identifies any wild type allele and does not distinguish the homozygote wild type from the heterozygous wild type (Parl, 2005).
Gel Electrophoresis

The PCR products are separated through polyacrylamide gel electrophoresis (PAGE). The size of the PCR product is confirmed through PAGE, and contaminations are detected if present.

In Study IV wild type GSTM1 and GSTT1 were evaluated by separating the PCR products on a 2% agarose or a 7.5% polyacrylamide gel (patient sample 1-355), Fig 6. Using the available agarose gel systems in the lab, 88 samples can be separated in one run, but the quality of the wells was not optimal, and we therefore switched to the acrylamide gel system. The acrylamide gels only contain 15 wells, thereby limiting the number of samples that can be separated in one run. Since more than 650 samples were finally analyzed, this was not an efficient method.

Figure 6  An acrylamide gel electrophoresis picture illustrating the size separation of the multiplex PCR product containing GSTM1, GSTT1 and β-globulin. Each lane represents either a patient sample or a control, i.e. size marker (lanes 7 and 8). Lanes 1 and 6 represent wild type for both GSTT1 and GSTM1. Lane 2, 4, and 5 represent wild type for GSTT1 and null for GSTM1. Lane 3 represents null for GSTT1 and wild type for GSTM1. No lane represents null for both GSTT1 and GSTM1 as there was no sample on this gel with only β-globulin.

High Performance Liquid Chromatography (HPLC)

In Study IV, absence of the two deletions, i.e. GSTM1 and GSTT1 wild type, were analyzed by reverse-phase ion-pair HPLC (patient samples 356-659), using the Transgenomic WAVE System (Transgenomic Ltd., Glasgow, UK). The PCR products, containing the β-globulin fragment and possibly a GSTM1 and a GSTT1 wild type fragment, are separated according to size. Samples are loaded onto a column – a DNASep Cartridge packed with C-18 alkylated polystyrene-divinylbenzene polymeric beads. Through the positively charged ion-pairing reagent triethylammonium acetate (TEAA) (Transgenomic Ltd) the negatively charged backbone of the DNA interacts with the hydrophobic matrix. By increasing the organic mobile phase con-
sisting of acetonitrile (Transgenomic Ltd) the DNA is eluated in a size dependent manner. The separation of the double-stranded DNA fragments takes place under non-denaturing conditions at 50°C. The DNA fragments are detected through UV, and the results are visualized using the Navigator software (Transgenomic Ltd). In the electropherogram the absorbance is plotted against time, Fig 7.

Using this technique approximately one percent of the product’s size can be resolved. This technique is straightforward, with limited hands-on time. However, each sample is analyzed for ~10 minutes in the machine, and a 96-well sample plate will therefore take approximately 17 hours to genotype.

![Electropherogram](image)

**Figure 7** An electropherogram illustrating the size separation of the multiplex PCR product containing GSTM1, GSTT1 and β-globulin. From the left, the first peak is the injection peak. Then follows (A) GSTM1 wild type, β-globulin, and GSTT1 wild type, (B) GSTM1 null, β-globulin, GSTT1 wild type, (C) GSTM1 wild type and GSTT1 null. No electropherogram shows the GSTM1 null and GSTT1 null genotype.

**Fragment Analysis**

Fragment analysis is a high resolution separation method that can distinguish between fragments that only differ in length by a single base. The fluorescently labeled PCR products are mixed with an internal size standard, Rox 500 (Applied Biosystems, Foster City, CA, USA), that provides fragments of varying sizes. The sample is then denatured and the single-stranded PCR products are separated through a matrix in an electric field according to size by capillary gel electrophoresis using the ABI3100, or after the upgrade in April 2006 the 3130xl Genetic Analyzer (Applied Biosystems). The results of the fragment analysis are evaluated using the Gene scan software (Applied Biosystems). The number of repeats is determined by the size of
the fragment, which is illustrated as a peak in the electropherogram, where the numbers of bases are plotted against fluorescence intensity, Fig 8.

The lengths of the DNA fragments are manually evaluated by comparing the repeat lengths with the sequenced control samples containing fragments of known lengths. After the upgrade of the system, samples were also automatically evaluated using the GeneMapper Software v.4.0 (Applied Biosystems). In Study I fragment analysis was used to evaluate the length of the IGF1 CA repeat. The number of CA repeats may be underestimated when evaluating samples against the internal size standard (as compared with sequencing) (Rodriguez et al, 2006). Our approach was to use sequenced samples of varying lengths as positive controls in the fragment analysis.

![Figure 8](image)

**Figure 8** An electropherogram illustrating the size separation of the PCR product containing the IGF1 CA repeat. A patient sample that is (A) homozygous for the 19 CA repeat allele, *IGF1*+19/+19 (B) heterozygote for the 18 and a 20 repeat allele, i.e *IGF1*-19/-19.

### Sequencing

The sequencing reaction is performed on amplified and purified DNA using the Big Dye Terminator Cycle Sequencing Kit v. 3.1 (Applied Biosystems) in accordance with the manufacturer’s instructions. The cycle sequencing reaction involves random incorporation of normal unlabeled nucleotides (dNTPs) and different fluorescently labeled dideoxy nucleotides (ddNTP). When a ddNTP is randomly incorporated by the polymerase the sequence is terminated, resulting in a pool of fragments of variable size and with fluorescent labeling. The fragments are then separated according to size by means of capillary gel electrophoresis and are detected us-
ing the laser fluorescence detection system of an ABI3100 or 3130xl Genetic Analyzer (Applied Biosystems). The results are analyzed using the Sequencher software, current version 4.5 (Applied Biosystems), Fig 9.

In Study II and Study III CYP1A2*1F was genotyped using the sequencing reaction. CYP3A4*1B, IGFBP3 (A-202C) and SULT1A1*2 were additionally genotyped using this method, even though the results are not presented in this thesis. Moreover, all the different genotypes that were found in the pyrosequencing (PSQ) reaction were confirmed by sequencing.

Sequencing is a reliable technique when evaluating DNA extracted from blood, though it is not a time-effective method for analyzing a single base substitution. The hands-on time is more extensive than with PSQ. However, some of the protocols were optimised in the lab before the PSQ technique was set up and the IGFBP3 (A-202C) was not successfully genotyped using PSQ.

Figure 9  Sequence results for CYP1A2*1F A patient sample that is (A) homozygous for the A allele (B) homozygous for the C allele or (C) heterozygous, i.e. A/C.

Pyrosequencing (PSQ)

The genetic region of interest is amplified through PCR, where one of the primers is biotinylated. PSQ is performed on amplified genomic DNA, using the PyroGold PSQ kit (Biotage AB, Uppsala, Sweden). The double-stranded DNA is denatured using sodium hydroxide, and separated into single-stranded DNA through biotin streptavidin interactions. The DNA is neutralized and then annealed with a sequencing primer. During the reaction nucleotides complementary to the template are incorporated one at a time by the DNA polymerase, resulting in formation of equivalent amounts of pyrophosphate (PPI). In the presence of adenosine 5′ phosphosulphate (APS) PPI is converted into adenosine 5′ triphosphate (ATP) by the enzyme ATP sulphurylase. ATP is then used in the conversion of luciferin into oxyluciferin, whereby visible light is formed equivalent to the amount of ATP generated. Light is detected using a charge coupled device (CCD) camera, and is ob-
served as a peak in the pyrogram, Fig 10. Unincorporated nucleotides during the reaction and ATP are degraded by the enzyme apyrase.

![Diagram of Pyrosequencing Technology](image)

**Figure 10** The Principle of Pyrosequencing® Technology. Reprinted with permission from Biotage AB.

The PSQ technique was used to evaluate the following SNPs: CYP2C8*3, CYP2C8*4, CYP2C9*2 and CYP2C9*3. We also used this technique for GSTP1*B. We attempted to analyze IGFBP3 (A-202C) using this method, but without success. This may be explained by the high GC content of this fragment (~75%), which may cause secondary structures. The fact that the PSQ reaction is carried out at 28°C supports this theory. Secondary structure formation can be prohibited by the addition of single-stranded binding protein in the PSQ reaction. This protein is now included in the reagent solutions. We were not, however, able to successfully genotype IGFBP3 (A-202C).

The major advantage of PSQ is that once the assay has been optimized it is a very time-effective technique. The hands-on time is limited, and in the PSQ run a 96-well sample plate can be analyzed within 10 minutes, i.e. ~1 minute/ incorporated base. The major drawback with PSQ is not in the technique itself, but rather in the preparatory PCR, where up to 50 cycles are required. When running 50 cycles nu-
merous copies of the template are generated, and this also increases the risk of contamination.

Other SNP Genotyping Methods and Limitations

In addition to the SNP genotyping assays described above there are other techniques, including the TaqMan-based allelic discrimination, the Illumina Golden Gate or Infinium Assay and the MassEXTEND assay.

The limitation of using SNPs as a genetic marker is the limited information (homozygous for the wild type or variant allele or heterozygous) that is received. Moreover, even if a patient is heterozygous for a given allele it is not guaranteed that both alleles are expressed, e.g. allele-specific expression (Minelli et al., 2005). The genomic ratio and the mRNA ratio are not always correlated. This may be the result of a SNP in a transcription factor binding site, in a transcript that affects mRNA folding and stability, in miRNA binding sites, or in a methylation site.

Enzyme Immunoassay

The levels of 2-hydroxyestrogens (2-OHE), i.e. 2-hydroxyestrone + 2-hydroxyestradiol + 2-hydroxyestriol and 16α-hydroxyestrone (16αOHE1), were measured in EDTA plasma by means of a monoclonal antibody-based enzyme assay (ESTRAMET™ 2 and 16, Immuna Care Inc, Bethlehem, PA). The competitive inhibition enzyme immuno assays (EIAs) for 2-OHE and 16αOHE1 in serum and plasma were developed from reagents and methods previously reported for measurement of metabolites in urine (Bradlow et al., 1998; Falk et al., 2000; Klug et al., 1994; Ziegler et al., 1997).

The sensitivity of this technique is less than 40 pg/ml serum for the 2-OHE and 16αOHE1. Within assays the variability of duplicates for positive control sera (150-560 pg/ml) is <10% for 2-OHE and 5% for 16αOHE1. Between assays variability is <15% for both assays.
Immunohistochemistry (IHC)

Tumors were analyzed at the Departments of Pathology at Lund University Hospital and Helsingborg Hospital in Study IV. In Lund, ER and PR status were determined by means of immunohistochemistry (IHC) using the Dako LSAB™ kit system (Dako, Glostrup, Denmark) and the antibodies M7047 (ER) and M3569 (PR) (Dako, Glostrup, Denmark). In Helsingborg ER and PR status were determined by means of IHC using the Ventana ultra view kit (760-500) (Ventana, Illkirch, France) and antibodies 790-4324 (ER) (Ventana, Illkirch, France) and NCL-L-PGR-312 (PR) (NovoCastra, NewCastle, UK). Tumors with more than 10% of positive nuclear staining were considered ER-positive or PR-positive. Receptor-negative tumors had a positive nuclear staining of 10% or less.

Validation

We attempted to reanalyze 25% of samples in separate PCR reactions using the same genotyping method. In Study I the first 352 samples were run in duplicate, in separate PCR and fragment-analysis runs. Thereafter every fourth sample was run in duplicate, and the concordance rate was 98.3%. The differences were not in the technique itself but in the manual selection of peaks in the electropherogram. In Study II all samples were analyzed for the estrogen metabolites in duplicate. In Study II and Study III one in four samples were validated for \textit{CYP1A2*1F}, and the concordance rate was 100%.

In Study IV the concordance rate was 100% for \textit{CYP2C8*3}, \textit{CYP2C8*4} and \textit{CYP2C9*3}, and for \textit{GSTM1} and \textit{GSTT1} samples size-separated using gel electrophoresis. However, the concordance rate was 97.9% for \textit{GSTM1} and \textit{GSTT1} when validating samples with HPLC that were size separated with HPLC. The concordance rate was 97.8% when validating 45 samples by means of HPLC that had earlier been evaluated using gel electrophoresis (n=355). For \textit{CYP2C9*2} the concordance rate was 97.2%. The PSQ software sometimes determined a homozygous reference sample as being heterozygous. When re-evaluating all heterozygous samples we conclude that they are heterozygous and not homozygous. The remaining discrepancy was due to a human error and not a problem with the technique \textit{per se}. This underscores the importance of validation with independent samples, preferably with newly extracted DNA from a new blood sample.

All different genotypes that were found using PSQ were confirmed by means of sequencing, i.e. reference samples in the PSQ reaction representing the different genotypes (when identified) were confirmed.
The concordance when determining ER and PR status between different pathology departments by different pathologists in Sweden is good (kappa=0.78 and kappa=0.72, respectively) (Chebil et al, 2003).

Haplotype Construction

The haplotypes in Study IV were constructed by setting up cross tables with each genotype, with its three possible combinations against each other. Using this approach we found that there were combinations that were not present in our study population.

Statistical Methods

Data analysis was performed using the statistical software SPSS 12.0 and 13.0 for Windows (SPSS Inc. Chicago, Illinois, USA) and SAS 9.1 (Cary, NC, US).

Two-group comparisons were performed using Student’s T-test or the non-parametric Mann-Whitney U-test. The latter was used when the sample size was small or when the underlying assumption of normality was not reasonably well fulfilled. The Spearman rank correlation ($r_s$) was used as a measure of dependence between two continuous variables and the Chi-square test was used to evaluate associations between discrete variables. Fisher’s exact test was used when less than five observations were expected in one or more cells of a contingency table.

Linear regression models were used to evaluate the association between genetic polymorphisms and lifestyle factors on continuous dependent variables. In multivariate linear regression models the dependent variable was adjusted for potential confounding factors.

Logistic regression models were used to evaluate the associations between genetic polymorphisms and lifestyle factors on a dichotomous dependent variable.

To test for gene-environment interactions, an interaction variable was calculated between the lifestyle factor and the genetic polymorphism of interest by multiplying these two variables. The interaction variable thus represents a combination of the two factors. The magnitude of the interactions was presented as $\exp(\beta)$ (i.e. Odds Ratio (OR)) in the logistic regression models and as $\beta$ in the linear regression models.

In Study I an interaction variable between multiparity and $IGF1$-19/-19 was calculated, and in Study III an interaction variable between coffee consumption and $CYP1A2*1F$ was calculated. When these two interactions were evaluated in relation to age at diagnosis reported as a continuous variable (years), a linear logistic model was
used. When age at diagnosis was dichotimized (<50 years versus ≥50 years), a logistic regression model was used.

The effect of the interaction variable was calculated using an exact logistic regression procedure (SAS), when there were no observations in a subgroup of patients. This was the case in Study III, where no patient with the CYP1A2*1F/A/A genotype and a low coffee consumption had an ER-negative tumor.

Kaplan-Meier estimates, log rank tests (unadjusted) and Cox regression models (adjusted) were used in Study IV to examine the association between genotype and breast cancer free survival.

All P-values were calculated based on two-tailed statistical tests. The significance level was set at 5%, i.e. a P-value <0.05 was considered to be statistically significant. By definition, one in 20 statistical tests performed under the null hypothesis will have a P-value <0.05, by chance. We have not adjusted for multiple testing. When using the Bonferroni correction the P-value obtained should be divided by the number of statistical tests performed. For example, for a study including 20 statistical tests a P-value needs to be lower than 0.0025 to be statistically significant.

When performing a large number of tests the chance of false positive results or Type I errors increases with the number of tests performed. In Study IV many statistical tests were performed. Not only one SNP but also a panel of polymorphisms as well as haplotypes were tested for prognostic factors and early recurrences. Based on the number of tests performed, the risk of a false positive result is evident. The results of Study IV should therefore be carefully interpreted, and should be considered hypothesis generating.

A false negative result or Type II error is most often the result of too small a study size. The results of the pilot study should therefore be carefully interpreted, since we might not have the statistical power to detect small but potentially important difference between two subgroups of patients.
Discussion of Results

Study I

IGF1-19/-19 distribution by age

We found that absence of the common 19 CA repeat allele (IGF1-19/-19) was more frequent in patients diagnosed prior to the age of 45 years and in patients diagnosed at the age of 55 to 59 years.

The combined effect of IGF1-19/-19 and multiparity on age at breast cancer diagnosis

We also found that there was an interaction between the IGF1-19/-19 genotype and multiparity on age at breast cancer diagnosis. The mean age of breast cancer onset was 53.1 (SD±9.5) years in multiparous patients with IGF1-19/-19, as compared with 59.0 (SD±11.5) years in all other patients, (P_{interaction} = 0.007) (see Table III).

Table III. Age at breast cancer diagnosis in relation to parity and IGF1 genotype. There was an interaction between multiparity and the IGF1-19/-19 genotype on age at breast cancer diagnosis.

<table>
<thead>
<tr>
<th>Multiparity</th>
<th>IGF1+19</th>
<th>IGF1-19/-19</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1 children</td>
<td>58.0 (±12.1)</td>
<td>62.3 (±10.9)</td>
</tr>
<tr>
<td>n = 115</td>
<td>n = 13</td>
<td></td>
</tr>
<tr>
<td>2+ children</td>
<td>59.3 (±11.3)</td>
<td>55.1 (±9.5)</td>
</tr>
<tr>
<td>n = 236</td>
<td>n = 39</td>
<td></td>
</tr>
<tr>
<td>P = 0.33</td>
<td>P = 0.006</td>
<td>(P_{interaction} = 0.007)</td>
</tr>
</tbody>
</table>

IGF1-19/-19 modified the association between multiparity and age at breast cancer diagnosis. Table III illustrates the effect of a gene-environment interaction. Multiparity was only associated with an earlier age at diagnosis in patients with the IGF1-19/-19 genotype. If we had not stratified patients according to IGF1 genotype this observation would have been missed. Age at breast cancer diagnosis was not further decreased with an increasing number of children from 2 to 5 in patients with IGF1-19/-19 genotype; \(r_s=-0.01\); \(P=0.95\), per additional child.
As shown in Fig 11A, IGF1-19/-19 is more frequent in the younger multiparous patients as compared with the nulli- or uni-parous patients.

In addition to multiparity, the high frequency of IGF1-19/-19 in the younger patients, irrespective of parity, could be associated with BRCA1 status. BRCA1 mutation carriers often present with an early disease (Couch et al, 1997; Narod et al, 2004; Shattuck-Eidens et al, 1997), and the risk for early-onset breast cancer further increases with parity (Jernström et al, 1999). Our group has reported on an association between IGF1-19/-19 and BRCA1 status (Henningsson et al, 2007; Jernström et al, 2005b). BRCA1 and BRCA2 mutations are present in approximately 9% of all breast cancer cases diagnosed ≤40 years of age in southern Sweden (Loman et al, 2001). In the present study we have not screened for BRCA1 mutations, which is a limitation. However, IGF1-19/-19 was not associated with either a history of breast cancer and/or ovarian cancer in a first- or second-degree relative in patients diagnosed ≤40 years of age. However, a family history of cancer was self-reported in the questionnaire and was not validated. Some patients reported cancers in the lower abdominal area in relatives, which may or may not be ovarian cancers, and the total number of ovarian cancers may therefore be underestimated.

IGF1-19/-19 was also more frequent in patients diagnosed at the ages of 55 to 59 years. The reason for the second peak had not been explored in the present study when published. However, upon re-analysis of our data the second IGF1-19/-19 frequency peak is present in uni-, nulli- and multi-parous patients. A potential explanation may be HRT use (Slattery et al, 2007), though there was no support for this hypothesis in our data. In patients diagnosed ≥50 years of age IGF1-19/-19 was present in 28% of patients who had used HRT and in 20% of patients who had not used HRT, Fig 11B.

Figure 11 A A frequency plot illustrating how the frequency of IGF1-19/-19 varies according to age at breast cancer diagnosis in nulli- and uni-parous patients (light grey bars) and in multiparous patients (dark grey bars). B A frequency plot illustrating how the frequency of IGF1-19/-19 varies according to age at breast cancer diagnosis in patients with prior HRT use (dark grey bars) and without prior HRT use (light grey bars).
We have noticed a mistake in the published paper. We have calculated the relative risk (RR) instead of the OR in two analyses. The first analysis compared the frequency of IGF1-19/-19 in accordance with multiparity and age at diagnosis. In the published paper we stated that the OR of IGF1-19/-19 among the younger multiparous patients as compared with all other patients was OR 1.7 (95% Confidence Interval (CI) 0.96-3.0). The correct OR is 1.9 (95% CI 0.94-3.7). The same analysis was then performed in patients diagnosed prior to the age of 45 years. The published OR was 2.3 (95% CI 1.3-6.5). The correct OR is 2.9 (95% CI 1.2-6.5; P=0.025). The combined effect of IGF1-19/-19 and multiparity was thus somewhat stronger than suggested in our publication.

The protective effect of parity

The protective effect of parity on breast cancer risk, or rather the possible mechanism, has recently been reviewed (Britt et al, 2007), and four different theories are now presented. The first theory concerns the altered levels of circulating hormones, e.g. prolactin and growth hormone, in parous women. Growth hormone exerts its effect through IGF-1, and IGF-1 levels are lower in parous women than in nulliparous women in the general population (Holmes et al, 2002). Since IGF1-19/-19 was associated with an idiosyncratic increase in IGF-1 levels in nulliparous OC users and an increase in breast volumes of parous women (Jernström et al, 2005b), we hypothesized that both IGF-1 levels and the breast tissue in women with this genotype may be differently affected by pregnancies than in women with at least one 19 CA repeat allele. We did not measure the IGF-1 levels in the present study, which is a limitation. A second theory concerns the finding that the breasts of parous women contain more differentiated cells that are less proliferative and therefore less susceptible to hormonal stimuli. The third theory concerns the decreased number of mammary stem cells and thereby the reduced number of cells that are susceptible to transformation. The fourth theory involves the altered estrogen responsiveness of parous glands. It is important to note that these theories are not mutually exclusive.

The functionality of IGF1-19/-19

The effect of not having the common 19 CA repeat allele has not been fully understood. However, this microsatellite has not only been investigated in breast cancer (Fletcher et al, 2005) but also in prostate cancer (Schildkraut et al, 2005), and in a subgroup of colon cancers (Slattery et al, 2005). The inconsistency of findings may be the result of not taking lifestyle factors into account. Since high-throughput techniques of microsatellite screening are limited, researchers are now analyzing tagging SNPs that may explain the altered IGF-1 levels and breast cancer risk (Al-Zahrani et al, 2006). Another approach in searching for genetic polymorphisms that may modify IGF-1 levels is a targeted strategy focusing on genetic polymorphisms in evolutionary conserved regions (Palles et al, 2008).
Early age at diagnosis

It is important to identify women who are at increased risk of early-onset breast cancer, since an early age at diagnosis is often associated with a poorer prognosis (Goldhirsch et al., 2007). Moreover, women who are diagnosed prior to the ages of 40 or 45 years will not have their tumors detected at mammography screening, since they are not routinely invited to screening.

Potential bias

The risk of recall bias when asking about parity is minimal, and the IGF1-19/-19 genotyping was validated with a high concordance rate. Age at breast cancer diagnosis depends on when the tumor was detected. Age at breast cancer diagnosis may be modified by patient or doctor delay. Mammography screening aims at earlier detection. Genetic variations in the IGF1 gene and IGF-1 levels have been studied in relation to breast density (Byrne et al., 2000; Diorio et al., 2008). Women with denser breasts have a lower chance of having their tumors detected through mammography screening, though no one has yet studied the absence of the IGF1 19 CA repeat allele in relation to breast density. In women diagnosed between the ages of 45 and 74 years, when screening is offered, we found no association between IGF1-19/-19 and screening detection, tumor size or spread to the axillary lymph nodes. We have no information on whether or not patients adhered to the screening guidelines, e.g. if the tumors were missed by mammography screening or found in women who did not attend screening. In addition, mammography screening does not detect all tumors.

In conclusion, it has been reported that IGF1-19/-19 modifies the effect of several breast cancer risk factors, and the results of the present study indicate that this may also be the case for multiparity. If our findings are confirmed, multiparous patients with IGF1-19/-19 may represent a subgroup of patients that would benefit from earlier screening. Similarly, women with the same genotype who are not multiparous may be offered later commencement of screening than is currently recommended, since they may be at a below-average risk of early-onset breast cancer.

Studies II & III

Pre- and post-operative 2-OHE/16αOHE1 ratio

In the pilot study (Study II) an increased 2-OHE/16αOHE1 ratio between the pre- and post-operative plasma samples was found in 69.5% of the 59 patients.

Non-genetic factors in relation to the 2-OHE/16αOHE1 ratio

Increasing alcohol consumption (β=0.131; P=0.006), a moderate to high coffee consumption (β=0.155; P=0.03) and concomitant use of tamoxifen and RT (β=0.189;
P=0.006) were all associated with an increase in the 2-OHE/16αOHE1 ratio between the two visits. Even though the results were statistically significant, the clinical relevance needs to be further elucidated. In the present study the metabolites were measured in plasma and not locally in the breast. However, the variation in 2-OHE/16αOHE1 ratio in tissue samples from the same patient varies widely, and may not be a representative measure (Bradlow et al, 2006). Smoking induces the CYP1A2 enzyme (Sachse et al, 1999), but did not seem to influence the plasma 2-OHE/16αOHE1 ratio in this study, which is in line with previous findings (Jernström et al, 2003b).

CYP1A2*1F and other genetic polymorphisms in relation to the 2-OHE/16αOHE1 ratio

The CYP1A2*1F C alleles was associated with a lower plasma 2-OHE/16αOHE1 ratio in samples taken both pre- (r_s=-0.20; P=0.13) and post-operatively in the same patient (r_s=-0.30; P=0.02). This is in line with the findings of others (Lurie et al, 2005). In addition to CYP1A2*1F, other polymorphisms in the estrogen metabolism – COMTVal158Met, CYP1A1*2A, CYP1A1*2B, CYP1B1Val432Leu and CYP17T27C – have also been studied in relation to the 2-OHE/16αOHE1 ratio (Lurie et al, 2005). However, the only SNP associated with the urine 2-OHE/16αOHE1 ratio was the CYP1A2*1FC allele. Since the correlation between plasma and urine metabolite ratios has been evaluated and the correlation was fair (Bradlow et al, 2006), their findings are compatible with ours.

CYP1A2*1F and other genetic polymorphisms in relation to breast cancer risk

The CYP1A2*1F C allele was associated with a lower plasma 2-OHE/16αOHE1 ratio, but it has also been associated with a lower risk of developing breast cancer (Le Marchand et al, 2005). In fact, this was the only SNP associated with breast cancer risk when evaluating a number of SNPs in the estrogen metabolism, including COMTVal158Met, CYP1A1*2A, CYP1A2*1F, CYP1B1Leu432Val, CYP3A4*1B, SULT1A1*2 and the Arg554Lys variant in AHR (a transcription factor for CYP1A1, CYP1A2 and CYP1B1) (Le Marchand et al, 2005). However, among BRCA1 mutation carriers CYP1A2*1F alone was not associated with breast cancer risk, but rather modified the association between coffee consumption and breast cancer (Kotsopoulos et al, 2007). The effect of CYP1A2*1F on breast cancer risk may thus be modifiable.

Coffee consumption according to CYP1A2*1F on breast cancer risk

In Study III patients with CYP1A2*1FA/A were more likely to drink ≥3 cups of coffee/day, OR 1.5 (95% CI 1.0–2.3; P=0.027). Moreover, patients with CYP1A2*1F A/A consumed almost a half cup of coffee more than patients with any C allele (3.3 versus 2.9 cups/day; P=0.008). Coffee consumption did not differ between patients with one or two C alleles (A/C or C/C), (2.9 versus 2.9 cups/day, P=0.84). We do
not know whether the coffee consumption reported is representative of the amount of coffee consumed earlier in life. However, coffee is a modifiable factor, and our results indicate that coffee may be protective in patients carrying at least one CYP1A2*1FC allele. The finding of a lower cancer risk in women with CYP1A2*1F any C allele and coffee has also been reported in BRCA1 mutation carriers (Kotsopoulos et al., 2007). Coffee consumption has also been associated with an increased risk of ovarian cancer in women with CYP1A2*1F A/A (Goodman et al., 2003).

Age at diagnosis and ER status according to CYP1A2*1F

In the subgroup of patients with CYP1A2*1F A/A we found no association between an early age at diagnosis and ER-negative tumors, which was unexpected. The association between an early age at diagnosis (<50 years of age) and ER-negative tumors was limited to patients with at least one CYP1A2*1FC allele, OR 4.2 (95% CI 1.9 – 9.3, P=0.0002). When including all 653 patients we found that the magnitude of this association increased with increasing number of C alleles.

Combination of coffee consumption and CYP1A2*1F on age at diagnosis

Moderate to high coffee consumption (≥2 cups/day) in patients who had never used HRT was associated with a later age at diagnosis than in patients with low consumption (48.0 years versus 57.7 years, P=0.001). Coffee consumption was not associated with age at diagnosis in patients with at least one C allele. The finding of a later age at diagnosis in patients with a higher coffee intake in patients with CYP1A2*1FA/A was in line with our hypothesis that predicted that coffee would slow the growth of ER-positive tumors, as a result of a higher plasma 2-OHE/16αOHE1 ratio and thereby weaker stimulation of the ER.

Combination of coffee and CYP1A2*1F on ER status

In the study by Le Marchand et al. the decreased breast cancer risk associated with the CYP1A2*1FC allele was even stronger in patients with ER-negative and PR-negative tumors (Le Marchand et al., 2005). In line with their findings, we found an association between a higher coffee consumption and a higher proportion of ER-negative tumors in patients with A/A (14.7% versus 0%, P=0.018). However, this may be a false positive finding. In the published data set of 458 patients, no patient with a low coffee consumption had an ER-negative tumor. With the present inclusion of an additional 188 patients, five patients with a low coffee consumption had an ER-negative tumor. This indicates that the interaction between coffee and CYP1A2*1F on ER status was not that robust. At the American Society of Clinical Oncology conference, results from a large prospective study of the risk of caffeine in relation to breast cancer were presented. Caffeine consumption was associated with an increased risk of developing an ER-negative and PR-negative tumor (Ishitani et al., 2008). However, the CYP1A2*1F was not considered in that study.
ER status is classified based on a threshold value, where tumors with >10% ER-positive cells are defined as ER-positive. This is a cut-off value and does not necessarily represent the biology of the tumor. It would be interesting to analyze our results using another threshold. Moreover, even though other forms of the ER have been discovered, only ERα was measured since this is the protein measured today in the clinical setting.

The associations between coffee and age at diagnosis and ER status differed according to CYP1A2*1F. However, even though patients with A/A and any C had different coffee consumption they did not differ in terms of age, weight, length, BMI, breast volume, parity, OC and HRT use.

Genotype considerations
In the present study the CYP1A2*1F genotype is referred to the single base substitution in intron 1. A correct genotyping would involve the sequencing of the entire gene, or genotyping of all known polymorphisms in the CYP1A2 gene (http://www.cypalleles.ki.se/). At present, there is no explanation according to which CYP1A2*1F affects the inducibility of the CYP1A2 enzyme. However, this finding has been shown in a number of studies (Djordjevic et al, 2008; Ghotbi et al, 2007; Sachse et al, 1999). Regulatory levels other than polymorphisms including methylation and miRNA binding sites have now been suggested (Ingelman-Sundberg et al, 2007). Even though the two coffee derivates caffeic acid and chlorogenic acid have been shown to inhibit methylation in vitro (Lee et al, 2006), no methylation or miRNA binding site seems to be present in intron 1 of the CYP1A2 gene (Methyl Primer Express Software v 1.0, Applied Biosystems).

Strengths and limitations
A strength of the pilot study was the study design, whereby estrogen metabolites were measured in the same patient on two different occasions, each women being her own control. This is the first study in which the plasma 2-OHE/16αOHE1 ratio has been evaluated in both the pre- and post-operative samples from the same patient. However, the pilot study is based on a very limited number of patients, and results should be carefully interpreted.

There are several limitations in Study II and Study III. The most obvious is the size of a cup of coffee. How much coffee constitutes a cup? The cut-offs used when evaluating coffee as a potential risk factor vary considerably from study to study (Baker et al, 2006; McLaughlin et al, 1992; Michels et al, 2002; Nkondjock et al, 2006; Phelps et al, 1988; Rosenberg et al, 1985; Stensvold et al, 1994; Vatten et al, 1990). In Study II the cut-off was set at ≥3 cups/day whereas in Study III it was set at ≥3 cups/day and ≥2 cups/day. Another limitation was that we only have specific questions concerning coffee consumption during the last week in our questionnaire. We also lack information on other caffeinated beverages, including tea, soft drinks and chocolate. In addition, we have no specific information on intake of cruciferous vegetables that may influence the 2-OHE/16αOHE1 ratio (Le Marchand et al,
1997). Moreover, the measure of alcohol intake is even more uncertain. Alcohol consumption is difficult to evaluate since the uncertainty lies not only in the size of a drink but also in the type of drink consumed. In the questionnaire we ask about “alcohol past week”, and a glass of wine and a drink are therefore equally reported by our study participants.

In conclusion, coffee had no effect on either ER status or age at diagnosis in patients with at least one \textit{CYP1A2*1F} C allele. In patients homozygous for the \textit{CYP1A2*1F} A allele, coffee consumption was associated with a later age at diagnosis and maybe also ER-negative tumors. The question as to whether coffee consumption is also associated with a lower breast cancer risk in women with the \textit{CYP1A2*1F} C allele and possibly an increased risk in women homozygous for the A allele warrants confirmation.

\section*{Study IV}

\textit{Frequency of the genetic polymorphisms}

The frequencies of the genotyped polymorphisms are in line with the frequencies reported by Yasar et al. in a Swedish control population (Yasar et al, 2002). These polymorphisms may therefore not be directly linked to an increased risk for breast cancer. However, we cannot rule out the possibility that these polymorphisms may be associated with an increased risk after certain exposures.

\textit{CYP2C8/CYP2C9 haplotypes}

Haplotypes between \textit{CYP2C8*3/*4} and \textit{CYP2C9*2/*3} were constructed (Table IV). \textit{CYP2C8/CYP2C9} haplotypes for \textit{CYP2C8*3} and \textit{CYP2C9*2/*3} have been reported in a Swedish population (Yasar et al, 2002). A very recent publication in a Spanish population has now included also \textit{CYP2C8*4} in the haplotype construction (Dorado et al, 2008). They also observed that \textit{CYP2C8*4} was unlikely to be present together with \textit{CYP2C8*3} or \textit{CYP2C9*2/*3}.

\begin{table}[h]
\centering
\begin{tabular}{lcccc}
\hline
 & \textbf{No allele} & \textbf{One allele} & \textbf{Two alleles} & \textbf{missing} \\
\hline
\textit{CYP2C8/9 *1/*1/*1/*1} & 42 (6.4) & 247 (37.8) & 362 (55.4) & 2 (0.3) \\
\textit{CYP2C8/9 *3/*1/*1/*1} & 641 (98.2) & 10 (1.5) & - & 2 (0.3) \\
\textit{CYP2C8/9 *3/*1/*2/*1} & 533 (81.6) & 112 (17.2) & 6 (0.9) & 2 (0.3) \\
\textit{CYP2C8/9 *3/*1/*3/*1} & - & - & - & 2 (0.3) \\
\textit{CYP2C8/9 *1/*4/*1/*1} & 567 (86.8) & 81 (12.4) & 3 (0.5) & 2 (0.3) \\
\textit{CYP2C8/9 *1/*4/*2/*1} & - & - & - & 2 (0.3) \\
\textit{CYP2C8/9 *1/*4/*3/*1} & - & - & - & 2 (0.3) \\
\textit{CYP2C8/9 *1/*1/*2/*1} & 624 (95.6) & 26 (4.0) & 1 (0.2) & 2 (0.3) \\
\textit{CYP2C8/9 *1/*1/*1/*3} & 571 (87.4) & 78 (11.9) & 2 (0.3) & 2 (0.3) \\
\hline
\end{tabular}
\caption{The frequencies of the most likely \textit{CYP2C8/9} haplotypes.}
\end{table}
**Genetic polymorphisms in relation to tumor characteristics**

GSTT1*0 was associated with a higher histological grade (P=0.026). The only SNP in the CYP2C8 and CYP2C9 genes that was associated with tumor characteristics in all patients was CYP2C8*4. Even though these patients have a higher histological grade (P=0.044), fewer PR-positive tumors (56% versus 71%; P=0.018) and similar ER status and tumor size, this polymorphism was associated with a lower frequency of axillary lymph node involvement as compared with the wild type (21% versus 38%, P=0.001). In the subgroup of patients with tumors larger than 20 mm, CYP2C8/9 *1/*4/*1/*1 was associated with a lower frequency of axillary lymph node involvement than in patients with the normal variants, OR 0.13 (95% CI 0.04-0.45; P=0.001). On the contrary, each allele of the CYP2C8/9 *3/*1/*2/*1 was associated with an increased risk of lymph node involvement OR 2.65 (95% CI 0.99-7.08; P=0.05), adjusted for tumor size, histological grade, age at diagnosis, ER and PR status. The difference in OR for lymph node spread between patients with CYP2C8/9 *1/*4/*1/*1 and CYP2C8/9 *3/*1/*2/*1 was substantial. However, whether these haplotypes can be useful as prognostic markers, they also need to be associated with risk of recurrence.

**Genetic polymorphisms in relation to early recurrences**

Although CYP2C8*4 was associated with less spread to the lymph nodes, this polymorphism was not associated with fewer early recurrences. Conversely, CYP2C8*3 was associated with shorter disease-free survival in all patients (Log-rank 8.62; 2df; P=0.013), adjusted HR 1.75 (95% CI 0.99-3.08; P=0.053). However, this association was much stronger in tamoxifen treated patients with CYP2C8*3, adjusted HR 2.93 (95% CI 1.25-6.85; P=0.013). Tamoxifen is a moderate CYP2C8 inhibitor and it is therefore plausible that polymorphisms that are associated with CYP2C8 activity influence response to tamoxifen treatment (Walsky et al, 2005). The molecular mechanisms behind the lower in vitro enzyme activity of CYP2C8*3 and CYP2C8*4 (Bahadur et al, 2002) are not known (Ingelman-Sundberg et al, 2007).

**Limitations**

Over 650 patients are included in the present study. Although this is a relatively large sample, a number of SNPs have been evaluated in smaller subgroups, e.g. different treatment groups. There is a risk of false positive findings due to multiple testing. Some of the findings are new and therefore warrant confirmation in independent materials. Moreover, when we evaluated SNPs in relation to early recurrences the follow-up time was still very short.

Tumor characteristics have been routinely evaluated by the departments of pathology in Lund and Helsingborg. The only factor that differed in the tumors of patients included in Lund and Helsingborg was histological grade. Since histological grade is not registered in the Regional Tumor Registry, we are unable to evaluate whether there is a difference between tumors from patients from Lund and Helsing-
The selected SNPs are involved not only in the metabolism of breast cancer drugs but also in the arachidonic acid metabolism, Fig 12. It would therefore be interesting to have had more extensive information on diet, including fatty acids, and on angiogenesis. Since blood samples are available, it is possible to analyze the lipid profiles of the patients.

**Figure 12** The three pathways of arachidonic acid metabolism. CYP2C8 and CYP2C9 are involved in the epoxygenase pathway. GSTs are involved in the lipoxygenase pathway in the formation of leukotrienes.

**General Discussion**

In this thesis we have identified a number of gene-environment interactions on tumor characteristics of prognostic importance and early recurrences in the BC blood case-only cohort.

**Genetic polymorphisms**

The genetic polymorphisms are used as markers for phenotype, since they are easier to evaluate than differences in hormones and growth factor levels between individuals. Genetic polymorphisms may reflect an altered enzyme function over a lifetime, whereas a blood sample may only reflect the levels at the time of sampling. In addition to genetic polymorphisms, it is important to note that there are other levels of variation that may affect risk, prognosis and treatment response. These variations include gene expression, copy number changes, miRNA binding sites, methylation sites, protein levels and possibly other as yet undefined variants. In addition, a potential source of misclassification when evaluating genetic polymorphisms is the spe-
cificity of the primers and the assay, since there are pseudo-genes that show great similarity to genes of interest, but possibly not at the site of the polymorphism. However, all our polymorphisms were present at frequencies similar to those reported by others.

Combination of IGF1-19/-19*multiparity and CYP1A2*1F*coffee on age at diagnosis

In Study I and Study III, interactions between genetic polymorphisms and lifestyle on age at breast cancer diagnosis were observed. When adding these variables to a multivariate linear regression model (Table V), the IGF1-19/-19*multiparity interaction remains significant and the CYP1A2*1F*coffee (≥2 cups/day) is of borderline significance. This suggests that the effects of these interactions on age at diagnosis are mutually independent. Multiparous patients with IGF1-19/-19 were diagnosed 10 years earlier than other patients, and patients with CYP1A2*1F/A/A who consumed ≥2 cups of coffee a day were almost seven years older at diagnosis.

Table V. Age at breast cancer diagnosis in patients with a moderate to high coffee consumption, the CYP1A2*1F/A/A genotype, the interaction between coffee consumption and CYP1A2*1F, multiparity, IGF1-19/-19 and the interaction between multiparity and IGF1-19/-19 (n= 403).

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>58.2</td>
<td>2.06</td>
<td></td>
</tr>
<tr>
<td>Coffee (≥2 cups/day)</td>
<td>-0.17</td>
<td>2.04</td>
<td>0.934</td>
</tr>
<tr>
<td>CYP1A2*1F/A/A</td>
<td>5.60</td>
<td>2.71</td>
<td>0.039</td>
</tr>
<tr>
<td>CYP1A2<em>1F/A/A</em>coffee (≥2 cups/day)</td>
<td>6.82</td>
<td>2.98</td>
<td>0.022</td>
</tr>
<tr>
<td>Multiparity</td>
<td>3.90</td>
<td>3.32</td>
<td>0.241</td>
</tr>
<tr>
<td>IGF1-19/-19</td>
<td>1.11</td>
<td>1.28</td>
<td>0.387</td>
</tr>
<tr>
<td>IGF1-19/-19*multiparity</td>
<td>-10.04</td>
<td>3.84</td>
<td>0.009</td>
</tr>
</tbody>
</table>

General interpretation of results

Reliability is the consistency of data. If multiple measurements generate similar results they are considered to be consistent, though this does not necessarily mean the results are accurate. In terms of SNP genotyping, rerunning samples using the same technique is a reliability test rather than a true validation. Validity is the correlation between the measurements obtained and the "true" value. This may be seen as a comparison between our genotype frequencies and the frequencies reported by others, though we cannot be sure that the other laboratories have the "correct" genotyping results.

False results – positive or negative – may be the result of several factors. When factors are studied that may be associated with age at breast cancer diagnosis, patient delay and doctor's delay are important. The fact that not all tumors are detected by mammography screening is a technical issue that also needs to be taken into consid-
eration. Once the tumor is operated on it is classified by the pathologist. The pathologist’s opinion may be subjective, and depends on the staining of the tumor, and whether the piece of the tumor is representative etc. In the laboratory, the technique used may also affect the genotyping results. Genotyping data and all information from questionnaires, patient charts, and pathology reports need to be entered in a database. Data entry may generate errors, and misuse of statistical tests may also result in wrong conclusions.

Association or causal effects?

This thesis presents the findings of a number of association studies. According to Hill’s criteria, several factors must be investigated in order to conclude whether or not an association reflects a causal relationship (Hill, 1965): size or power of the association, consistency of findings, specificity of findings, temporality, biological gradient, plausibility, coherence, experimental consideration and analogy.

In our association studies only a few of these criteria are met, and our results are therefore only deemed to be associations. However, I will comment on some of these factors below. In the pilot study the sample size was restricted and the power of the association was limited. In addition, Study IV was exploratory and a substantial number of statistical tests were performed. The results of Study II and Study IV should therefore be carefully interpreted. In Study III all analyses but the one between coffee and CYP1A2*1F on ER status, remained statistically significant when adding another 188 samples. This may indicate that the ones that remained significant are fairly robust, and the association between coffee and CYP1A2*1F on ER status should be more carefully interpreted. The association between coffee consumption and CYP1A2*1F any C allele on breast cancer risk is consistent with the findings in BRCA1 mutation carriers (Kotsopoulos et al., 2007). In ovarian cancer, the effect of coffee was stronger in women with CYP1A2*1F A/A (Goodman et al., 2003). Since all studies were performed in a case-only cohort, we can only evaluate temporality in relation to treatment response and recurrence. However, genetic polymorphisms are constitutional and thus precede the cancer. A polymorphism that alters the activity of an estrogen metabolizing enzyme may reflect altered endogenous estrogen levels throughout life. In terms of lifestyle factors, e.g. coffee, we assume that the consumption reported reflects the consumption at the time the tumor was initiated. The results need to be validated in a prospective case-control study. In Study III there was a biological gradient, whereby the association between ER-negative status and a young age at diagnosis increased with an increasing number of CYP1A2*1F C alleles. The finding that no patients with GSTM1*0 and GSTTI*0 had experienced a recurrence is plausible. It is reasonable to think that patients lacking these two enzymes are unable to deactivate their drugs, and will therefore have an improved treatment response and prognosis. Although our group mainly focuses on epidemiology, we have recently initiated experimental in vitro studies of caffeine and caffeic acid in different cell lines in order to elucidate the effect of coffee on ER status. Others have reported that a specific repeat length of a microsatellite denoted
GGN in the androgen receptor (AR) is associated with a specific transactivation in response to testosterone as compared with other repeat lengths (Lundin et al, 2007). By analogy, we think that IGF-1 levels are differently regulated by exogenous and endogenous estrogens, depending on the presence or absence of the common 19 CA repeat allele in the \textit{IGF1} gene.

It is important to note that there are associations for which not all of Hill’s criteria are met, but these associations are still considered to be causal, for example the strong association between smoking and lung cancer.

\textit{Confounding factors}

An association in one study population may not be true for another population, because of differences in genetic or environmental background. Ethnic background is a very important confounder, especially when constructing haplotypes, as in Study IV. In the BC blood study we have no information on ethnicity, which is a limitation. However, the vast majority of participants in the BC blood study are ethnic Swedes. Even though gender may be a confounding factor, all patients included in the BC blood study and in this thesis are women. Moreover, age is also a potential confounder. Age was the output variable in two of the studies in this thesis, and we have adjusted for age. Although the information on the patients in our questionnaires is extensive, we are always lacking some information. Even if we have information on concomitant medications, this information has not yet been analyzed in combination with SNPs. The more information there is, the larger the study populations that are required in order to perform more statistical tests.
Conclusions

In conclusion, this thesis contains the first gene-environment interaction studies performed in the BC blood study in southern Sweden. Specifically we conclude that:

**Study I:** *IGF1*-19/-19 was most frequent in patients diagnosed prior to age 45 years and in patients diagnosed between the ages of 55 and 59 years. *IGF1*-19/-19 and multiparity was associated with an earlier age of breast cancer onset. The combination of *IGF1*-19/-19 and multiparity may thus be useful when selecting a subgroup of patients for earlier breast cancer screening.

**Study II:** Patients with tamoxifen during radiation therapy, increasing alcohol and a high coffee consumption had an improved estrogen metabolite profile between the pre-and post-operative visit. *CYP1A2*1F also modified the 2-OHE/16αOHE1 ratio. Since a higher 2-OHE/16αOHE1 ratio has been associated with an improved prognosis, the presented factors may influence prognosis.

**Study III:** In patients with *CYP1A2*1FA/A, a moderate to high coffee consumption was associated with a higher proportion of ER negative tumors and a later age at diagnosis as compared with patients with a low consumption. Coffee consumption was not associated with age at diagnosis or ER status in patients with *CYP1A2*1FA/C or C/C. Coffee consumption was higher in patients with the *CYP1A2*1FA/A genotype as compared with patients with A/C or C/C. Since coffee consumption does not differ according to genotype in the general population, coffee may be considered a potential modifier in breast cancer.

**Study IV:** The frequency of *CYP2C8*3/*CYP2C8*4/*CYP2C9*2/*CYP2C9*3/*GSTM1*0/*GSTT1*0 were in accordance with frequencies reported for the selected polymorphisms in European populations. *CYP2C8*/*CYP2C9* haplotypes were constructed. *CYP2C8*4 was only present with the wild type SNPs in the remaining three positions. We confirmed the previously reported high linkage between *CYP2C8*3 and *CYP2C9*2. Patients with *CYP2C8/9*1/*1/1/*1 had a lower odds for lymph node involvement whereas patients with *CYP2C8/9*3/*1/2/*1 and tumors ≥21mm had a higher odds for axillary lymph node involvement. In addition, *CYP2C8/9*3/*1/2/*1 may identify patients that recur early when treated with tamoxifen. No patients with the GSTM1/GSTT1 homozygous double deletion had
a recurrence. The *GSTM1/GSTT1* homozygous double deletion may thus be associated with good prognosis.

Although the studies have been performed on a limited number of patients with a limited number of recurrences, most of the findings are new and will be of importance for breast cancer patients if confirmed in independent studies.
Future

The BC blood study is an ongoing study whereby the power of the studies performed increases over time as the number of patients included and the follow-up time increase. In addition, the numbers of patients that suffer relapse also increases over time, and this will increase the power of our pharmacogenetic studies. The BC blood study in Lund is a gold mine that is ripe for exploration.

This thesis involves genetic testing, which is interesting from an ethical viewpoint. Nowadays people can order genetic tests over the internet. A genetic test that gives people a risk estimate not only regarding breast cancer but also regarding other diseases can be both good and bad. However, it is unlikely that today’s commercially available tests will provide people with clinically relevant risk data, and they may cause distress.

When it comes to pharmacogenetics, the applications may be more straightforward. If a patient will not benefit from a therapy such as tamoxifen because she cannot activate the drug owing to polymorphisms in the activating enzymes, alternative treatments can hopefully be prescribed.

Clinical relevance

Finally, when is a finding ready to move into the clinical setting? BRCA1 mutation carriers currently have as much as a 60-80% lifetime risk of developing breast cancer (Easton et al, 1995; Ford et al, 1994; Ford et al, 1998; Narod et al, 2004). These women are offered prophylactic mastectomies in order to prevent breast cancer. Women living in Sweden have an approximately 10% risk of developing breast cancer prior to age 75 years. With one affected first-degree relative the woman’s risk is doubled, i.e. there is a 20% risk. According to the protocol from state of the Art in 2004, women in Sweden with a lifetime risk of over 20% and aged 25 to 30 years should be examined through palpation, should be taught self-examination and should undergo annual screening. Our findings need to be confirmed before they can be used in the clinical setting or before any recommendations on coffee consumption can be made.
Populärvetenskaplig sammanfattning (Swedish)

Bröstcancer är den vanligaste cancersjukdomen hos kvinnor i Sverige. Årligen insjuknar ca 7000 kvinnor och ca 1500 kvinnor dör i sin sjukdom.

Bröstets normala tillväxt regleras av ett flertal könshormon, bland annat östrogen och progesteron, samt tillväxtfaktorer som insulin-like growth factor-1 (IGF-1). Dessa har därför studerats i relation till bröstcancer. Livsstilsfaktorer som påverkar bröstcancerrisken omfattar bland annat reproduktiva faktorer som graviditet, p-piller och hormoner mot övergångsbesvär, samt alkohol och eventuellt även kaffe.

Bröstcancer kan även orsakas av genetiska faktorer. Mutationer i de kända ärftliga bröstcancergenerna BRCA1 och BRCA2 utgör dock endast en liten del av all bröstcancer. Nio av tio kvinnor som får bröstcancer har ingen nära släkting med bröstcancer och merparten av all bröstcancer förklaras kanske av ett flertal gener, eller genetiska variationer, med mindre genomslagskraft. Exempel på genetiska normalvariationer är substitutionen av en enstaka bas i arvsmassan, en ”single nucleotide polymorphism” (“SNP”), upprepandet av en eller flera baser, en mikrosatellit, eller förlust av hela genen eller delar av genen, en deletion. En SNP innebär att en av DNAts fyra baser (A, C, G, T) varierar normalt hos minst 1% av befolkningen på vissa positioner.

I denna avhandling har jag studerat kombinationen av genetiska normalvariationer och livsstilsfaktorer och hur dessa kan modifiera bröstcancerisk, prognos och risk för återfallet.

Eftersom vi inte vet vilka kvinnor som kommer att utveckla bröstcancer rekommenderas alla kvinnor i Sverige i åldrarna 40-74 år mammografiscreening, även om inte alla landsting följer detta. När tumören detekterats klassificeras tumören utifrån prognostiska faktorer som inkluderar; ålder vid diagnos, spridning till lymfkörtlar, histologisk grad, tumörstorlek samt uttrycket av hormonreceptorer med vilka hormoner och tillväxtfaktorer kan interagera. Baserat på dessa faktorer bestäms hur patienten skall behandlas.

Den primära behandlingen av bröstcancer är operation, varefter en tilläggsbehandling (adjuvant behandling) i form av strålning, cellgifter (cytostatika), hormonell eller antikroppsbehandling ofta ges. Men, en stor andel av de adjuvant behandlade patienterna (beroende på hur aggressiv tumören är) får inte en förbättrad
prognos. Antingen är patienterna redan botade av operation eller så svarar de inte på given behandling. Dessa patienter kommer inte att ha nytta av sin behandling men får kanske ändå de biverkningar som behandlingen medför. Idag finns således inte tillräckligt med information för att läkaren skall kunna ge rätt behandling till rätt patient.

Samtliga studier i denna avhandling har utförts i bröstcancer blodstudien, BC blodstudien, i Lund och Helsingborg. Patienter som väljer att delta i denna studie lämnar vid upprepade tillfällen blodprov samt fyller i ett formulär med frågor om kända livsstilsfaktorer som kan påverka bröstcancerrisk och prognos. Information om tumören och behandlingen erhålls från patientens patologutlåtande och journal. De genetiska analyserna har utförts vid den onkologiska forskningsavdelningen i Lund, medan hormonanalyserna i studie II utförts av Immuna Care Inc., i USA.

I studie I har vi undersökt hur avsaknaden av en genetisk normalvariation i den som kodar för tillväxtfaktorn IGF-1 påverkar åldern vid bröstcancerdiagnos efter multiparitet (dvs fött två eller flera barn). Patienter med avsaknad av denna normalvariant och som fött två eller flera barn hade en 5,9 års tidigare ålder vid bröstcancerdiagnos jämfört med övriga kvinnor.

I studie II undersökte vilka faktorer som var associerade med en gynnsam östrogenmetabolitprofil i blodprov tagna innan och efter operation från samma patient. Vi fann bl.a. att en genetisk normalvariation; CYP1A2*1F, var associerad med denna östrogenmetabolitprofil. Dessutom fann vi ett samband mellan en kombinationsbehandling (tamoxifen under stålbehandling), samt att alkohol och kaffe påverkade östrogenprofilen i en gynnsam riktning.


I studie IV har frekvensen av ett flertal genetiska normalvariationer, främst i läkemedels metaboliserande gener, studerats. Visa normalvarianter är kopplade till varandra, så kallade haplotyper. Haplotyper mellan polymorfier i CYP2C8 och CYP2C9 har satts samman. Dessutom har dessa genetiska variationer, samt deletionerna GSTM1*0 och GSTT1*0, studerats i relation till tumörkarakteristika och tidiga återfall. Patienter med CYP2C8*4 men normalvarianten i både CYP2C8*3, CYP2C9*2 och CYP2C9*3 hade i lägre utsträckning spridning till lymfkörtlarna i armhålan, vilket indikerar att dessa tumörer är mindre spridnings benägna. Dessutom fann vi att patienter med CYP2C8*3 och CYP2C9*2 men med normalvarianten i CYP2C8*4,
och CYP2C9*3, hade en ökad risk för spridning och en högre risk för tidiga återfall, speciellt vid tamoxifen behandling.

Sammanfattningsvis så har jag identifierat ett flertal samband mellan vanligt förekommande genetiska normalvariationer och livsstilsfaktorer som var associerade med ålder vid diagnos samt andra tumör egenskaper. Dessa studier är baserade på ett relativt litet antal bröstcancerpatienter och resultaten är inte ”sanna” förrän de bekräftats i oberoende studier. Men, våra resultat indikerar på att genetiska normalvariationer, eller avsaknaden av dem, kan i kombination med vanliga ”exponeringar” i form av barnafödande, samt kaffekonsumtion påverka bl a ålder vid diagnos. Du föds med dina gener, men du kan påverka din livsstil. Om våra resultat bekräftas kan de i framtiden användas för att identifiera patienter för tidigare screening.
Acknowledgement

I would like to start by thanking everybody involved in my education who inspired me and helped me make this thesis possible. I would especially like to thank:

My supervisor Helena Jernström. I think this has been a very interesting time for us both. It was not only your first time as a PhD supervisor but also my first time as a PhD student. I have learnt a lot, and I hope you have too. When we worked together we achieved greatness. You have always been there for me when deadlines have been tight.

My co-supervisor Mårten Fernö. You have been both a friend and an excellent coach, who believed in me and cheered me on through rough times.

My co-supervisor Åke Borg. It has been an honour to have your excellent input, especially in my thesis writing. Thank you for being "Mr Cool".

Dick Killander, Bo Bäldetorp, and Håkan Olson for running the Department of Oncology and for providing a stimulating research environment.

My co-authors: Carsten Rose, Christian Ingvar, Thomas Klug, and Per-Ebbe Jönsson for fruitful and productive collaborations and for enlighten me with clinical aspects.

Everyone involved in the BC blood study: all the patients and the research nurses in Lund and in Helsingborg; Karin Henriksson, Annette Möller, Anna Weddig, Linda Ågren, Ulrika Midelund, Arnhild Nilsson, and Karina Sandström. Everyone at the Department of Clinical Chemistry in Helsingborg. Sol-Britt Olsson for taking care of the BC blood bank and Maj-Britt Hedenblad for data entry. I would also like to express my gratitude to everyone who has ever handled a BC sample, as I know how demanding they can be.

Everybody at the genetics lab for great technical support: Therese Törnblom, Camilla Person, Lina Tellhed, Gunilla Sellberg, Ulla Johansson, Steina Aradottir, Eva Rambeck, Anna Isinger, and Chris Persson. Thanks for sharing not only your time and experience but also your patience (PSQ can be very demanding!) and friendship. You may not have taught me everything you know in the lab, but everything I know is thanks to you!

Mef Nilbert, who gave us PhD students some extra attention and discussed the important aspects of research in a broader context. Hopefully someone will take over from where you left off.

Pär-Ola Bendahl, who has put extra effort into my education – being enthusiastic not only about statistics but also about soccer finals.
All the PhD students who joined me in the ‘PhD student club’. To mention just a few: Karolina Holm, Gabriella Honeth, Sara Brommeson, Eleonor Olsson, JosefFernebro, Katarina Domanska, Helena Persson, and Katja Backenhorn. Being the only student in my group was quite lonely, and I really enjoyed our meetings and discussions. I hope that you will find the strength to start it up again.

The new recruits in our group: Maria Henningon, Maria Hietala, and Ann Rosen- dah. We have had many interesting discussions about anything and everything (including research, of course!). Maria Henningon, not only have you confronted Sir Bruce Ponder but you have also baked the most delicious chocolate cakes in the world.

Carina Strand, who stayed with me when all my other friends either got pregnant (why did you not let me in on the plan?) or went to San Diego with their husbands. Karin Renstam, you have been very supportive about almost everything.

My roommates Anna Linger and Susanne Magnuson, for putting up with a sometimes stressed-out and always talkative neighbour =).

Anita Schmidt Casslén, I have never met such a colourful person. You are a true inspiration on many levels and I am very glad to have got to know you.

Susanne André, is it possible to clone you? I think you always have the answer to everything. I owe you big time!

Lotta Welinder, for “hooking me up” with the right people, which resulted in a job.

Britt-Marie Lundh at the Regional Tumor Registry for always taking your time.

Eric Dryer for proofreading all my manuscripts.

My yoga instructor Daniel Strauser and all my yogi friends, especially Karolina Holm and Markus Clemmedson for giving me new perspectives. I also wish to take the opportunity of thanking “Mr Govinda” for providing me with excellent food fit for goddesses.

Jonas, you have at many times disturbed my “academic romantic working-late-nights dream”, and I so love you for that! How could I even come up with the idea of spending more time at my workplace than with the one who loves and appreciates me the most?

My dear parents, Gunilla and Jörgen. You have been supportive, loving and understanding. Now that I’m ending my PhD studies I hope to find the time to visit you more often.

My siblings Jessika and Henrik, for being the best of friends. A Thai massage and a couple of ginseng teabags are exactly what a stressed-out sister needs! Thank you.

This work was supported by grants from the Swedish Cancer Society, the Mrs Berta Kamprad Cancer Foundation, the Ingabritt and Arne Lundberg Foundation, Lund University Hospital Fund, the Crafoord Foundation, the G. Nilsson Foundation, the Swedish Research Council (K2001-27GX-14120-01A and K2008-68X-20802-01-3), the GA’s Donation for Breast Cancer Research, the 1049 Fund at the Lund Oncology Clinic, the Region Skåne ALF, the Medical Faculty of Lund University and an unrestricted grant from Novartis.

Finally, I would like to thank myself! Well done!
References


58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer 87: 1234-45.


