Common genetic variation and mammographic density: risk factors for breast cancer

By

Stine Margarethe Biong

Cancer Genome Variation Group
Department of Genetics
Institute for Cancer Research
Oslo University Hospital
Radiumhospitalet

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1361

ISBN 978-82-8264-354-2

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Table of contents

Acknowledgements

The work presented in this thesis has been carried out in the Department of Genetics, Institute for Cancer Research, at The Norwegian Radium Hospital from 2008 to 2011, and at McGill Centre for Bioinformatics, McGill University, Montreal, Canada, from March 2010 to March 2011.

I would like to express my gratitude to the University of Oslo and the Faculty of Medicine for admitting me into the PhD program. I would also like to acknowledge the Research Council of Norway (FUGE), South-Eastern Norway Regional Health Authority and Radiumhospitalets legater for the financial support that made this PhD possible.

I am very thankful to the Head of Department, Professor Anne-Lise Børresen-Dale for the privilege of being part of a laboratory with such inspiring and exiting research. I have thoroughly enjoyed working alongside all the talented scientists in the department.

I am very grateful for the opportunity my supervisor Professor Vessela Kristensen has given me in taking me on as a PhD student. It has been a time of scientific as well as personal growth from which I will greatly benefit in my future endeavors.

Kind regards to my co-supervisors Giske Ursin and Bettina Kulle for their involvement during my PhD. With their expertise in the field of mammographic density and statistics their involvement was essential.

To all my collaborators and co-authors: thank you for spending your valued time on the projects, your input and feedback has been important for me to improve my research.

In my daily work I have been fortunate to have a friendly and cheerful group of fellow students and colleagues who have created a pleasant work environment. I am especially thankful to Hege L. H. for being a wonderful colleague and friend with whom I have shared many scientific discussions as well as laughs during my PhD. As part of the Cancer Genome Variation Group I have had the opportunity to work with talented researchers and I am especially grateful for all the scientific input from Grethe and Hege E. In addition to your knowledge you have both provided me with a sense of stability in the occasional rough sea of a PhD student.

Åslaug and Vilde made the MDG samples presented in this thesis available to me, and also introduced me to the field of gene expression analysis, and for this I am very thankful. I am also grateful to Inger Torhild for providing me with the TMBC samples and scientific input.

A special thanks goes to the members of the pancreas group at the department of Genetics for being so friendly and including. I am especially grateful to Professor Elin Kure from whom I received valuable support and who inspired me and cheered me on during the final stages of writing this thesis, and Dr. Inger Marie Bowitz Lothe, who has shared with me her extensive knowledge and introduced me to the exciting field of pathology. Your kindness has been overwhelming.

I was fortunate to have the opportunity to collaborate with the McGill Centre for Bioinformatics and spend a year in Montreal, Canada during my PhD. I am very grateful to Professor Michael Hallett for welcoming me to his laboratory and for subsequently making me join the lab ice hockey team - giving me an interest I was sure I would never acquire! Thank you Vanessa for being a friend and colleague from home, away from home. I am also incredibly grateful to Matt for taking me under his wing and teaching me about statistics and programming, your talent and determination is truly impressive. I was in many ways introduced to Canadian life by Rob, Julie and Sean who included me in their everyday life and taught me the enjoyment of the national sport; ice hockey. It all went so fast, but it was a fantastic time.

I would also like to acknowledge all the women who donated their blood, tissue and time for without them this project would not be possible.

To my friends, especially my childhood friends Cecilie and Julie, thank you for always being there for me, through difficult times as well as good times. Kathrine, thank you so much for helping me with the modification and making of Figure 2.

My academic journey started many years ago with the encouragement from my parents who advised me to go and see the world while getting an education. I am incredibly grateful for this advice that has taken me to all corners of the world. It has truly been a fantastic journey so far! Thank you mum for always providing me with a helping hand whenever I need it and for all the times you have encouraged me to continue when I have felt it was impossible. You also helped me immensely and gave me peace of mind by proofreading this thesis. And to my brother and sister, Kristian and Erica, thank you for the optimism and the good times filled with laughter, I always look forward to spending time with you.

Oslo, November 2011

Margarethe B.

Aims

The objective of this study is to identify genetic predictors for increased mammographic density, and by way of this contribute to the knowledge of breast cancer development from normal, healthy breast through the potentially premalignant condition of dense breast to breast cancer.

1. Identify single nucleotide polymorphisms (SNPs) with putative impact on mammographic density and/or hormone levels.

2. In combination with expression data, identify markers with regulatory effect associated with mammographic density and hormone levels.

3. Apply original statistic analysis to compare mammographic density groups in order to identify genetic risk factors for increased mammographic density and breast cancer.

List of papers

Paper I

Margarethe Biong, Inger Torhild Gram, Ilene Brill, Fredrik Johansen, Grethe I.G. Alnaes, Toril Fagerheim, Yngve Bremnes, Stephen J. Chanock, Laurie Burdett, Meredith Yaeger, Giske Ursin, Vessela N.Kristensen. ***Genotypes and haplotypes in the insulin-like growth factors, their receptors and binding proteins in relation to plasma metabolic levels and mammographic density.*** BMC Medical Genomics 2010 Mar 19;3:9.

Paper II

Vilde D. Haakensen, Margarethe Biong, Ole Christian Lingjærde, Marit Muri Holmen, Jan Ole Frantzen, Ying Chen, Dina Navjord, Linda Romundstad, Torben Lüders, Ida K. Bukholm, Hiroko K Solvang, Vessela N. Kristensen, Giske Ursin, Anne-Lise Børresen-Dale, Åslaug Helland, ***Expression levels of uridine 5'diphosphoglucoronosyltransferase genes in breast tissue from healthy women are associated with mammographic density.*** Breast Cancer Research 2010 ;12(4):R65

Paper III

Margarethe Biong, Matthew Suderman, Vilde D. Haakensen, Bettina Kulle, Paul R. Berg, Inger Torhild Gram, Vanessa Dumeaux, Giske Ursin, Åslaug Helland, Michael Hallett, Anne-Lise Børresen-Dale, Vessela N. Kristensen. ***Candidate SNP analysis integrated with mRNA expression and hormone levels reveal influence on mammographic density and breast cancer risk.*** Submitted to Cancer Epidemiology Biomarkers and Prevention.

1 Introduction

"*We cannot hope to develop new, more effective strategies for cancer prevention if we do not understand how the factors that increase the breast cancer risk affect the development of the normal human mammary epithelium*"

- Anderson, Clarke and Howell, 1998 - [1]

The breast is the most prevalent site of cancer in women. Early detection is key to improving therapy and prognosis, and as a result more focus is now placed on the development and understanding of the normal breast. The aim is to detect signs of cancer already in the first steps in the transition from a healthy to a malignant state.

1.1 The breast

1.1.1 Breast anatomy

The breast is one of the few, if not the only, human organ that is not completely developed at birth [2], and unlike other ramiform organs most of the branching occurs during the adolescent years [3]. Central in the development of the breast are the female hormones, which play a key role from puberty through adulthood to menopause. The hormones exert their effect through local signaling cascades and stromal-epithelial interactions to elicit tissue reorganization, differentiation and specific activities that define each phase of development. The main function of the breast is milk production, and thus the breast has an "open" structure which ensures room for production and storage of milk during pregnancy and lactation (Figure 1). Structurally the mature female breast is made up of 1) adipose tissue, 2) connective tissue, 3) vasculature and 4) epithelial tissue [4]. The first three are included in the so-called microenvironment, while latter category includes the major functional mammary units, the lobules and ducts. The lobules, which are capable of producing milk, are drained by the ducts which transport milk to the nipple and the two are collectively referred to as the terminal duct lobular unit (TDLU) [5,6] (Figure 2). The area known as the TDLU was so called due to a disagreement between surgical and comparative pathologists. While the first claimed that breast cancers arose in the ducts, the latter saw that it originated in dilated lobulo alveolar units. Due to rigidity towards changing the basic concepts in surgical pathology, the area was politically correctly termed TDLU [5]. The TDLUs have gained

much attention due to the fact that they have been recognized as the site of origin of preneoplastic lesions such as atypical ductal hyperplasia, which may evolve to ductal carcinoma in situ (DCIS), which in turn may progress to invasive carcinoma [7].

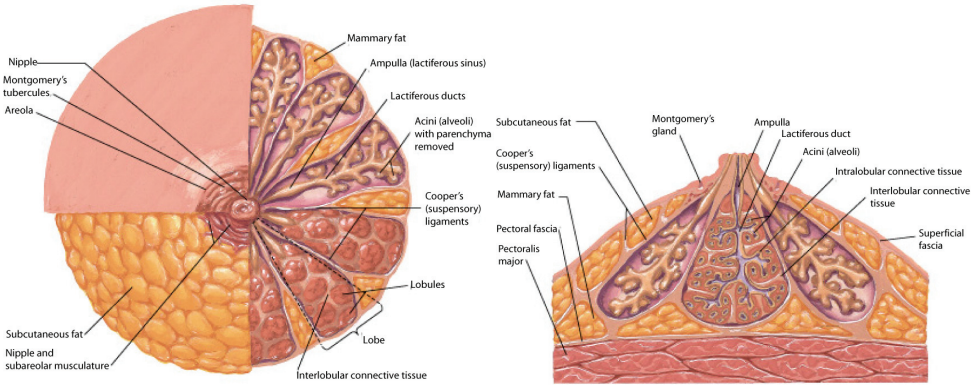


Figure 1: Structural anatomy of the breast [8]. ***Netter Illustration from www.netterimages.com ©Elsevier inc. All rights reserved***

Although some refer to the TDLU as the equivalent of lobule type 1 [7], others [6] and also herein, the TDLU is defined as the duct and lobule combined, thus the structure of the TDLU varies depending on the different lobule stages 1 to 4 (Figure 2:B). According to studies by Russo and Russo, lobule type 1 is the least differentiated lobule with approximately 6-10 ductules, lobule type 2 evolves from the previous and is comprised of approximately 47 ductules which may progress to lobule type 3 with approximately 80 ductules [7,9]. Lobule type 4 is the final and most differentiated stage and is only reached in the event of pregnancy when the number of ductules increases further in addition to increasing in size [10]. Women who have never been pregnant (nulliparous) tend to have breasts mainly composed of lobule type 1 and 2, while parous women who have reached the most differentiated state have breasts mainly composed of lobule type 2 and 3 [10]. In the mammary gland, the lobules are separated by loose connective tissue (intralobular stroma) and are arranged in lobes each with their separate ductal system. The lobes, which can be seen as slightly triangular structures (Figure 1), are separated from each other by adipose and dense connective tissue (interlobular stroma) [6].

1.1.2 Breast histology

The TDLU is the major histological unit of the breast [1] and is made up of epithelial and myoepithelial cells and a basement membrane (BM) (Figure 2, C). The BM is the outermost layer that forms the physical barrier separating the myoepithelium from the intralobular stroma. Adjacent to the myoepithelial cells we find the epithelial cells which line the ductal and lobular lumen. Little focus has been given to the myoepithelial cells compared with the other cells in the breast, especially the luminal epithelial cells [11,12]. However, the myoepithelial cells are fundamental in maintaining homeostasis in the breast. Apart from producing the BM, the myoepithelial cells influence luminal epithelial cell differentiation, polarity and proliferation in addition to invasion and migration of adjacent luminal epithelial cells [13]. In the lactating breast the myoepithelial cells have contractile properties which enables the flow of milk through the treelike structure of the ducts [14] and contributes to milk secretion [15]. The luminal epithelial cells which can be subdivided in ductal and lobuloalveolar epithelium, line the ducts and the alveoli respectively. The epithelial cells located in the alveoli become milk-secreting cells during pregnancy [14].

At the tip of the TDLU the epithelial cells are less differentiated, and in addition to being responsible for most of the mammary growth and branching [16] these cells have been suggested to be progenitor cells or epithelial stem cells, [17,18] (for review see [19]). Although much is still unknown regarding stem cells, these epithelial stem cells are thought to give rise to both the mature luminal epithelium and the mature myoepithelium (Figure 3) [20]. Most of the knowledge on stem cells is derived from studies using cell culture or rodents, however the human mammary stem cell *in situ* are likely to be similar to those present in mice or rats [21]. Experiments performed on mice and rats have revealed “cap cells” as progenitors of the myoepithelium [16,22] and “body cells“ (Figure 2, C) as progenitors of the luminal epithelium [16] (for review see [23]).

The breast is a dynamic organ, and in parallel with the rapid proliferation of stem cells forming epithelial and myoeptithelial cells, apoptosis takes place to form the ductal structures seen throughout the mammary gland (Figure 2). It has been suggested that cells in contact with the BM are protected from cell death signal whereas the cells in the centre of what is to become the

lumen undergo apoptosis [23]. Programmed apoptosis is also central during remodeling of the gland after pregnancy and at menopause.

The luminal epithelial cells are most frequently reported as the origin of tumors, however, the myoepithelial cells are also capable of giving rise to tumors, reviewed in [12,24]. The potential to develop breast cancer may depend on high amounts of normal stem cells with mutation potential, or because replicating cells have started a malignant pathway [25].

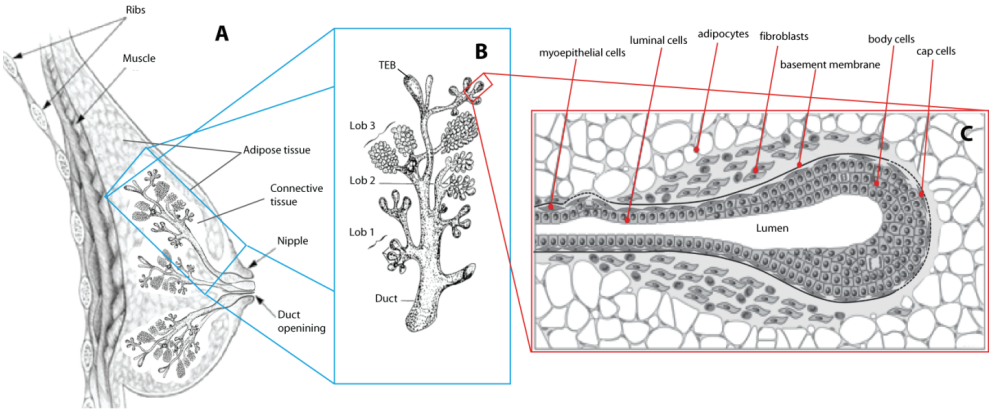


Figure 2: A) Structure of the mammary gland, B) the lobular structures comprising the mammary gland. Lob:lobule, TEB: terminal end bud, C) diagram of a TDLU. Modified from [10,26].

The microenvironment

"Tumor initiation and progression as well as response to therapy depend on the interplay between the cancer and its host - the microenvironment"

- Gonzales-Angulo, Hennessy and Mills, 2010 - [27]

Regulation of the proliferation, differentiation and survival of the epithelial cells of the mammary gland is provided by the stromal cells in which they are embedded [28]. Epithelium and stroma interactions have been suggested, due to the existence of indistinct boundaries between the epithelium and a very loose stroma at the tip of the TDLU [16]. The microenvironment

surrounding the TDLUs is made up of a distinct intralobular stroma [19] which is composed of a cellular component and an extracellular component/extracellular matrix (ECM). In the cellular component we find the loose connective tissue made up of fibroblasts, adipocytes, vasculature (endothelial cells) and immune and defense cells such as lymphocytes and macrophages. The ECM, which is a matrix of organic material or "ground substance", is a semi fluid gel which is produced and maintained by the fibroblasts. Collagenous, reticular and elastic fibers are embedded in the ground substance which provide structural support [29]. By loosely binding tissue fluid, the ground substance provides a medium for passage and exchange of materials and metabolites throughout the connective tissue and with the circulatory system respectively [14].

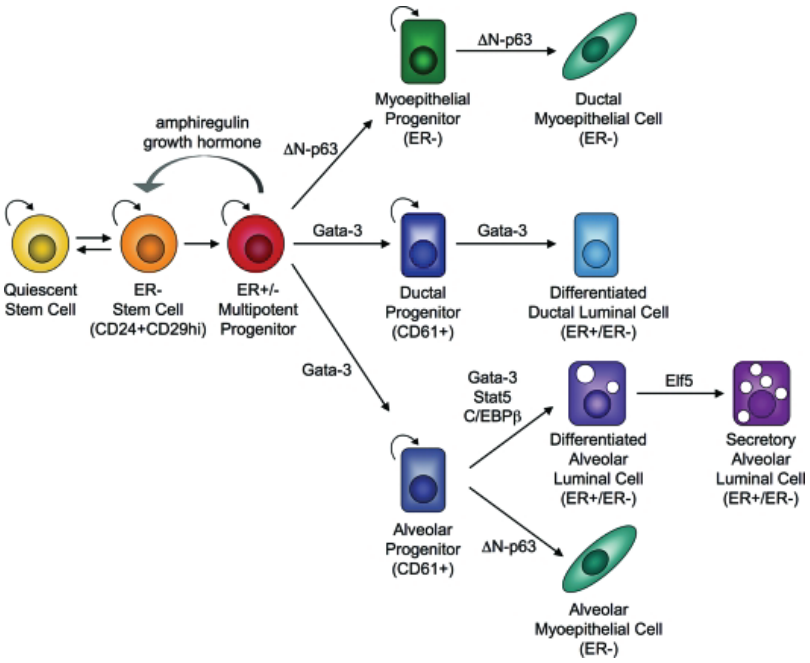


Figure 3: Suggested progression and regulation of a stem cell dependant on hormones, growth factors and transcription factors. Green: basal restricted stem cells (myoepithelial). Dark blue: luminal restricted progenitor cells. Violet: Alveolar restricted lineage in response of pregnancy [30].

1.1.3 Normal breast development and physiology

"The hormonal control of development, growth and differentiation in vivo of the human breast is, unfortunately, still an enigma "

- Jean McManus & Clifford Welsch, 1984 - [31]

The breast is a hormonally responsive organ, and apart from the first trimester in the womb, the development of the breast is guided by steroid hormones from fetus to puberty, through adulthood and until menopause.

From fetus to puberty

The first formation of breast tissue starts already in the womb and at the 5th week of gestation in both males and females, and up to the 32nd week of pregnancy the process is independent of steroid hormones [9]. In brief, two parallel milkridges starting at the groin and ending in the armpit (axilla) form the primary mammary structures. Although most of these ridges fade away, small portions are left in the chest region to form the primitive nipples, with a mammary fat pad containing 15-20 strips of epithelium that will later become the secretory units of the mammary gland [32]. During the adolescent years leading up to puberty the development of the mammary gland is limited but keeps pace with the general growth of the body [33].

Puberty and adulthood

The female mammary gland undergoes substantial changes as a result of the onset of puberty and the consequent increase in female sex hormones. Cyclic changes in mammary morphology are observed due to changes in epithelial proliferation induced by the hormone production in the ovaries [1]. Both the epithelial tissue and stroma of the breast are influenced by the fluctuating levels of the sex steroid hormones estrogen and progesterone. Estrogen is the main driver of ductal differentiation in the breast while progesterone is the main driver of lobular development, reviewed in [34]. Both hormones are produced in the ovary and while estrogen is formed in the Graafian follicles (mature vesicular follicle), the progesterone is produced in the corpus luteum that is the remainder of the follicle after the release of the oocyte. Progesterone may induce changes in the breast epithelium on its own, however, the greatest changes in breast morphology are seen when the two hormones work together to stimulate full ductal-lobular-alveolar

development [35,36]. An increased complexity of the ductal-lobulo-alveolar structures is seen during each menses, nevertheless, the drop in both estrogen and progesterone levels at the start of each menses is accompanied by a regression in the development of the alveolar clusters. However, the mammary development reached in one cycle never completely returns to the starting point of the previous and thus each consecutive cycle, until approximately the age of 35, ensures a gradual accumulation of epithelial tissue [33,37].

The hormones are the drivers of development and differentiation of the mammary gland in addition to local controls [16]. In the developed mammary epithelium the TDLU is the major hormone sensitive area [38,39]. Studies on terminal end buds (TEB), which are mouse analogs to human TDLU [5], showed that proliferation and regulation of the TEB cells, are regulated by systemic hormones such as estrogen, progesterone and growth hormone (GH), and also locally acting growth factors such as epidermal growth factors (EGF) transforming growth factor alpha and beta (TGFα/TGFβ) and insulin like growth factor 1 (IGF1), reviewed in [23]. Accordingly, the latter growth factors are suggested regulators of programmed cell death in the TDLU due to their ability to inhibit or initiate cell growth [23].

Additional hormones are involved in growth regulation and differentiation of the TDLUs of the breast during pregnancy and include placental lactogen, prolactin, glucocorticoids, growth hormone, insulin, chorionic gonadotropin, oxytocin and the sex hormones. During lactation, the hormone oxytocin induces milk ejection [26] through activation of the contractile properties of the myoepithelial cells [19]. Once the milk synthesis has begun the lactogenic hormones and milk proteins may directly regulate growth factors and epithelial proliferation in the breast.

Involution

There are two main types of involution; one is experienced during cessation of lactation and the other at menopause. Involution is programmed cell death, also termed apoptosis, which occurs as a result of the withdrawal of the steroid hormones from mammary epithelial cells. The result is a reduction in glandular tissue and an increase in stroma or adipose tissue in premenopausal or postmenopausal women respectively.

Post-lactational involution is two-phased. The first phase starts when the weaning process reaches its final stages. At first there is an accumulation of milk in the lumen and little apoptosis of

epithelial cells. However after 1-2 days the process is irreversible with the major event being epithelial apoptosis. During the second stage the BM is degraded, alveoli collapse and macrophages infiltrate and the gland is restructured to a virgin-like state. At the end of the second stage 50-80% of the alveoli have been cleared [40] and the breast consists of 70-90% lobule type 3 [9]. As a result of the post-lactational involution the breast tissue in the parous premenopasal women is remodeled and prepared for a new pregnancy [40]. The ducts are not involved in the aforementioned process, in contrast to the postmenopausal involution in which the number of both lobules and ducts is reduced [41].

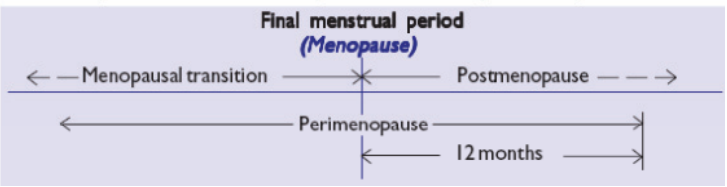


Figure 4: The status of the ovary surrounding the menopause. ***Illustration: WHO***

Involution by ageing occurs during the transition from premenopause to postmenopause which takes place over several years and includes the perimenopause and menopause (Figure 4). Perimenopause commences at the end of a woman's reproductive period and includes the period before menopause and the year following menopause, at which point a woman is classified as postmenopausal. Menopause which is usually experienced by the average woman at the age of 50, marks the withdrawal of 99% of the 400 000-700 000 ovarian follicles that are present in a 5 month old female fetus [33,42]. At the time of menopause the ovaries have already had a reduced function for one or two years and according to Russo and Russo this event results in a regression of lobule structures type 2 and 3, and an increase in the number of lobule type 1 and adipose tissue in the mammary gland. The regression is more marked in parous women but is also observed in nulliparous women, and both groups will eventually have breasts mainly composed of lobule type 1 [7]. However, although the breasts of parous and nulliparous postmenopausal women may be at the same lobular stage, the hormonal and physiological changes in the mammary epithelium of parous women may have altered their risk for developing breast cancer

[7,43]. Hence, pregnancy is postulated to give a protective effect against breast cancer development. Throughout the postmenopausal years the endogenous hormones, including estrogens and androgens, are derived mainly from the adipose tissues (Figure 5).

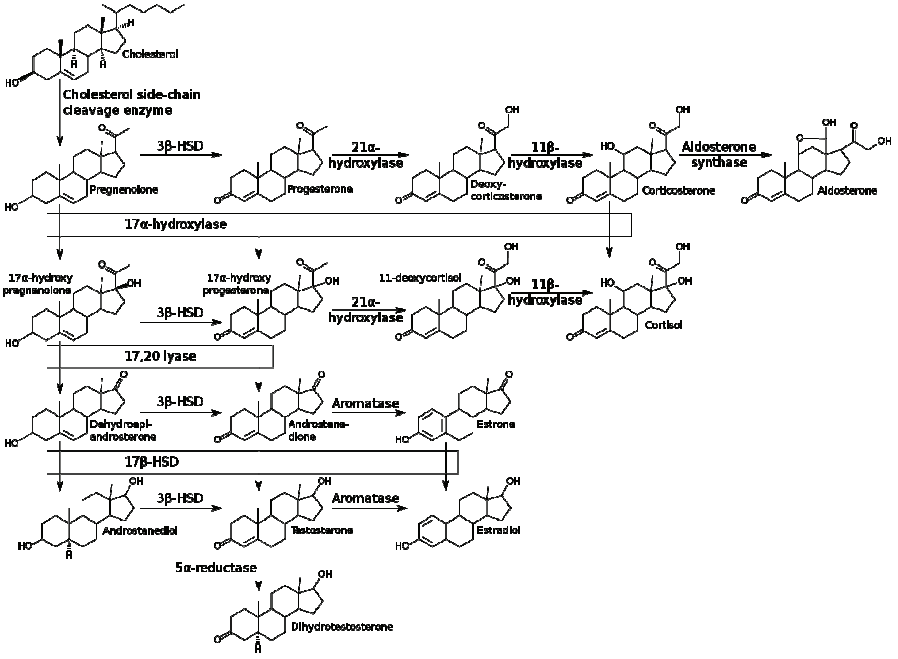


Figure 5: The conversion of steroid hormones from cholesterol (steroidogenesis). ***Illustration by Michael Häggström***

1.1.4 Steroid hormones

The word hormone is derived from the Greek word horman and means to "set in motion". Hormones, which are made by nearly every organ or tissue type in the body, serve to send signals from one cell to another to instigate processes such as cell growth. In women the steroid hormones, also known as sex hormones, are mainly produced in the adrenal cortex and the ovaries.

Biosynthesis of steroid hormones

As depicted in Figure 5 the steroid hormones are synthesized from cholesterol. The cholesterol is synthesized from either acetate, from pools of cholesteryl esters in stereoidogenic tissue, or more commonly from low density lipoprotein taken up from diet. In premenopausal women the LDL is taken up into the ovarian follicles, a process which is facilitated by follicle stimulating hormone (FSH) and lutenizing hormone (LH). In the follicle the LDL is degraded enabling the release of the cholesterol. The synthesis of steroid hormones from cholesterol is dependent on one or several of six different cytochrome P450s, which are encoded by different *cytochrome P450* (*CYP*) genes. The *CYP* genes are expressed in a cell and tissue specific manner, enabling cell and tissue specific synthesis of steroid hormones. The ovarian cortex produce androgens, and as previously mentioned, estrogen is synthesized in the Graafian follicles, and progesterone in the corpus luteum. The synthesis of hormones is carefully regulated due to the fact that they cannot be stored in the body [44]. In circulation the steroid hormones are bound to the sex hormone binding globulin (SHBG), which serves to inhibit their functions through decreased bioavailability. In turn, the level of SHBG is decreased by high levels of androgen, insulin and insulin like growth factor 1, while high levels of estrogen and thyroxine increase SHBG levels. Other important factors for the levels of circulating hormones include the uridine 5'-diphospho-glucuronosyltransferase (UGT) enzymes and the 3 beta-hydroxysteroid dehydrogenases (HSD3B) and the sulfotransferase family (SULT). The UGTs catalyze the glucoronidation of endogenous estrogens and androgens thus making them water soluble and more easily excreted from circulation. Likewise the members of the SULT family are enzymes located in the liver and adrenal gland and are seen to convert drugs and endogenous compounds into water soluble sulfate conjugates, thereby enabling their clearance from circulation. Enzymatic activity of the HSD3B is essential in the biosynthesis of all steroid hormones in which it catalyses several conversion processes [45].

The hormones are divided into the estrogens; estrone (E1), estradiol (E2) and estriol (E3), the progestagens; progesterone and the androgens; androstenedione, dehydroepiandrosterone, dihydrotestosterone and testosterone (Figure 5). Estrone is synthesized from androstenedione, estradiol from estrone in premenopausal women and from testosterone in postmenopausal women and estriol is synthesized from estradiol. Estrone, which is a weaker estrogen than estradiol, is present in higher levels in postmenopausal women. Progesterone is derived from pregnenolone

and its activity is modulated by the progesterone receptors. Dependent on the receptor they bind the hormones may be grouped into; glucocorticoids, mineralcorticoids and the steroid hormones.

Steroid hormone receptors

Estrogen and progesterone exert their effects through the estrogen receptor (ER) and the progesterone receptor (PR). Both ER and PR belong to the nuclear receptor family (NR) and like other members of this family they are ligand-activated transcription factors that regulate gene expression [46]. Upon binding of the hormones to their respective receptors, two ligand-receptor complexes may dimerise. The result is increased phosphorylation of the receptors, subsequent binding of specific hormone responsive elements (HRE) in the promoters of the target genes and interaction with coactivatior proteins and transcription factors needed for the transcription of DNA to mRNA.

The estrogen receptor consists of two proteins, namely ERα and ERβ which are transcribed from the *ESR1* and *ESR2* genes respectively. In the presence of ligands the receptors may dimerise to form ER-αα, ER-αβ or ER-ββ receptor variants. Although the receptors have high structure homology they are expressed in various ratios depending on the tissue. ERα is expressed in bone, liver, normal breast and breast cancer cells, CNS, cardiovascular system and the urogenital tract, in addition to the endometrium and ovarian stroma cells. ERβ is expressed in bone, kidney, breast, the brain, lungs, intestinal mucosa, prostate and endothelial cells. Due to differences in their transactivational domains the receptors may activate different genes. While ERα is targeted in cancer therapy, ERβ may interfere with the anti-proliferative effects of antiestrogens on tumors and may enhance tumor aggressiveness (for an extensive review on the ER see [47]).

Both estrogen and progesterone act through the ER to induce expression of the PR. Like the estrogen receptors the progesterone receptors may also dimerise in the presence of ligands and form the homodimers AA and BB or the heterodimer AB. In contrast to the ERs, the PRs are transcribed from a single gene (*PGR*) with two alternative promoters. Ratios of the different isoforms vary in response to developmental and hormonal status and also carcinogenesis, for an extensive review see [48]. While removal of PR-A show no effect on the mammary gland, removal of PR-B reduce pregnancy-associated mammary gland morphogenesis [48].

Steroid hormone and growth factor crosstalk

Extensive crosstalk has been reported between IGF and estrogens [49], which suggests that estrogens act through ER and induce expression of IGF1 and subsequent phosphorylation of IGF1R. In turn IGF1 initiates cascades which involve activation of the ER. In addition the serine /threonine protein kinase *Akt* and phosphatidylinositol 3 kinase (PI3K) are seen to play a key role in ER-IGF1 crosstalk with *Akt* mediating the effects of IGF1 [49]. In breast cancer cells E2 and IGF1 cooperate to stimulate cell cycle progression [50].

The IGF family of growth factors is, as the sex hormones, highly involved in the development of the breast, and has been suggested to be involved in the development of both MD and BC. The members of the IGF family include the insulin-like growth factor (IGF) ligands such as IGF1 and IGF2, insulin-like growth factor binding proteins (IGFBPs) that prolong the IGFs half-life in circulation such as the IGFBP1-6, and acid labile subunit (ALS) [51]. The insulin-like growth factor receptors (IGFRs) mediate the signals of the ligands that bind them. In mammary ductal morphogenesis the presence of insulin-like growth factor is necessary for the estrogen and growth hormone (GH) induced ductal growth. While estradiol and IGF1 may independently provide some ductal branching normal TEB formation by estradiol and GH is dependent on IGF1 [52,53]. Other known effects of the IGF pathway include cell proliferation and inhibition of apoptosis [54] thus linking it to cancer development. High circulating levels of IGF1 are shown to be associated with increased risk of breast cancer amongst young women (<50) [55] while IGFR1 may play a substantial role in the regulation of breast cancer cell growth [56,57]. Conversely, low levels of IGFBP3 are found associated with BC risk in premenopausal women [58].

1.2 Single nucleotide polymorphisms

On average the human genome of two individuals is 99.9% correlated [59]. However, variations in the DNA sequence can be detected approximately every 300 bp. These variations are usually on the single base level in the form of deletions, insertions or substitutions and are collectively referred to as single nucleotide polymorphisms (SNPs) (Figure 6, a). The human genome is estimated to encompass more than 11 million SNPs which renders them the most common form of genetic variation [60]. By definition, the difference between a SNP and a mutation is their population frequency, SNPs have a frequency of > 0.01 while mutations have a frequency of < 0.01. Although most SNPs have no described biological consequence, the ones that do, make up the diversity seen among humans [61]. SNPs have gained increasing interest due to, amongst other things, their ability to influence protein folding and function through change in amino acid composition. SNPs may also influence mRNA expression through changed affinity of transcription factors to their binding site.

Haplotypes

A haplotype is a particular combination of alleles in a genomic sequence which is inherited together as one unit (Figure 6, b). These combinations may occur with varying frequency depending on recombination events and mutations. The human genome can be divided into blocks with limited variability in haplotypes [62]. Such haplotype blocks are useful in genetic studies since they allow for genotyping of a single SNP or a small group of SNPs, namely haplotype tagging SNPs (htSNPs) (Figure 6, c), for the identification of all the alleles within the given haplotype block. In theory, this means that the identification of a specific allele marks the presence of all the other alleles, thus reducing the list of SNPs required for genotyping and also the cost. The degree of allelic association within a haplotype may be estimated by calculating linkage disequilibrium (LD). LD does not designate linkage or lack of equilibrium, but rather refers to a “disequilibrium in a genetic linkage analysis” caused by the nonrandom association of SNPs at two or more loci. Combinations of SNPs in LD are referred to as a haplotype. If several haplotypes are in LD they are referred to as haploblocks On a genomic level LD reflects the rate of mutations, recombinations and natural selection during evolution and on a population-based level LD is indicative of the pattern of geographic subdivision and breeding system reflecting population history [63]. The most common measures of LD are r^2 and D which both range from 0 to 1. LD equal to 1 denotes SNPs in complete LD.

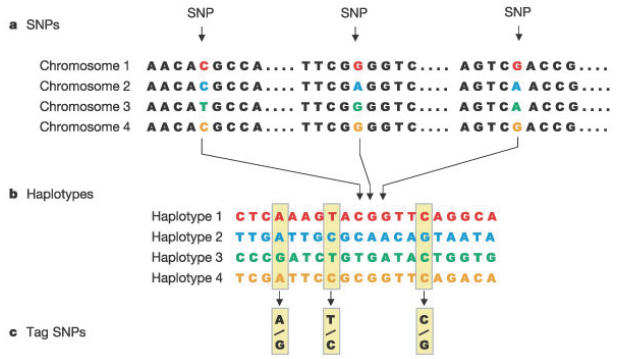


Figure 6: Schematic diagram of a) SNPs identified in samples from four individuals (1-4), b) The co-inheritance of SNPs in the four individuals resulting in haplotypes, c) The identification of htSNPs for the unique identification of haplotypes within the four individuals [64].

The International HapMap project

The International HapMap project [65] has made a large contribution to the knowledge of common genetic variation in humans. The project started in 2002 and was a collaboration between scientists in six countries: Japan, The United Kingdom, Canada, China, Nigeria and the United States. The goal of the project was to genotype in three years, 1 million common SNPs in 270 individuals for the identification of common genetic variations, such as SNPs and haplotypes, across chromosomes in ethnically different individuals. The individuals, who were selected to be a representation of the world population, were divided into the four categories; Nigerian from Yoruba (YIR), Japan from Tokyo (JPT), Han Chinese(CHB) and Utah residents with ancestry from northern and western Europe (CEPH). At the time of writing this thesis the HapMap project has genotyped 3.1 million human SNPs, representing 25-35% of the common SNP variations in the selected population [66].

1.2.1 SNPs in the estrogen pathway

Due to the importance of the estrogen pathway in the development and diseases of the mammary gland it has been granted much interest. As a result, research concerning this pathway has been performed in studies of both the normal breast and diseases of the breast. An array of genetic variations such as SNPs representing genes of the estrogen signaling pathway have been found associated with conditions of the breast including increased mammographic density and breast cancer, and include amongst others, the estrogen receptor [67], *UGT* genes [68], *SULT* genes [69] and the *HSD3B* gene family [70,71] (Figure 5).

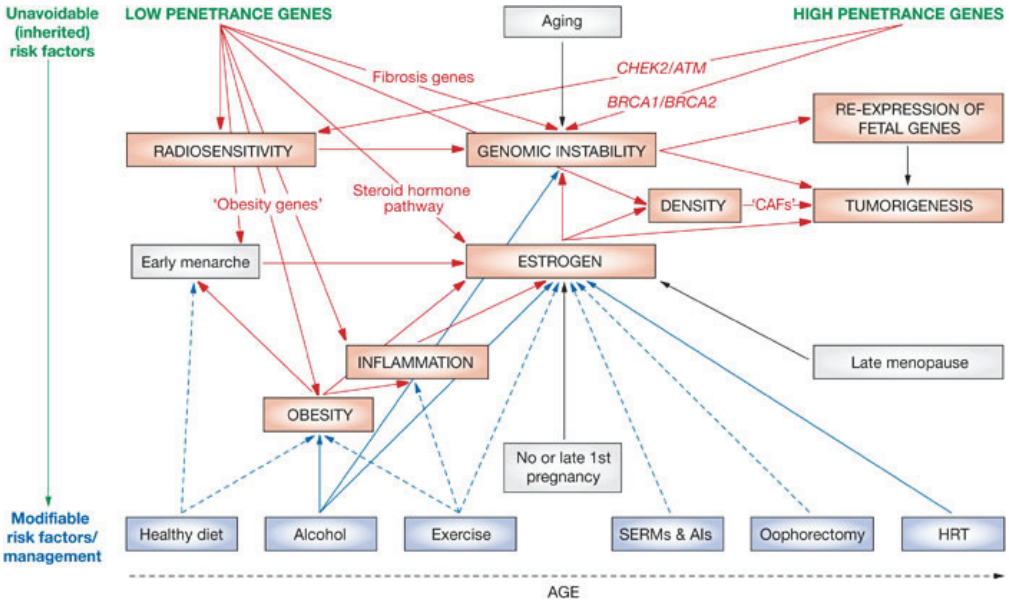


Figure 7: Diagram depicting the factors influencing BC risk. Red: genes and pathways, Blue: modifiable risk factors, Gray: reproductive life style events, Green: inherited and un-modifiable risk factors. From [72].

1.2 Breast cancer

"Tumors are wounds that never heal"

- Harold Dvorak 1986 - [73]

In general, cancer can be described as the uncontrolled growth of cells in the body, leading to lumps of cells or tumors that can disrupt tissues and organs in the body resulting in a lethal outcome.

It has been known since 1896 that breast cancer (BC) is a hormonally dependent disease [74], and for a quarter of a century that cancer is a disease with dynamic changes in the genome [75] (Figure 7). During the last decades scientists have been working on understanding the molecular process that underlies cancer initiation and growth, and even though much is still unknown there has been great progress. Hanahan and Weinberg have defined the acquired capabilities of cancer into six parts: 1) self-sufficiency in growth signals, 2) insensitivity to anti-growth signals, 3) evading apoptosis, 4) limitless replication potential, 5) sustained angiogenesis, and 6) tissue invasion and metastasis. These capabilities may occur in parallel or successively and cause variation in time necessary for complete tumorigenesis [75].

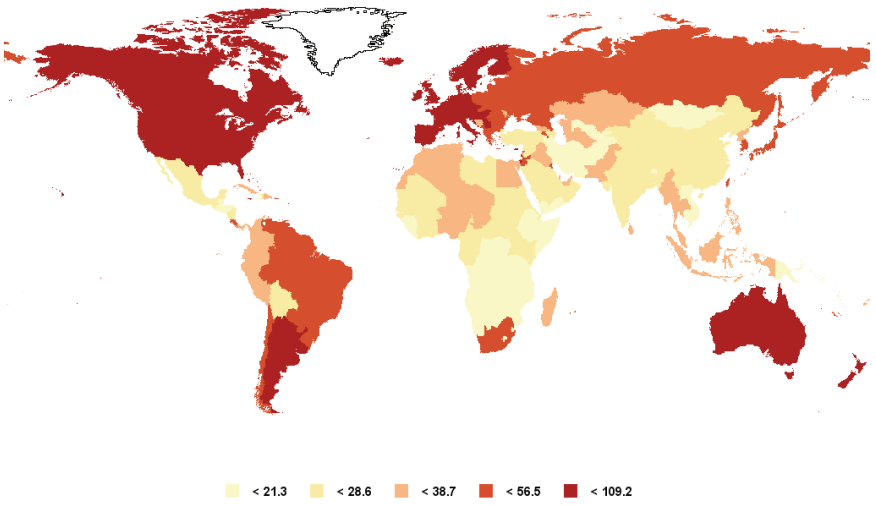


Figure 8: Age-standardized breast cancer incidence rate per 100,000 among women worldwide in 2010 [76].

1.2.1 Incidence and mortality

Breast cancer is the most common cancer in women worldwide with 1.38 million new cases each year, accounting for 23% of all new cancer cases and 14% (458,400) of all cancer deaths in 2008 [77] (Figure 8). Also in Norway it is the most predominant cancer among women with 2753 new cases registered in 2008, followed by colorectal and lung cancer. Breast cancer has the highest cumulative risk[1] based on numbers from 2004-2008 and it is estimated that one in 12 Norwegian women will develop breast cancer by the age of 75. Of all cancers developed by women between the ages of 25 and 69, one out of every third will be diagnosed with breast cancer. In Norway, breast cancer is ranked as the third cause of cancer death in women, only surpassed by lung and colon cancer [78]. Increasing incidence indicates that the disease continues to be a serious problem for women's health (Figure 9). The increase in western countries the last two decades is most likely due to a change in reproductive pattern and lifestyle in addition to postmenopausal hormone therapy use and increased detection due to screening [76]. The increased incidence seen in developing countries however is explained with a shift in reproductive patterns and a change towards a more western lifestyle [79].

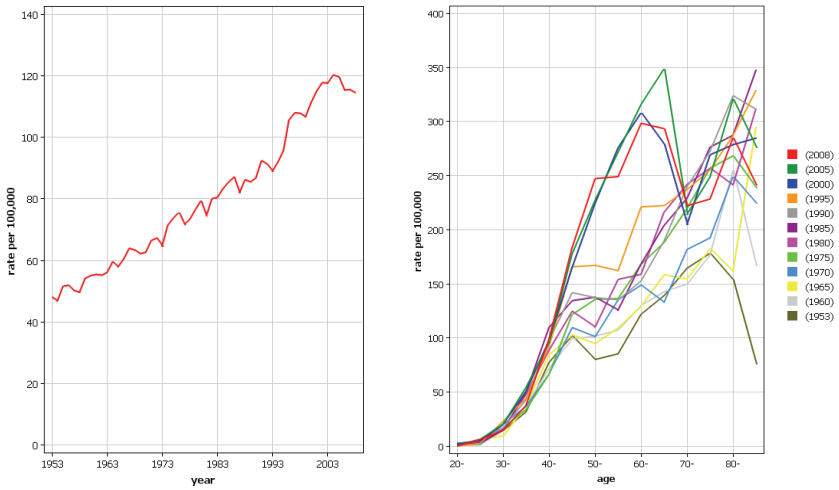


Figure 9: A: Breast cancer incidence rates in Norwegian women (1953 - 2008), B: Age adjusted breast cancer incidence rate, Norwegian women (1953 - 2008) [80].

[1] Cumulative risk of breast cancer refers to the effect of age on breast cancer risk in which each additional year increases the risk.

1.2.2 Mammographic screening program

The Norwegian mammographic screening program started in 1995/96 as a pilot project with the long-term goal of reducing breast cancer mortality among Norwegian women between the ages of 50-69, and also to test the infrastructure needed for such a program. The pilot was successful and the mammographic screening program was implemented throughout all the Norwegian counties by the end of 2004. The goal of the program is to reduce breast cancer mortality [78].

1.2.3 Breast cancer risk

The epidemiology of breast cancer has been known since the 1970s [81]. The main risk factors for breast cancer include gender, age, reproductive factors, genetics, alcohol consumption [82], and mammographic density [82-87] (Figure 7 and Table 1).

Age

For most cancers a linear relationship is observed between incidence and age. The most likely explanation being that accumulation of non-reversible DNA damage over time causes genomic instability leading to cancer [88]. However for breast cancer, the relationship between age and incidence is not linear [89,90]. The model proposed by Pike and colleagues [91] explains this non-linear relationship as breast tissue aging due to hormonal exposure over time. The model includes the parameters age at menarche, first full term pregnancy (FFTP) and menopause (Figure 10).

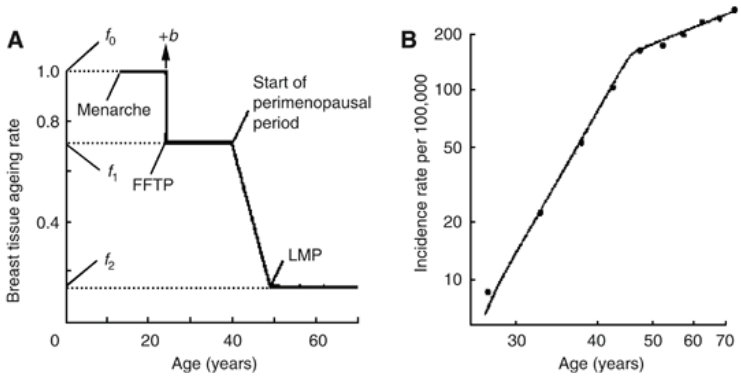


Figure 10: A) Pikes model on breast tissue aging. B) Age-specific breast cancer incidence rate. LMP: last menstrual period. FFTP: first full term pregnancy [91].

Hormonal influences

Reproductive life style events leading to increased breast cancer risk include nulliparity [10,92-94], late full-term pregnancy, early onset of menarche [81,92,95] and late menopause [92]. However, a protective effect is observed with increased numbers of children [96]. Additionally, increased duration of lactation is protective for premenopausal BC risk [97]. The common denominator for these factors is that they all influence the ovarian hormones previously shown to be important for the development of the breast. The hormone dependency of breast cancer (BC) was first discovered by Beatson in 1896 who implemented removal of the ovaries to treat inoperable BC [74]. It is believed that the cumulative exposure of the breast tissue to hormones, especially the ovarian hormones estradiol and progesterone, plays a key role in determining their effect on breast cancer risk [91,98].

Pregnancy may either promote or protect against BC dependent on the age at FFTP [99]. FFTP late in a woman's life increases the risk of breast cancer [100-102], while earlier age at FFTP decreases the risk [93,96,100,102]. Increased parity decreases BC risk substantially, with each subsequent birth resulting in 10% reduction [96,103]. It has been suggested that the protective effect is predominantly mediated through the stroma [104]. It is reasonable to believe that the mammary epithelium has reached a development stage with an altered potential to develop BC [93], reviewed in [20], and that the protective effect is derived from fully differentiated mammary epithelial cells [105,106]. In addition a reduced number of mammary stem cells, an altered response to estrogen and reduced levels of circulating hormones may also play a role, for extensive review see [106].

Early onset of menarche may lead to an earlier establishment of regular ovulatory menstrual cycles and a longer lifetime exposure to the endogenous hormones. In addition, early menarche leads to elevated levels of hormones throughout a woman's reproductive years [107,108]. Similarly, a late menopause prolongs the duration of which these hormones exert their effect. The protective effect of parity was first described in 1926 [103] and later confirmed by a large number of studies reviewed in [109].

Both oral contraception and hormone therapy (HT) use are linked to BC risk [110,111]. In premenopausal women, current use of oral contraceptives confers at most a slightly increased risk of BC which endures until approximately 10 years after termination of use [112,113].

Postmenopausal hormone therapy (HT) use increases the risk of BC substantially [111,114], and while estrogen alone causes increased risk, combining estrogen with progesterone has shown to confer an even greater risk [115-117]. HT associated BC risk increases with duration of use and does not regress to baseline until 5 years after cessation [110]. Interestingly, a recent study reported favorable tumor characteristics and better survival among women with BC who had used HT before diagnosis [118].

Anthropometric factors

Increased risk of breast cancer is associated with increased height [92,119], increased weight and BMI [92]. Increased BMI in older women increases breast cancer risk [119]. Most studies show that higher body fat/weight in adulthood is associated with an increased risk of post menopausal BC, but lower pre-menopausal BC risk. However, it is acknowledged that the relationship between BC risk and weight is complex and not completely understood [120].

Diet

Because diet is modifiable, understanding the role of diet in cancer etiology is important. A change in diet could alter levels of endogenous hormones and growth factors, and also reduce BC risk through antioxidants. However, to date there is no consistent evidence of dietary involvement in cancer [121-123], with the exception of alcohol. Alcohol consumption has been reported to be associated with BC incidence, conferring increased risk with increased consumption [123,124]. Total fat intake has been given much attention in cancer research but most studies have not been able to confirm the hypothesis that increased fat intake presents a greater risk of BC [125]. Reactive oxygen species (ROS) also known as free radicals, have gained much attention due to their ability to cause oxidative stress resulting in cellular damage through reactions with protein and nucleic acids [126-128]. Due to accumulation of DNA damage, ROS may induce cancer [128]. Increased levels of ROS may arise from inflammation, infection, extreme exercise and environmental factors such as pollution, tobacco smoke and radiation, for review see [129]. Antioxidants such as vitamin A, C, E, beta-carotene, lycopene, and selenium, have been shown to protect against reactive oxygen species and reduce oxidative stress and DNA damage [127,129]. Accordingly, it has been suggested that taking supplements of zinc, beta carotene, vitamin C and E protects against breast cancer [130]. On the genomic level, genetic alterations in genes related to oxidative stress have been found associated with risk of BC [131].

Table 1: Breast cancer risk factors. From [132]

Risk factor	Direction of effect[a]
Well-confirmed risk factors	
Family history in first-degree relative or genetic predisposition (e.g. *BRCA1*)	↑↑
Height	↑
Benign breast disease	↑↑
Mammographically dense breasts	↑↑
Parity	↓
Age at first birth > 30 years versus at < 20 years	↑↑
Lactation (longer durations)	↓
Menopause at > 54 years versus at < 45 years	↑↑
High endogenous estrogen levels	↑↑
Postmenopausal hormone use	↑
Ionizing radiation exposure in childhood	↑↑
Menarche at < 12 years versus at > 14 years	↑
High body mass index (postmenopausal)	↑
High body mass index (premenopausal)[b]	↓
Alcohol use (~ 1 or more drinks/day)	↑
Probable relationship exists, based on substantial data	
Current oral contraceptive use	↑
Physical activity	↓
Limited study to date	
High prolactin levels	↑↑
High premenopausal insulin-like growth factor I levels	↑↑
In utero exposures	↑
Nonsteroidal anti-inflammatory drug use	↓

[a] Arrows indicate approximate magnitude of the relationship: ↑, slight to moderate increase in risk; ↑↑, moderate to large increase in risk; ↓, slight to moderate decrease in risk; ↓↓, moderate to large decrease in risk.
[b] In Western countries – data are less consistent in other lower risk populations.

Genetics: family and personal history of BC

Having a personal or family history of BC elevates the risk substantially [133], the latter being a strong indication that BC is a genetic disease. High penetrant genetic mutations (high individual risk) in genes such as *BRCA1* [134], *BRCA2* [135], *PTEN* [136] and the tumor suppressor protein *TP53* [137] have been identified to account for increased risk in predisposed individuals. Although, with the vast amount of research being performed, new tumor suppressor genes[2] and oncogenes[3] are still being identified. The known genetic mutations account for approximately 20-25% of familial BC cases [138,139], suggesting that sporadic and lower penetrant genetic variations such as SNPs, together with environmental factors [140] are at play in determining the susceptibility to BC. The identification of SNPs conferring increased risk for a disease such as BC promise a better understanding, prevention, early detection and treatment of the disease. Numerous studies have identified SNPs associated with increased BC risk [141-149]. These include SNPs in the oxidative stress related genes *CYBA*, *MT2A* and *TXN* [131], genes related to the hormone biosynthesis pathway; *CYP1A1, CYP3A4, CYP1B1, CYP17, CYP19, SULT1A1, AHR, HSD17B, GSTM1, GSTT1,GSTP1,GSTM, COMT* [140,149] and *CYP19A1* [150] and in the insulin-like growth factor gene family: *IGF1* [151,152], *IGFBP3[153]* [154], *IGF1R* [155]

Other SNPs with moderate to low-penetrance effect that are identified associated with BC risk include: *TOX3*, *MAP3K1*, *LSP1*, *CASP8*, *FGFR2*, 8q and 2q35 [156], *XRCC1* [146], *ICAM1*, *ICAM4* and *ICAM5* [143], *LRP1*(lipoprotein receptor related 1) [144], CHEK2 1100delC; a checkpoint kinase resulting in a frameshift alteration [141], transcription growth factor beta (*TGFβ*) [147], *DPF3* [148], locus 9q31.2 (*RAD2)*)and *ACTL7A*, locus 6q25.1 (*ESR1*)[157]. Due to the power needed to detect low penetrant and low frequency SNPs a study of small sample size is not enough, thus a new trend is emerging in which several studies are assembled into one large study. The most frequently used approaches for SNP association identification include genome wide association studies (GWAS) and candidate gene studies. Recent GWA studies based on such large pooled studies have identified additional SNPs in genes which may contribute to the risk of BC. These genes include; *RAD51L1* [158], locus 5p12 [159], *FGFR2*, *TNR9*, *MAP3K1*, *LSP1* [139], *CASP8* [160], thus confirming some of the findings previously mentioned.

[2] A gene which protects a cell from acquiring the mutations required for tumor initiation
[3] A gene that might become cancer causing upon mutation.

Growth factors in the microenvironment

"Changes in epithelial plasticity permit a dynamic cell migratory response dependent on the constitution of the cell, its gene expression and on input from the local environment. "

Micalizzi et al, 2010 [161]

The microenvironment does not only maintain tissue architecture, inhibit cell growth and revert malignant phenotype, but may also promote and induce cancer. As mentioned in section 1.1.2 the microenvironment provides a passage of materials and metabolites, which includes the endogenous hormones, and growth factors. The result is dynamic signaling produced and received by the cells in the microenvironment [15]. The development of the breast is dependent on, and regulated by, these signals during growth and differentiation (Figure 11) [15]. Albeit, the stimuli might also induce malignant growth and the circulating levels of endogenous steroid hormones [114] and growth factors such as the IGFs [162] have been implicated in BC.

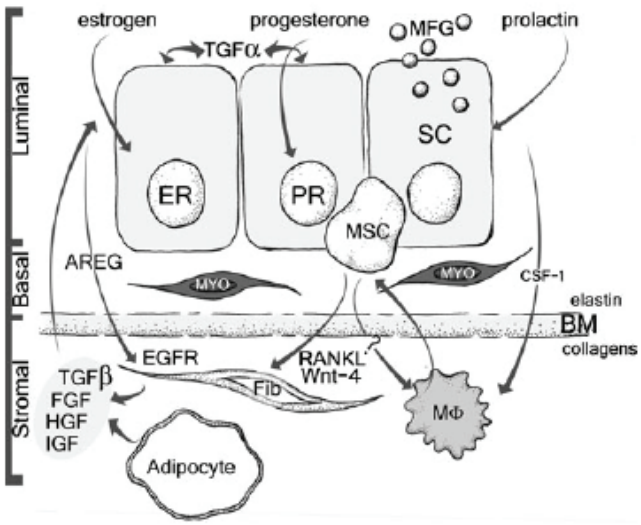


Figure 11: Schematic illustration depicting some of the signaling that occurs in the normal microenvironment of the mammary gland development [162,163]. AREG: amphiregulin, BM: basement membrane, CSF-1: colony stimulating factor-1, EGFR: epidermal growth factor receptor, ER: estrogen receptor, FGF: Fibroblast growth factor, Fib: mammary fibroblast, HGF: hepatocyte growth factor, IGF: Insulin-like growth factor, MΦ: macrophage, MFG: milk fat globule, MSC: mammary stem cell, MYO: myoepithelial cell, PR: progesterone receptor, RANKL: receptor activator of NF-κB ligand, TGF: transforming growth factor.

Epithelial-mesenchymal transition

The epithelial cells of the breast are derived from progenitor cells (see section 1.1.2). During development, the progenitor cells are able to gain an epithelial cell phenotype and vice versa, a process referred to as mesenchymal-epithelial transition (MET) or epithelial-mesenchymal transition (EMT) respectively [161]. The EMT provides the flexibility required during development, in particular during embryogenesis, but also during remodeling due to wound healing and regeneration of fully differentiated tissues [164,165]. In normal breast development, EMT processes can be seen during ductal branching through regulation by the epidermal growth factor (EGF), hepatocyte growth factor/Scatter factor (HGF/SF) and proteases such as the matrix metalloproteinases (MMPs) [166]. EMT is also observed in breast cancer, where it contributes to cell plasticity and metastasis. The mesenchymal phenotype acquired by tumor cells permits local invasion and escape from the primary tumor [161]. In addition to regulating features of the tumor cells, the EMT also regulates the tumor cell-microenvironment interaction, including the immune cells. Interestingly, interleukin signaling profiles are observed in invasive cancers, moderately in cancers in situ and only weakly or not at all in normal tissue [167].

Based on the dynamic nature of the normal breast, it is clear that there are numerous opportunities for the development of cancer through EMT related processes. Although breast cancer is known to mostly arise in the epithelial cells, it is becoming evident that also the tumor stroma and the microenvironment are involved in cancer progression. However, the mechanisms underlying their influence are still unknown. Finak et al investigated the gene expression signatures in stroma of cancer patients and found a stroma-derived prognostic predictor (SDPP) enabling stratification of disease outcome [168] and thus emphasizing the importance of stromal biology in breast disease.

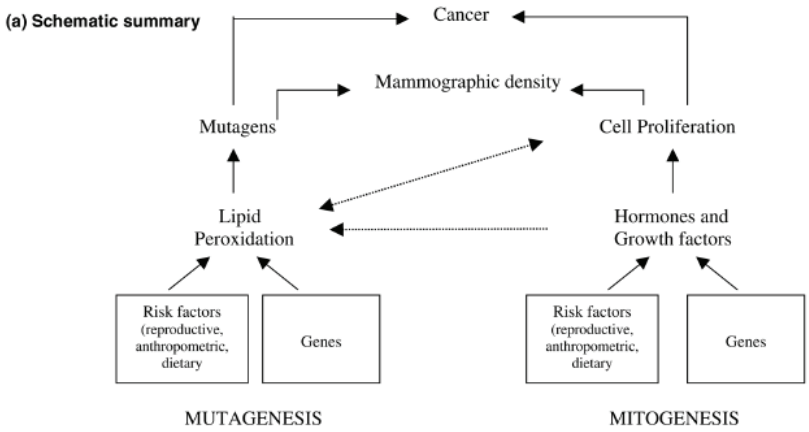
Mammographic density

"Mammographic density is perhaps the most undervalued and under-utilized risk factor in studies investigating the causes of breast cancer"

- Celia Byrne, 1997- [169]

Mammographic density (MD) was first proposed as a risk factor for breast cancer in 1976 by Wolfe [83,170], and is today accepted as an established risk factor for breast cancer [84,87,171-175] for both premenopausal and postmenopausal women [176]. Studies on women with invasive breast cancer have shown that MD may also predict local recurrence [177]. The effect of MD is estimated to persist for 10 years subsequent of MD assessment [171]. Women with density comprising more than 75% of the total area of the breast have a 4-6 fold increased risk of BC compared to women with less than 5% breast density [85,87,171]. It is not completely understood how density of the breast parenchyma influences the risk of developing breast cancer. Although the question of a possible masking effect of MD on BC lesions was raised early [178], this effect has been found to be small [175,179] and cannot explain the association with breast cancer risk.

Studies on the biology behind this link suggest there might be histological [180-182], environmental [183] or genetic factors (see section 1.3.2), or a combination of these (Figure 12). Most likely the latter is true. It has become increasingly evident that BC is a heterogeneous disease with several subtypes on the molecular [184], and pathological level, reviewed in [185]. Whether or not similar subtypes exist in MD, or if MD can affect risk of certain subtypes remains uncertain. However, as with BC, associations of MD with ER and HER2 status have been reported [176,186-189]. Association of MD with pathologically determined luminal A and triple negative breast cancers has been reported [86].



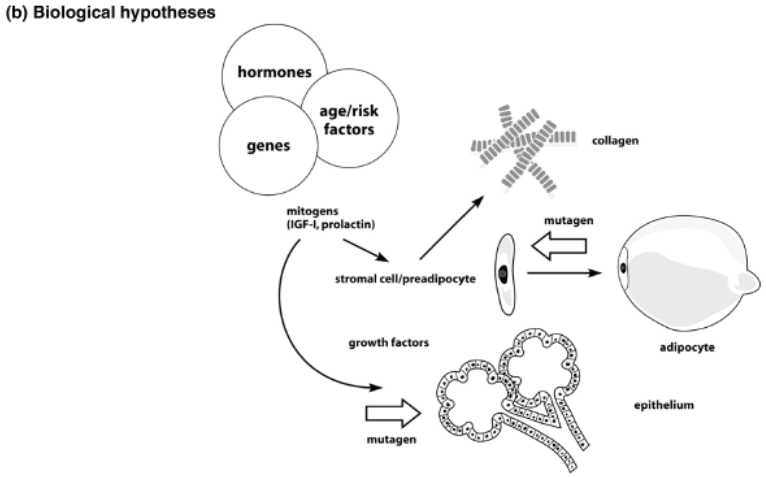


Figure 12: Hypotheses of BC risk associated with MD. (a) schematic summary of the underlying processes in MD that may lead to BC, mitogenesis is the disruptive effect on cell proliferation while mutagenesis is genetic damage by mutagens increasing genome instability (b) Diagram showing the biological hypothesis involving the tissue components responsible for MD (epithelial cells, stromal cells, collagen and adipocytes). From [190].

1.3 Mammographic density

Mammographic density (MD) is determined at mammography by evaluation of the radiological appearance of the tissues comprising the breast. While adipose tissue appears dark, the tissue referred to as "mammographically dense" is mostly comprised of stroma and epithelial tissue and appears white on exposure to x-rays (Figure 13). Breasts with increased density have been found to have greater numbers of epithelial and stromal cells compared with the less dense counterparts [180,191]. There are several methods for evaluating and classifying MD and depending on the method the variation in estimated BC risk is evaluated to be marginal [192-194].

1.3.1 Mammographic density classification

Mammographic density can be estimated from either analogously or digitally obtained mammograms. Both methods are based on x-ray technology, and patients undergo the same procedure for both. However, analogue mammography uses film to produce a static fixed image while digital measurement uses detectors that change the x-rays into electrical signals. The electrical signals are converted to numbers in a digital receptor which also processes these numbers to generate an image that can be displayed. The latter method, although more expensive, has been shown to be more flexible and advantageous, and at the same time better at detecting cancers in mammographically dense tissue [195]. The mammograms are evaluated in accordance with one of several methods for determining BC risk by mammographic density, mainly these entail qualitative and quantitative methods.

Qualitative methods

Initial studies investigated the parenchymal patterns and texture in the mammography image. Of such methods Wolfe's classification [83] was the one most commonly used [196]. Wolfe's parenchymal pattern method classifies risk of BC from mammography images based on four groups; N1, P1, P2 and DY ranging from tissue mainly composed of fat to tissue with an increased density respectively [170,197]. Tabár's method [198,198] is another quantitative method. Although it is not widely used, Tabár focuses on four "building blocks" in describing breast composition, nodular density, linear density, homogenous fibrous tissue and radiolucent adipose tissue. Depending on the individual involvement (%) of these four building blocks, a mammogram will be classified as having pattern: I, II, III, IV or V. The BIRADS (Breast Imaging Reporting and Data System) [199] method started out as a qualitative method by

classifying MD according to four categories (ACR1-4), but has later changed the classification to a quantitative-based method using percent density quartiles [194].

Quantitative methods

A number of quantitative methods exist. The simplest involves a subjective classification approach in which a radiologist evaluates mammographic density based on predefined categories. One such method was developed by Boyd [84,85,200] who classifies the mammograms according to six class categories (SCC) according to percent density; Class1: 0%, Class 2: 0-10 %, Class 3: 10-25%, Class 4:25-50%, Class 5: 50-75% and Class 6: 75-100% (Figure 13). As previously mentioned BI-RADS also has a quantitative approach. BI-RADS quantitative system uses percent density in 4 categories: < 25%, 25-50%, 51-75% and <75% density [199]. Additionally, computer-assisted methods exist which are the most widely used method for assessing MD in epidemiological studies [196]. Such computer-assisted methods include Cumulus [201,202] and Madena [203] and both entail the use of specially developed software and digital images to determine MD according to a given threshold [196]. The computer-assisted method with the use of the Madena method is described in detail in section 2.4.3. Briefly the method requires an experienced reader and computer software for the handling of the mammography image which is digital, or if analogue, digitized. For the purpose of studying mammographic density in relation to epidemiological and molecular factors, it is common to use MD as a continuous variable, either as absolute density (ABDEN) or as percent density (PDEN). ABDEN is the number of pixels within the area defined as dense while PDEN is the proportion dense tissue in relation to non-dense tissue (adipose) in the breast.

Volumetric methods

The methods described so far are based on two dimensional images. Volumetric methods offers a three dimensional view and may therefore improve BC risk estimates. Methods under current development include magnetic resonance imaging (MRI) [204] and ultrasound tomography [205] to mention some.

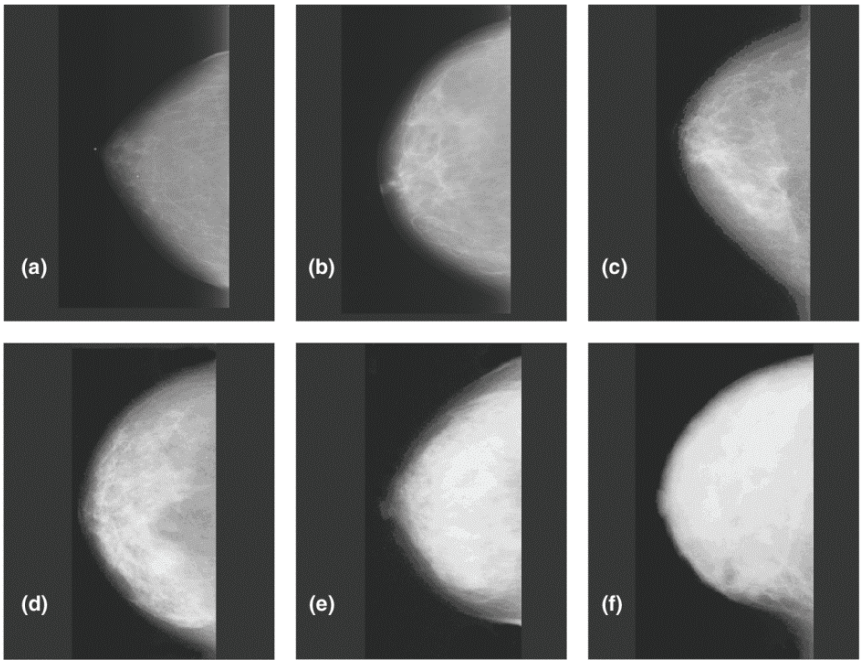


Figure 13: Images depicting mammographic density according to Boyd's six-category system based on fibroglandular tissue content. a) 0%, b) <10%, c) 10-25%, d) 26-50%, e) 51-75%, f) >75% [206].

1.3.2 Causes of MD variation

For the most part, the factors that influence breast cancer risk also cause variation in mammographic density. With regard to the reproductive factors, increased MD is seen with increased birth weight [207], premenopausal status [207,208], nulliparity, HT use [208-210], late age at first birth and no children [208]. The relationships are observed across ethnic groups [211]. Additionally, cyclic changes in the proliferation rate of breast epithelial cells are seen in premenopausal women due to the production and release of estrogen and progesterone from the ovaries [1]. Breast epithelial and stromal cells attenuate X-rays, and increased proliferation is positively associated with MD [180,191]. Hence, MD may be seen to change according to the menstrual cycle.

Anthropometric factors also influence MD; increased MD is seen with increased height [212-215] and while BMI is positively associated with breast cancer risk (in postmenopausal women) it is inversely associated with MD [208,216]. This suggests that BMI is a negative confounder of

the association between MD and breast cancer risk. High BMI implies a higher amount of adipose tissue which will appear translucent on the radiological examination and thus increase the area of non-dense tissue. This is mostly an issue when measuring PDEN as opposed to ABDEN, when a ratio between total breast area and dense area is reported using percentage compared to crude measurement of the dense breast area in pixels respectively. Additionally, in postmenopausal women most of the steroid hormones are derived from the adipose tissues, hence increased levels of hormones are found in postmenopausal women with high BMI. While estrogen levels are seen to increase BC risk slightly [217], it is inconsistently found associated with MD and while some find a positive association [87,218,219] others do not [220].

Diet

Diet may influence the level of MD and it is debated whether increased intake of fat may increase MD levels [221]. There is some [222,223] but not consistent evidence [224,225] that diets high in fat are associated with higher MD. While some have reported increased levels of MD with increased alcohol consumption [226,227] others have not [228]. High intake of vitamin D has been shown in some [229,230] studies to reduce MD while others see no association [231-233].

Genetic influence

In addition to the epidemiological factors, there is strong evidence of genetic influence in the development of MD based on results from studies on family history [234], familial aggregation [235] and twins [236,237]. Most of the genes responsible for MD are unknown, however genetic variation in the form of SNPs have been reported associated with MD levels.

Polymorphisms in genes coding for the vitamin D pathway have been studied for association with MD and while some have been positive (*VDR*) [238] others have found no association (*VDR*, *CYP27B1*, *DBP*) [239].

As previously mentioned association of MD with the level of sex hormones has been reported and studies on polymorphisms residing in genes belonging to the estradiol pathway are numerous; SNPs in *COMT* were associated by some [240-242] but not all [243]. Associations of MD with estrogen-metabolism related SNPs residing in *SULT1A1* [244], *UGT1A* [243,244], *HSD17B1* [241], *CYP1A2* [242], *CYP1B1* [240,241], *CYP19A1* and *CYP1A1* [240] with MD have been reported. However, others have found no association between MD and the SNPs in the

genes; *CYP1A1*, *CYP1B1* and *CYP17* [242]. Other members of the estradiol pathway in which SNPs have been associated with MD include; *ESR1* [240,245] and *HSD3B1* [70,71]. Conversely, one study found no effect of 239 SNPs in 34 estrogen metabolic genes on MD [246], and no association has been seen between SNPs in the androgen receptor (*AR*) gene and MD [243].

Due to its mitogenic and antiapoptotic effects the IGF pathway has also been of interest in BC and MD research and SNPs in the genes *IGF1* [247-250], *IGF1R* [250], *IGFBP1* [247] and *IGFBP3* [251] have been found associated with MD to mention some. In addition, it is often seen that the levels of the respective gene products are also associated with MD [250,252]. In a more recent study, a meta-analysis of five GWAS studies revealed a SNP in *ZNF365* associated with MD [253]. Consistent with being associated with BC risk the SNPs in *LSP1*, 8q, *MAPK3K1*, *H19*, *CASP8* [254] and *TOX3* [255] are associated with MD.

2 Materials and methods

2.1 Sample population

Table 2: Table describing the sample population, name of study, number of samples from healthy women and women with breast cancer, a short description of the material and which papers they appear in.

Material name	N (healthy)	N (breast cancer)	Description	Paper
Tromsø Mammography and Breast Cancer study (TMBC)	964*/ 433**		Healthy norwegian postmenopausal women in the age group 55-71 years. Information on diet, menstrual and reproductive factors are obtained through an extensive questionnaire . All women have negative mammograms with no sign of cancer.	I* & III**
Mammographic Density and Genetics (MDG)	120	66	Norwegian pre- and postmenopausal women in the age group 22-87 years, with and without breast cancer. Information obtained on menstrual and reproductive factors, including serum hormone levels and mammographic density.	II& III

2.1.1 TMBC

The Tromsø Mammography and Breast Cancer study (TMBC) consists of samples collected from Norwegian women living in the municipality of Tromsø, Norway (Table 2) [192]. The samples were collected in the spring of 2001 and 2002 as part of the population-based Norwegian Breast Cancer Screening Program (NBCSP) at the University Hospital of North Norway. The women, between the age of 55-71 years, were all postmenopausal and healthy with no sign of cancer. At the time of mammography screening the women were interviewed by a trained nurse regarding reproductive and menstrual factors, previous history of cancer, and the use of HT and other medications. For women not currently using HT, serum levels were obtained for estradiol, testosterone, DHAE, Vitamin D and prolactin and glycoprotein SHBG. All women signed an informed consent. The study was approved by the National Data Inspection Board and the Regional Committee for Medical Research Ethics.

2.1.2 MDG

The Mammographic Density and Genetics (MDG) study consists of samples from women of Norwegian origin (Table 2) [256]. The samples were collected between 2002 and 2007 from women either recruited through the NBCSP or a breast diagnosis centre based on referral for a second examination due to irregularities. The women, 22-87 years of age, included both premenopausal and postmenopausal who were healthy or had breast cancer. To be eligible the

women could not have breast implants, be pregnant or lactating or currently on anticoagulant therapy. Women who had a history of breast cancer but no current malignancy were included in the group of healthy women. However, for these women the biopsy was collected from the contra lateral breast. Mammography images were obtained for each woman in addition to information about weight, parity, HT use and family history of breast cancer. The study was approved by the Regional Committee for Medical Research Ethics, IRB approval number: S-02036.

Genetic variations and mutation may reside in the inherited germline DNA or in somatic cells, and may be identified in blood and tissue respectively. For SNP genotyping we collected blood samples, and for gene expression analysis we used core biopsy samples from breasts of controls and cases.

2.2 Blood samples and plasma analyses

2.2.1 TMBC

On the day of the mammographic screening non-fasting venous blood samples were drawn in two 9 mL citrate vials for plasma extraction. The tubes were centrifuged at 3000 rpm for 15 minutes and deposited into 2 mL cryo tubes prior to storage at -70°C.

The plasma samples were shipped to the Nutrition and Cancer Group at IARC[4] in Lyon, France, who specialize in hormone assays. The levels of[5] IGF1 and IGFBP3 were measured by Enzyme-Linked Immunosorbent Assay (ELISA) [257]. Estradiol, estrone, androstenedione and prolactin were measured by direct double-antibody radioimmunoassay whereas testosterone and DHEAS were measured by direct radioimmunoassays. SHBG was measured by direct "sandwich" immunoradiometric assay. The general principles of radioimmunoassay are described in detail elsewhere [258]. For the women who had concentrations that were not detectable, the lower detection limits for the respective assays were recorded.

[4] IARC: International Agency for Research on Cancer

[5] DHAE(S): Dehydroepiandrosterone (sulfate), SHBG: Sex Hormone Binding Globulin, LH: Luteinizing Hormone, FSH: Follicle Stimulating Hormone, IGF1: Insulin-Like Growth Factor 1, IGFBP3: Insulin-Like Growth Factor Binding Protein 3.

2.2.2 MDG

Three blood samples were obtained from each woman. The blood was drawn in SST tubes with gel, which were left for 30 minutes before they were centrifuged at 2000 G for 10 minutes. The serum was subsequently aliquoted and stored at -20°C. The serum levels of estradiol, testosterone, progesterone, SHBG, LH, FSH and prolactin were measured with electrochemiluminescence immunoassay (ECLIA) at the Department of Medical Biochemistry, Oslo University Hospital Rikshospitalet, a laboratory accredited according to ISO-ES 17025. The levels of estradiol, LH and FSH were used in combination with information on HT use and age, in order to determine menopause status.

2.3 Biopsy collection

2.3.1 MDG

The women participating in the MDG study had two biopsies taken with a 14 gauge needle. For the healthy women the biopsies were taken from an area with some mammographic density identified by ultrasound, to avoid biopsies consisting of purely adipose tissue. For the women with breast cancer the biopsy was taken from the tumor. Biopsies were collected at 6 hospitals with slightly varying methods. Oslo University Hospital (OUS) Radium Hospital snap froze biopsies which were subsequently frozen at -80°C. The other hospitals soaked one biopsy in RNA later and the other in 70% alcohol for RNA and DNA extraction respectively. Once the biopsies reached the Department of Genetics at OUS Radium Hospital they were stored at -20°C.

2.4 Mammograms

2.4.1 TMBC

The mammograms were obtained at the NBCSP in Tromsø. For each woman two screens for each breast were performed, one craniocaudal (CC) and the other medio-lateral oblique (MLO) according to NBCSP manual [259]. The left breast was chosen because most women are right handed in which case the pectoralis muscle is more visible in the mammogram of the right breast. However, the concordance between the right and left breast has been found to be high [260]. For assessment the images were digitized using a Cobrascan CX-812 scanner.

2.4.2 MDG

The mammograms were obtained from local radiologists according to routine practice. Up to two CC mammograms were obtained for each woman. In the cases where two mammograms were obtained, both were assessed and the average used. For assessment the images were digitized using a Kodak Lumisys 85 scanner (Kodak, Rochester, New York).

2.4.3 Mammogram assessment

For both the TMBC study and the MDG study the Madena computer-based threshold method developed at the University of California [203] was used to assess the mammograms. This method utilizes the digitized image which is viewed on a screen. In brief, the Madena program works as follows: with the use of an outlining tool an experienced reader defines the total breast area, which is then converted through calculations into a value in pixels. To assess density, a region of interest (ROI) is defined, which includes areas of density but excludes radiodense artifacts such as the pectoralis muscle, prominent veins and fibrous strands. The reader then uses a tinting tool to apply a yellow tint to dense pixels within the region of interest that has grey levels at or above some threshold X and below a pixel value of 255. The reader searches for the best threshold where all pixels X within the region of interest are considered to represent mammographic densities. The mammographic density may be presented as either absolute density (ABDEN), which is the total number of pixels in the ROI or as percent density (PDEN) which is the ROI (absolute density) divided by the total breast area multiplied by 100. The latter was the measurement used for subsequent analysis referred to as mammographic density (MD). For the MDG study the absolute density reproduction rate was 0.99, similarly, for the TMBC study the intra-reader agreement rate was 0.86, and the inter-reader agreement was 0.86.

Box-Cox

Statistical methods may require the data tested to conform to certain assumptions. Parametric methods assume that the data has a normal like distribution for the modeling of the relationship between variables. This might however not always be the case, which can be seen as a non-linearity in scatter plots of the two variables, or as a heterogeneous variance. The solution to the problem is to transform the data to make it more normally distributed. Common methods involve log or power transformations or both used together known as Box-Cox transformations. The Box-Cox transformation was developed by the statisticians George E. P. Box and Davis Cox in 1964

[261]. Briefly the method utilizes the parameter λ as the variable exponent (eg X^{λ}). When λ=0 the log transformation is used. Different λ values are applied to the data and one way to select λ is to use the λ that maximize the logarithmic of the likelihood function λ, or minimizes the sum of squares residuals. In paper I we performed a Box-Cox transformation of the mammographic density variable and the optimal λ was estimated by solving non-linear least square problems.

2.5 In silico analyses: candidate gene and SNP selection

In paper III we used a candidate gene approach. Candidate gene approaches utilize the *a priori* knowledge of our phenotype of interest in order to select genes with an increased probability of affecting the phenotype. This type of approach ensures that a limited set of genes is tested and may thus limit the need for multiple testing correction and also reduce the cost. Due to the importance of estrogen in both the development of the normal breast and breast cancer we chose to focus on the estrogen pathway for the selection of the candidate genes. Subsequent to the gene selection, known genetic variations in the form of single nucleotide polymorphisms (SNPs) in or surrounding these genes were identified.

2.5.1 Gene selection

We searched the literature and molecular databases for candidate genes and SNPs. Entrez Pubmed was used for the literature search with the following key words: (1)"Estradiol and mammographic density", (2)"Estradiol and ER", (3)"SNPs and (1)&(2)", (4) "Mammographic density", (5) "Breast density", and (6)"Single nucleotide polymorphism and Estrogen/Progesterone". In addition we looked for genes in the estradiol and estrogen receptor pathway using CGAP [262] (provided by Biocarta), iPATH™ [263] and PathwayAssist/ PathwayStudio® [264] (licensed software by Ariadne Genomics). Through these processes a total of 281 genes were determined to be of interest, and carried over to the following SNP selection step.

2.5.2 SNP selection

SNPs residing in the 281 genes were identified using Ensembl [265] and SNPper [266]. These SNPs were analyzed in SIFT [267] (Sorting Intolerant From Tolerant) developed at Fred Hutchinson Cancer Centre in Seattle [268] to identify SNPs with impact on protein function based on sequence homology and amino acid property. HapMap [64] was also used to retrieve

SNPs from the candidate genes. We selected the European population for identification of SNPs within the candidate genes, since this population would resemble our study population the most. The SNP data obtained from HapMap was subsequently analyzed by De Bakkers "tagger"-test [269] which is implemented in the software Haploview [270]. The test returns information about the SNPs and their correlation or linkage disequilibrium (LD). If the SNPs are in LD, one SNP within a haplotype is chosen, often arbitrarily, and defined as the htSNP. With these approaches we selected a total of 1001 SNPs in the 281 genes for further analyses. (Figure outlining the gene and SNP selection process is presented in the supplementary of paper III).

2.6 SNP genotyping

Over the past decade there has been a rapid development in the genotyping field. From Botstein's experiments on restriction fragment length polymorphisms in 1980 [271] to today's whole genome wide association studies (GWAS) of up to 4,3 million SNPs [272]. SNP genotyping has become a widely used tool for the identification of genetic markers that may predispose for complex disorders [61,273] of which some have already been identified [274].

2.6.1 Taqman real-time PCR genotyping

The amplification of a selected piece of DNA is made feasible with the Polymerase Chain Reaction (PCR), a procedure developed in 1983 by Kary Mullis. All PCR rely on the same principles of DNA denaturation, primer annealing and primer extension/elongation steps. In the Taqman real-time PCR procedure fluorescent dyes are incorporated to detect DNA amplification while it occurs, rather than detecting total amplification product at the end of amplification as in conventional PCR. The method is described in detail elsewhere [275], in brief it works as follows: an oligonucleotide probe containing a reporter fluorescent dye on the 5' end and a quencher dye on the 3' end is constructed. Due to the close proximity, the quencher dye reduces the fluorescence from the 5' reporter dye through fluorescence resonance energy transfer (FRET). The probe is designed to anneal downstream of the target DNA sequence. A primer specific for the DNA sequence attaches and elongates up to the site of the probe and as it elongates the probe is cleaved thus separating the reporter dye from the quencher dye and increasing the fluorescent signal from the reporter. In addition the probe is removed from the target DNA strand allowing the primer to continue to the end of the strand. For each cycle additional reporter dye molecules

are cleaved and thus the intensity of the fluorescence reflects the amount of amplicon produced [276].

2.6.2 Sequenom technology and genotyping

The genotypes analysed in paper II and III were generated using the Sequenom [277] MassARRY® platform and iPLEX genotyping assays [278]. Briefly the iPLEX genotyping technology works as follows. The first step involves up amplification of the sample DNA using PCR to enable SNP detection. During the amplification, special pre-designed MassExtend primers anneal up to the polymorphic site and extend dependent on the polymorphism. Either a one or two base extension is incorporated, although, since the masses that separate these bases are small an additional mass modified terminator is incorporated. The nucleotide molecular weight differences make it possible to distinguish them on the mass spectrometer utilized in the MassARRAY platform (Figure 14).

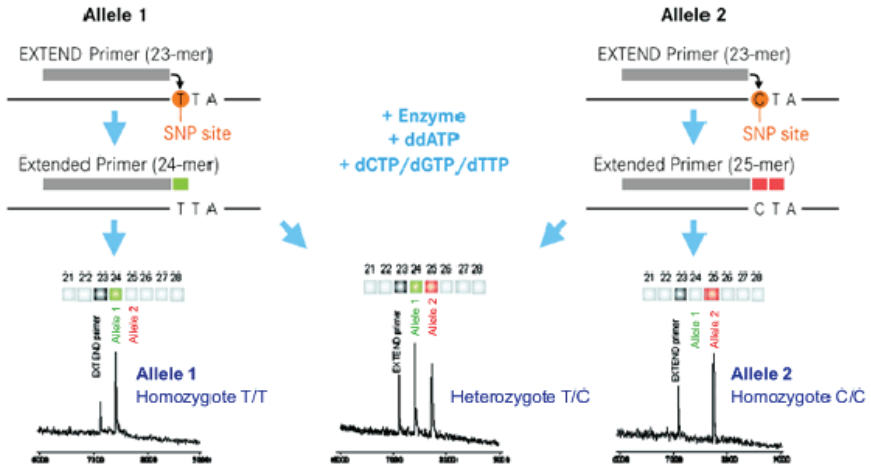


Figure 14: The detection of homozygote or heterozygote genotype with the iPLEX genotyping assays using allele specific primers. The product consists of either a one or a two base extension. From [277].

The underlying technology of the MassARRY® platform is matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). In brief this technology works as follows: an ionization source pulses laser onto the amplified sample DNA which is embedded in a crystalline structure called the matrix which in turn is deposited on a conductive chip often made of metal. We used the spectro-CHIP spotted with matrix in a 384 format. The sample which is in vacuum and under the influence of a strong electric current will, due to the laser, evaporate into gas. The plume drifts through the electric field and passes a mass analyzer before it reaches the detector. The mass analyzer separates and resolves the ions in the plume based on their mass/charge ratio and their time of flight through the electric field. Lighter ions travel faster than heavier ions and will be detected first. The last step is the detector which transforms the ion current into a mass spectrum through determination of the abundance and mass of the components in the sample. The sequenom genotyping method is recognized as cost-effective, high throughput, flexible and for producing high quality data with a high level of reproducibility [279].

Genotype processing: quality control

SNP genotyping quality control involves removal or flagging of SNPs or samples that have not passed the quality control threshold specified. In paper II and III SNPs that were estimated not to be in Hardy Weinberg equilibrium (HWE) were flagged. HWE has similarities with Linkage Equilibrium (D=0), the opposite of LD (D=1), in that it implies that SNPs at different loci are randomly associated [63]. HWE is a rarely obtained ideal which implies that the genotype frequency in a population remains constant, or in equilibrium, from generation to generation. Deviation from HWE is due to disturbing factors such as mutations, selection, random genetic drift, inbreeding or genotyping error [63]. Small populations, however, are seen to have genotype frequencies that are not in line with the expected frequencies defined by HWE [280]. Additionally, SNPs with a call rate of less than 80% (more than 20% missing) and a minor allele frequency of less than 5% were removed. Similarly, samples with more than 30% missing genotypes were removed.

2.7 Microarray technology

With the introduction of DNA microarray technology, genome wide scans have been made possible which means that up to 10^5 to 10^6 genomic loci can be tested simultaneously in one sample from a single individual [281]. Currently arrays for the detection of a variety of genetic variations exist and in paper II and III we used gene expression and SNP array technology.

2.7.1. Gene expression

In paper II and III we used the Agilent Human Whole Genome Oligo Microarrays 44K (G4110A) from Agilent technologies to obtain gene expression from biopsies obtained in the MDG study. The micro array is a two colored array (Figure 15) spotted with oligonucleotides of 60-mer length, referred to as probes. A two-colored array has a reference DNA placed on the array along with the sample. We used a common reference on all the arrays and to ensure a general background for the test samples a commercially available reference made from 10 different cell lines was chosen. In the experiment the reference and the samples are labeled differently (red and green fluorescence) before they are mixed in solution and placed on the microarray slide. During 12 hour hybridization incubation the reference sample and test sample compete for hybridization to the 60-mer probes on the array. Thus, if there is equal expression of a gene in both, this will be seen as approximately equal binding to the probes. However, if the test sample has regions that are deleted only the fluorescent dye of the reference will be present at the spots representing these loci and vice versa.

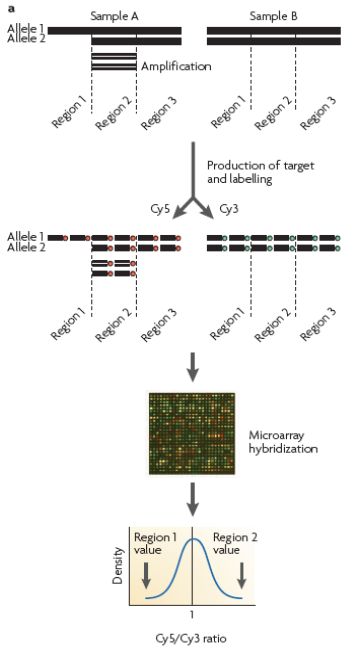


Figure 15: Schematic diagram of a two colored gene expression array, A is the test sample, and B is a control/reference sample. Both sample A and B are up-amplified and labeled with Cy5 or Cy3 making their detection on the micro array feasible. From [281]

Gene expression data processing: normalization and imputation

Gene expression data are obtained from a set of microarrays, and one sample or more are analyzed per array. However, since not all samples will fit on one array, several are run. This gives rise to batch effects and other artifacts dependent on the environment in which a given array was processed. Additionally, differences between samples may be observed due to differences in sampling and RNA quality. To be able to analyze the gene expression data from the arrays together the expression data needs to be modified to account for these intra- and inter-gene expression differences. This is referred to as normalization of microarrays. Locally weighted scatterplot smoothing (Lowess) [282] was used for this purpose. The Lowess method fits simple regression models to local subsets of the expression data and point by point builds up a function to describe the variation in the whole dataset. A constant defined by the regression models is subtracted from the original data giving rise to new normalized values.

2.7.2 GWAS

In paper II we used a genome wide SNP genotyping array from Illumina to perform a genome wide association study (GWAS). We used Illuminas Human 1 Genotyping Beadchip which contained assays for 109,365 SNPs (Figure 17). As described elsewhere [283], the array utilizes the following technologies: a) whole genome amplification (WGA); b) hybridization of sample DNA to a specific and sensitive oligonucleiotide probe array; c) array-based enzymatic SNP scoring assay and; d) sensitive signal amplification, enabling readout (Figure 16).

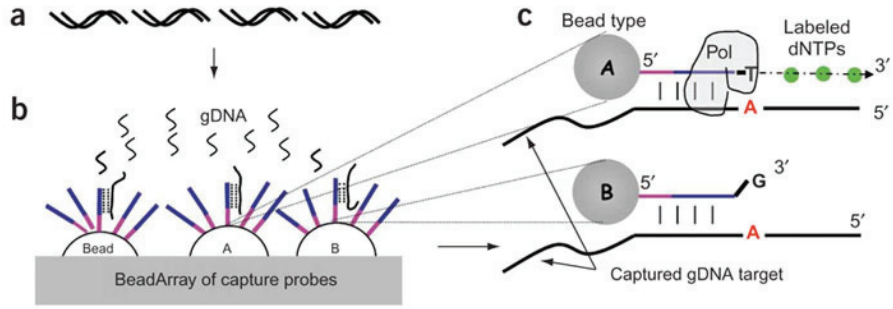


Figure 16: The Illumina 109K array for whole genome genotyping [283]. a) The amplified sample DNA is fragmented by enzymes prior to b) hybridization to the specific and sensitive oligonucleotide probes on the BeadArray. c) Allele specific bead types are used to discriminate against allele A and allele B. The probes are only different in the 3' end which enables allelic discrimination during polymerase extension following hybridization to a perfectly matched allelic target. The intensity ratio of the two bead types determined the genotype of a given sample. Illustration from [283].

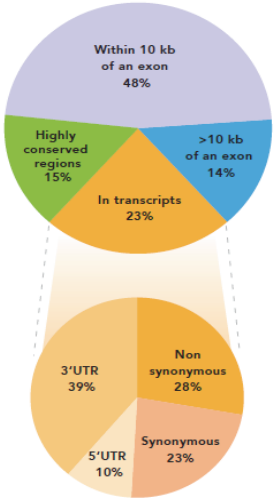


Figure 17: Pie chart depicting the categories of the SNPs included on the Illumina 109K Human 1 genotyping Beadchip array. The SNP selection on the array is based on two approaches; a direct approach and an indirect approach [61]. The direct approach involves SNPs that are located in promoter or coding regions while the indirect approach features htSNPs that are evenly spaced throughout the genome.

From www.illumina.com.

2.8 Statistics

2.8.1 ANOVA

In its simplest form, Analysis of Variance (ANOVA) is a statistical method that tests a null hypothesis (H_0) of no difference in means between two or more groups. When comparing two means, ANOVA is a generalized form of a t-test which utilizes the student's t-distribution in order to compare 2 groups with respect to the mean. However, when two or more groups/means are tested at one time it is referred to as an F-test which utilizes the F-distribution. This one-way ANOVA (one factor) can test whether there is an influence of the factor of interest on the phenotype, thus leading to a conclusion for the whole factor. The differentiation of the groups and possibly pair wise post-hoc tests need to be performed in order to specify the structure of the factor levels. In paper II a two-tailed t-test was used while in paper III a one-way F-test was used to analyze the significance of the association between SNPs and MD. The aforementioned tests are so-called parametric tests which assume normal distribution of the variable analyzed.

2.8.2 Correlation

A correlation analysis tests for association between two continuous variables. Pearson's partial correlation was used in paper III. The advantage of partial correlation is the ability to adjust for covariates which may influence the observed relationship between our two variables of interest [284].

2.8.3 Chi-square and Fishers exact test

When testing the association of categorical variables, chi-squared and Fishers exact tests are suitable. In a chi-square test the data is placed in a two-way frequency table. A test statistic X^2 is calculated based on the observed and expected frequencies, the latter is estimated from the data [284]. The H_0 is that there is no difference between the (treatment) groups. When the sample size is too small or has low cell counts the Fisher exact test should be used. This test keeps the marginal frequencies fixed and generates all possible table frequencies that would result in such totals. A probability for each table is calculated under the assumption of a true H_0. The overall probability is subsequently calculated for the observed data based on these probabilities [284]. Chi-squared and Fishers exact test were used in paper II.

2.8.4 Gene expression analysis: SAM

Significant Analysis of Microarray (SAM) [285] is a statistical test to identify genes significantly expressed in a set of microarray data. In paper II we used Quantitative SAM for the identification of differentially expressed genes according to MD as a continuous variable. Briefly, the method entails the identification of genes with significant changes in expression through a set of gene specific t-tests. A score is given to each gene based on the change in gene expression relative to the standard deviation obtained from repeated measurements of that gene. According to a set threshold the score is used to determine the significance of the gene. Due to the fact that analysis of microarrays entails testing a large number of genes and thus increases the chance of detecting false positives, multiple-testing correction is necessary. The FDR method for multiple-testing correction is incorporated in SAM and applied to data permutated from the original data. In paper II we performed 500 permutations. Based on the FDR threshold the size of the gene sets vary from small to large [285]. To test the robustness of the results obtained in SAM analysis, regression analyses were performed on selected probes with MD as the response variable.

2.8.5 SNP analysis: haplotype estimation

SNP data does not tell us the phase, but haplotypes can be estimated by different approaches like E-M-algorithm and PHASE [286]. Using haplotypes for association testing can be more powerful than testing each single SNP [287]. In paper I we looked at haplotypes associated with levels of MD. Since the data was collected at the SNP level the haplotypes had to be estimated. For this purpose we used the SAS/GENETICS and the PROC HAPLOTYPE procedure. The implemented expectation maximization (EM) algorithm [288] estimates the maximum likelihood of the haplotypes and their frequencies in a given population with the assumption that they are in HWE. The level of LD between SNPs within the haplotypes was tested through generation of a likelihood ratio test with the null hypothesis of no association between SNPs. In brief the method works as follows; the two locus haplotypes are first estimated, the ones that have a lower frequency than a given threshold are discarded, and then the remaining haplotypes are expanded to the next locus by formation of all possible three locus haplotypes. As the number of loci increases the number of possible haplotypes grows exponentially. We used a haplotype frequency cutoff threshold of 0.05, which meant that haplotypes with a lower frequency in our population were not carried through to the next haplotype trend regression analysis in SAS 9.1.

2.8.6 Regression analysis

Regression analyses were used in all the papers in different forms. Regression is in its simplest form a straight line and is referred to as a linear model, representing the relationship between a dependent variable and an independent variable. The equation for a dependent variable Y as a linear function of the independent variable X is given by:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Where α is the intercept, β is the slope of the line, ε_i is the individual residual error term and i is the individual sample. For example, the genotype for an i'th individual is given by X_i, and explains the changes in the i'th individual's MD which is given by Y_i, when assuming the additive model.

SAS genetics

In paper I we used SAS® 9.1 programming software and SAS/GENETICS in combination with the SAS/STAT procedure PROC REG for haplotype-trend regression for the association analysis of SNPs and MD. In SAS/GENETICS the estimated haplotypes are processed and formatted so they can be used in the regression procedure in SAS/STAT. The haplotype trend regression was performed using the PROC REG procedure which used the data obtained on the estimated haplotypes for association analysis of each haplotype and our response variable MD. In addition to the analyses of the haplotypes, single SNPs were analyzed for association with MD using general linear models (PROC GLM). All analyses were adjusted for age and BMI due to their importance for MD and results were rendered significant if the p-values were < 0.05.

eQTL

In paper II we used the programming software R 2.9.0 and the package "eMAP" v1.1 for the expression quantitative trait loci (eQTL) analysis. The package has the option of analyzing associations of distant SNP and expression probes (*in trans*) as well as SNPs and probes within a limited distance of each other (*in cis*). *In cis* analyses were performed which meant that the SNP and gene probe analyzed had to lie within 10^6 base pairs of each other. From the equation above the gene expression is given by Y, and the SNPs are given by X. The genotypes were coded into 0, 1and 2 (major homozygote, heterozygote, minor homozygote) and analyzed for a linear relationship with the gene expression probes, also referred to as the additive inheritance model.

Although a time-consuming analysis, the eMAP package is written in the programming language C and implemented in R which makes it faster and more efficient than many other methods.

SNPassoc

In paper III, R 2.10.1 programming software was used with the package "SNPassoc". The inherent function WGassociation utilizes generalized linear models to estimate the association between SNP and MD under five different inheritance patterns (Figure 18): dominant, recessive, additive, overdominant and codominant. In addition to the analyses involving SNP and MD, we modified the program to be able to run SNPs and gene expression data and thus performing an eQTL analysis. For comparison, the previously mentioned regression analyses so far have only involved analyses under the additive model.

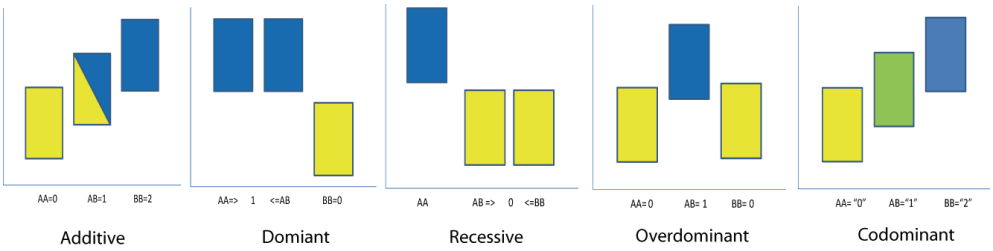


Figure 18: The five inheritance models applied in WGassociation in SNPassoc. Phenotype on the y-axis.

Multivariate analysis

In addition to the previously mentioned univariate analyses, a multivariate analysis was applied to the SNPs in paper III. The approach somewhat resembles a haplotype analysis in that it analyses a predefined set of SNPs according to a response variable, however, that is where the similarities end. The multivariate analysis does not assume that SNPs in a given set of SNPs are in LD, hence, SNPs that reside on different chromosomes and with great distance between them can be placed in the same group. The purpose is to determine if the SNPs in combination can influence the phenotype at a greater magnitude than each SNP alone. This type of analysis is useful when testing specific pathways of interest and groups of genes such as those from a cluster analysis. The multivariate analysis was performed in R 2.12.0 with the package "Globaltest" [289]. The equation for a multivariate regression model fits several variables and is given by:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_n X_{ni} + \varepsilon_i$$

However, the multivariate approach may hold more power than a simple univariate analysis, and unless many groups of markers are tested the multiple testing correction problem is circumvented.

2.8.7 FDR multiple testing correction

With the increasing amount of genetic markers being tested in genetic studies, there is an increased probability of detecting false positives (type 1 error). This has established the need for multiple testing correction procedures, to correct the level at which we define a result as significant. A p-value threshold may be used to define the statistical significance of a test, that is, the probability of observing the test statistic or one that is more extreme given that the null hypothesis of no association is true. Hence, theoretically, by using the common threshold 0.05 suggested by Fisher in 1925 [290], we expect that 5% of our results are false positives. This does not pose a major problem in relatively small studies when 10-20 markers are tested, in which one would expect 1 significant result to be a false positive. However, larger studies and whole genome scans testing 10^5-10^6 markers are becoming increasingly common and the amount of false positives increases linearly with the number of markers tested. To correct for this phenomenon a variety of statistical methods have been developed. The most common methods control the family wise errors rate (FWER) or the false discovery rate (FDR). FWER is the probability of making one or more false discoveries, while FDR is the expected proportion of false positives among all significant discoveries [291]. FDR is thus considered a less conservative method than the FWER methods (e.g. Bonferroni) which may result in an increased amount of false negatives. FDR was used in paper II and III.

2.9 Gene ontology databases

2.9.1 DAVID

In the SAM analysis conducted in paper II a list of genes was identified as potentially significant. To investigate the biology behind the gene list, Database for Annotation, Visualization and Integrated Discovery (DAVID) [292-294] was used. This database provides information about genes in the form of molecular function, biological processes and pathway visualization, amongst others. The inherent pathway visualization is offered by Kyoto Encyclopedia of Genes and Genomes (KEGG) [295].

3 Results in brief

We studied SNPs from selected pathways for involvement in MD development in paper I, II and III, while in paper II we also used SNPs from a whole genome analysis approach. Additionally, gene expression data was integrated with SNP analyses in paper II and III to further investigate the molecular background in MD.

Paper I

Genotypes and haplotypes in the insulin-like growth factors, their receptors and binding proteins in relation to plasma metabolic levels and mammographic density

(Biong et al. BMC Medical Genomics 2010)

The development of the breast and MD are both influenced and increased by various growth factors. In this paper we focused on the influence of the insulin-like growth factor genes on MD. The IGFs have previously been reported to be of importance for normal mammary growth, MD and also cancer of the breast. The ability of *IGF1* and *IGF2* to stimulate cell proliferation and inhibit cell death, may account for the reported implication of *IGF1* in several cancers and MD. Insulin-like growth factor binding proteins (IGFBPs) are as the name implies proteins that bind IGFs and prolong their half-life; IGFBP3 is the principal carrier of IGF1 and IGF2. The resultant complex may bind an acid labile subunit (ALS). The addition of ALS to the binary complex creates a tertiary complex which prolongs the half life of IGFs from 30-39 minutes to more than 12 hours. IGF1 and IGF2 exert their growth promoting effects through their receptors, and while both may bind to insulin growth factor receptor 1(IGF1R), only IGF2 binds to insulin growth factor receptor 2 (IGF2R).

In paper I, selected SNPs (n=24) harboring the genes; *IGF1, IGF2, IGFBP3, IGFALS, IGF1R* and *IGF2R* were analyzed in a population of 964 Norwegian postmenopausal women. Both single SNPs and haplotypes were analyzed with respect to MD, IGF1 and IGFBP3 serum levels, using linear regression methods. The results from the single SNP analysis revealed one SNP in *IGF1R* associated with the plasma level of *IGFBP3* while three other SNPs in the same gene were associated with MD. Two SNPs in *IGF2R* were associated with both the level of *IGF1* and

IGFBP3. One SNP in *IGFALS* and one in *IGFBP3* were both associated with serum levels of *IGFBP3*, the latter indicating association *in cis*.

Haplotype analysis revealed the existence of haplotypes in each gene. Analysis of these haplotypes revealed an association of a haplotype variant in *IGF1* with MD, however, after stratification by HT it was no longer significant. A rare haplotype in *IGF2* was associated with increased levels of IGF1 and IGFBP3. Two haplotypes in *IGF2R* were found associated with decreased levels of IGF1, although one association was no longer significant after HT stratification. The two haplotypes in *IGFBP3* were both associated with the levels of IGFBP3, giving more power to the results found in analysis of the single SNPs. HT stratification revealed that most of the significant results remained significant either for HT users or non-users but seldom both groups.

We conclude that there are haplotypes in the studied genes and that four of in total six haplotypes were associated with serum levels of the gene products studied and/or MD. Of special interest were the results from analyses of SNPs and haplotype variants in *IGFBP3* which were found associated with serum levels of IGFBP3, presenting a strong indication of regulatory properties of these SNPs *in cis*.

Paper II

Expression levels of uridine 5'diphosphoglucoronosyltransferase genes in breast tissue from healthy women are associated with mammographic density.

(Haakensen, Biong et al. Breast Cancer Research 2010)

Gene expression technology holds promise of a better understanding of the molecular underpinnings of diseases that have so far not been sufficiently explained through other methods. Breast cancer is one such disease and although the causative agents have been identified for some variants of this disease (i.e. *BRCA1/2* mutation) most are unidentified. Detection of early events in breast cancer development is important for good prognosis and for this purpose MD has been suggested used as a proxy phenotype. In paper II we aimed at elucidating the nature of the difference in gene expression pattern seen in healthy breasts with low and high MD, which may explain the relationship between MD and BC risk. Gene expression data was analyzed

statistically using the method quantitative Significant Analysis of Microarrays (SAM) in combination with linear regression. Additionally, involvement of SNPs on expression pattern was also estimated through expression quantitative trait loci (eQTL) analysis.

Of the 9767 expression probes analyzed, 25 probes representing 24 genes were differentially expressed between two groups consisting of healthy women with high and low MD. Upon analysis of these genes with the gene ontology tool DAVID, no terms or pathways were overrepresented in a significant manner. However, of particular interest it was noted that three genes from the *UGT* gene family (*UGT2B11, UGT2B10* and *UGT2B7*) were represented in four of the 25 probes. The *UGT* gene family is involved in detoxification of toxic compounds including hormones by clearing them from circulation, thereby protecting tissues from the growth promoting effects. These particular *UGT* genes were seen to be down-regulated in the samples with the highest MD, which was found to be the case for the tumor samples as well. Upon stratification for hormone therapy use and age it was revealed that the association remained significant for young women <50 and women >50 currently using HT. Thus, decreased *UGT* expression in breasts of women currently under the influence of steroid hormones may increase MD and possibly also BC risk. Genotypes and expression were subsequently analyzed for the identification of eQTLs. Only associations *in cis* were considered which resulted in the identification of one SNP (rs1828705) in *UGTB10* associated with two expression probes in *UGT2B10* and *UGT2B7* respectively. This suggests that variations on DNA level may influence the associations made on the mRNA level.

In summary, a set of genes was identified to be differentially expressed with regards to whether the biopsies were taken from breast with high or low MD. The nature of certain genes, such as the *UGTs*, within this list could greatly influence estrogen metabolism in healthy women who are under the influence of female hormones. The fact that the *UGT* genes were significantly down-regulated in women under the influence of steroid hormone (premenopausal or user of HT) with high MD, indicates that these two parameters may act together and give rise to high MD.

Paper III

Candidate SNP analysis integrated with mRNA expression and hormone levels reveal influence on mammographic density and breast cancer risk

(Biong et al. submitted to Cancer Epidemiology Biomarkers and Prevention)

There is a growing need for the identification of genetic markers that can predict whether a woman will develop breast cancer at some stage in her life. Mammographic density is a well-established risk factor for breast cancer, but the biology underlying MD is complicated. MD is influenced by reproductive life style factors, diet, age, growth factors and genetics. The genetic component of MD is estimated to be between 30-60%.

Based on this we set out to identify SNPs with potential impact on MD which may also be used as early markers of BC. The estradiol pathway was chosen for the selection of candidate genes, in which candidate SNPs were selected. A total of 257 candidate SNPs were analyzed with regard to MD and serum hormone levels. In addition we obtained mRNA expression in biopsy from non-diseased breast tissue which was added to the genotype analyses. With the addition of the mRNA to the genotypes we aimed at gaining power to identify low-penetrant SNPs and also possible mediation effects.

Two independent sample materials made a discovery/verification approach feasible. The SNPs were analyzed univariately under different inheritance models using regression method. Multivariate analyses of the SNPs were also performed with the use of Globaltest to reveal sets of SNPs associated with MD. Univariate analysis results revealed associations of 28 SNPs with the levels of MD of which seven SNPs were also associated with gene expression transcripts *in cis*. SNPs harboring or in the vicinity of genes which are directly influencing the estradiol pathway were of special interest such as the uridine 5'diphosphoglucoronosyltransferase (*UGT*), sulfotransferase (*SULT*) and hydroxysteroid dehydrogenase (*HSD*) gene members. These gene families were represented by six of the 28 SNPs which were associated with MD in healthy postmenopausal Norwegian women. In paper II we discovered an association of the expression of *UGT* family members with MD, an association confirmed at the DNA level in the current paper. We also confirmed the association of the *HSD3B1* SNP (rs1047303) which is previously associated with MD in several other independent studies. Also SNPs in other interesting genes

such as the matrix metalloproteinase 2 (*MMP2*) and breast cancer antiestrogen resistance 1 (*BCAR1*) were found associated with MD. The multivariate globaltest analysis was run on sets of SNPs defined from univariate analysis results. Based on HT use and two defined p-value thresholds, six SNP sets were analyzed. Of these six, two sets consisting of 7 and 35 SNPs respectively were significantly associated with MD.

The findings in this paper are based on studies of small sample size and their statistical significance is based on replication and not on multiple testing corrections. However, the nature of the genes and the importance of the estradiol pathway in MD and BC warrant their further investigation.

4 Discussion

4.1 Discussion of main findings

4.1.1 Involvement of hormone and growth factor pathways in MD

The interaction between mammographic density and genetic variation in members of the insulin like-growth factor pathway were analyzed in paper I while members of an extended estradiol pathway were analyzed in paper III. The interrelation between these two pathways is illustrated in Figure 11 and Figure 19.

Main findings in estradiol pathway

In papers II and III several members of the estrogen pathway were identified associated with MD, and included SNPs in *UGT*s, *SULT2A1* and *HSD3B1* as well as the gene expression of *UGTs* and *ESR1*. As previously mentioned these may all be important in determining the level or action of estrogens in the breast tissue. As a consequence these are often targeted in studies regarding abnormalities of the breast.

In paper II we found a set of 24 genes differentially expressed according to MD in breast tissue of healthy pre- and postmenopausal women. Amongst these genes three *UGT* genes were identified (*UGT2B11*, *UGT2B10* and *UGT2B7*). These *UGT*s were inversely associated with MD levels. Interestingly, comparing the women with low MD, high MD and women with BC showed that the two latter groups had similar expression of these *UGT*s, suggesting that these two groups are more alike than the counterpart with low MD. Additionally, stratifying on age and HT use revealed that women older than 50 years currently using HT where the only group were *UGT* genes were still differentially expressed according to MD. Confirming these results in regression analyses using stepwise regression resulted in *UGT2B10* being the only one remaining in the model. The association remained significant for women aged <30 and women aged >50 and currently taking HT when stratifying the regression analysis according to age and HT use. This suggests that low expression of *UGT*s in an environment with high levels of female hormone may increase MD and possibly also BC risk. Further studies would be needed to determine if the *UGT*s confer an independent risk of BC or if the risk is mediated by MD.

Due to the strong evidence of heritability in MD we set out to investigate the genetic variability in the *UGT* genes whose expression was found associated with MD. One study has so far reported an association of SNPs in UGT genes associated with MD [244]. Upon analysis of SNPs we identified one *UGT* SNP associated with the expression levels of two *UGT* transcripts differentially expressed according to MD. However upon analyzing SNPs and expression transcripts in paper III, the association of *UGT* SNPs with *UGT* transcripts did not hold. There might be several reasons for this, with the main reasons being that the *UGT* genes analyzed in paper II were represented by SNPs derived from both GWAS and candidate gene study in which none of the *UGT* SNPs were represented on both in addition to low power due to few samples. In paper III only SNPs from the candidate gene approach were analyzed. The analysis approach in the papers also differed. In paper II we selected to only study *UGT* SNPs according to *UGT* transcripts while in paper III we investigated a larger set of SNPs according to a larger set of transcripts utilizing a verification approach. In paper III we identified SNPs in *UGT* genes (*UGT2B28, UGT2A1, UGT2B15*) associated with MD, and two of these (*UGT2B28, UGT2B15*) were associated with transcription factors *in cis* that were also associated with MD. The findings of SNPs in these *UGT* genes through validation analyses with an independent sample set strengthen the finding in paper II. However, larger epidemiological studies with the appropriate power would be needed to verify these findings. Also, of special interest is the association of a SNP in *UGT2B15* with the expression of *H2AFJ*. As part of the histone H2A superfamily *H2AFJ* may play a role in transcription regulation, DNA replication and repair and chromosome stability [45]. The amplification of *H2AFJ* in breast tumors has led to the suggestion that this gene is an oncogene [296]. We observed a relationship between the expression of *H2AFJ* and MD in both papers II and III, suggesting that the expression levels of this gene influence the levels of MD.

In paper III we found a SNP in *HSD3B1* associated with MD which is supportive of previous reports [70,71]. This is an interesting finding due to the involvement of the HSD3B family in the estrogen metabolism. Also, the fact that we managed to verify this association suggests that our study approach may provide enough power to detect causative alleles. In paper II we observed an association of increased *ESR1* expression with decreased MD - a finding inconsistent with previous reports and opposite of what one would expect. ERα which is encoded by *ESR1,* is able to set off signaling cascades involved in the growth of hormone responsive organs such as the breast and is therefore suggested to be involved in the development of increased MD. However,

reports of such an association have been inconclusive [297]. This raises the question as to what extent MD level is influenced by non-hormonal factors.

Steroid hormone and growth factor cross talk

As previously mentioned there is a crosstalk between members of the estrogen pathway and IGF pathway. The microenvironment is the primary compartment for exchange of growth related signaling. The developmental stages and growth of the mammary gland is directed by these signals which may be in the form of growth factors, hormones, cell-cell or cell-basement membrane interactions [15]. The metabolites and hormones are transported in the ground substance directed to or from target cells. The crosstalk between hormones and growth factors may be prominent for the achievement of signaling induced growth and differentiation of the epithelium comprising the ducts and alveoli.

In paper I we identified one haplotype in *IGF1* and three SNPs in *IGF1R* associated with MD. This suggests that, in addition to circulating levels, genetic variation in the IGF signaling pathway may also play a role in development of MD. In the normal breast *IGF1R* is found expressed in the cytoplasm of breast epithelial cells but not the BM, and increased expression confers an increased risk of BC [298]. Additionally, IGF1R is and anti-apoptotic agent found to be highly over expressed in malignant tumors, enhancing survival [299]. Recently it was discovered that IGF1 and IGF1R were involved in the early transformation of mammary cells [300,301]. In paper III we used the same sample material as in paper I, in addition to an independent verification study to analyze the estradiol pathway. In paper III we found a SNP in *AKT3*, a member of the AKT family, associated with the levels of MD. The AKT family is involved in cell proliferation, differentiation, apoptosis and tumorigenesis amongst others and upon activation by IGF1, AKT3 may play a role in cell survival [45]. Akt is the downstream mediator of PI3K, often referred to the PI3K/Akt signaling pathway due to the many extracellular signals they mediate, and is a pathway that is often dysregulated in human cancers [302].

In paper II we analyzed gene expression for association with MD. Amongst other findings we saw that the gene *PIK3R5* was differentially expressed according to MD. *PIK3R5* stands for PI3-Kinase regulatory subunit 5 and is thus a subunit of PI3K. The PI3Ks are divided into three classes; I, II, and III of which PIK3R5 is class I which is the only class involved in oncogenesis [303]. Among the 257 SNPs in paper III we analyzed one SNP residing in the gene PI3KC2B but

found no association. This gene however is part of class II [45] and may therefore not hold the same potential as the class I kinases or the SNP had no causative effect. However, mutations in PIK3CA (class II) are frequent in BC and are associated with ER gene expression, lymph node metastasis and poor survival. A SNP within the PIK3CA was found to be associated with breast cancer risk in two large studies but failed to be verified in a third, suggesting that this common variation may confer some risk but does not have a strong influence [304].

As previously mentioned the expression of *ESR1* was significantly down-regulated in women with high MD, a finding contrary to what one would expect. However, due to crosstalk between the growth factors and steroids, interaction between the signaling pathways may occur at several levels to increase proliferation. As previously suggested, in the loss of steroid receptors, one could hypothesize that increased cellular expression of other signaling molecules would provide sufficient signal through the growth factor stimulated pathways [305].

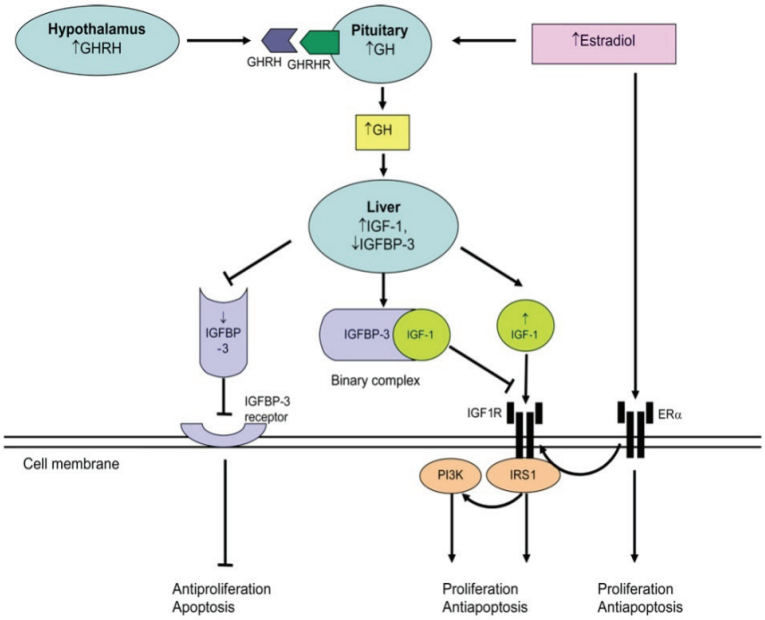


Figure 19: Crosstalk between estrogen and IGF pathway signaling. ERα: Estrogen receptor alpha, GH: Growth hormone, GHRH: Growth hormone releasing hormone, IGF1: Insulin-like growth factor 1, IGFBP3: Insulin-like growth factor binding protein 3, IGF1R: Insulin-like growth factor receptor 1, IRS1: Insulin receptor substrate 1, PI3K: phosphatidylinositol 3 kinase. From [306].

The propagation from dense tissue to BC might be induced by the microenvironment.

In paper III we found an association of a SNP in matrix metalloproteinase 2 (*MMP2*) with MD. This suggests that in addition to the crosstalk of hormones and growth factors, there might be an involvement of stromal matrix proteins in MD. MMP2 has the ability to degrade the extracellular matrix and is involved in tissue remodeling and metastasis [45]. Studies on the involvement of growth factors and stromal matrix proteins in MD have revealed that women with higher mammographic densities compared to their counterparts, had larger areas that were stained for IGF1 and the metalloproteinase TIMP3, the latter is an inhibitor of MMP2 [181]. Matrix metalloproteinases ensures proper branching morphogenesis by loosening the extracellular matrix (ECM) and clearing the path for the growing TDLU [26]. Women with higher stromal ECM density in the breast also have a stiffer matrix, due to greater deposition of the ECM [307]. As established in paper III involvement of the MMPs and SNPs in *MMPs* in cancer have been reported which makes them interesting candidates in studies on BC risk involving MD. The *MMP2* SNP would need to be analyzed in larger sample sets to conclude its involvement in density of the breast.

4.1.2 Integrated analyses and complex diseases

Our results suggest that it is imperative to not only test SNPs in coding regions and that collectively SNPs may together cause variability in phenotype. This was also seen in paper III upon performing multivariate analysis (Globaltest) of selected SNPs. It signifies the importance of interactions between loci. These interactions are postulated to hold the answer as to which genes underlie complex diseases, which has been suggested to be an average of 40 genes [308]. The hypothesis is that the function of a genetic factor is mediated through complex mechanisms involving other genes and also environmental factors. Accordingly, the effect of the genetic factor might be missed by isolating it from the system [309]. Several methods and software packages aim to reveal these interactions, and aim to either increase power to detect effect or detect statistical interactions between loci to explain the biology underpinning complex disease. With our analysis approach we aimed to achieve both in paper III in which we used candidate genes /SNPs, discovery-verification and inclusion expression data. In papers II and III we verified the association of *UGT* genes with MD on both the SNP level and the mRNA level, further strengthening the hypothesis that the integrated analysis may contribute to identification of associations.

4.2 Methodological considerations

4.2.1 Heterogeneity within and between studies.

Study heterogeneity is a problem because it causes variation among the samples which influences the statistical analysis. One might argue that the MDG study is heterogeneous because it contains samples from women who have attended follow-up screening due to suspicious findings. Hence the study may be enriched for women whose BC risk is not linked with MD. McCormack *et. al*. performed a meta analysis and reported that the association between breast density and breast cancer risk only holds when the control group represents an asymptomatic population, and that underlying breast disease might result in inconsistency in findings [175]. The inclusion of controls with suspicious findings in the MDG study raises the question as to whether or not these samples completely conform to an asymptomatic population. In addition, women in the MDG study who had previously had cancer were labeled as healthy in the clinical file. For biopsy of the breast this might be sufficient but for germline mutations it could potentially introduce noise. That is, if a woman who has a history of cancer has low MD, the main cause of BC can be assumed not to be linked with the MD level. However, of these women only one was included in the analyses and we conclude that the potential noise from this sample among the other 50 samples included from MDG is minimal and that our results are not affected substantially. TMBC on the other hand is a study of only healthy women, with no sign of disease. The combination of TMBC and MDG could therefore add noise to our analyses, resulting in no associations. However their differences could also be a strength. If we hypothesize that the samples in the MDG study lie closer to our trait of interest (BC) it could enforce our attempt of finding the specific SNPs with risk for increased MD which confer BC risk. On the other hand, all the women in the TMBC study have breast densities ranging from low to high density, and we cannot exclude the possibility that also some of these women may also have some irregularities. The difference between the studies is in the inclusion criteria, but in the end, the studies might not be too different. We know too little about BC progression to safely say that the samples are healthy just based on the mammographic image. As we have seen there may be many factors on the molecular level that come into play during tumor initiation and progression. A longitudinal study is needed to investigate all of these factors at the time of good health and after a woman has developed BC.

4.2.2 Study selection bias

Selection bias is an issue caused by the selection of subjects for the study and as a result the associations made in the study are dependent or biased by the participants. The women participating in the TMBC study were picked from the Norwegian breast cancer screening program (NBCSP), but only those who responded to the questionnaires could be included. Thus our samples could represent a selection bias. However, the TMBC study is a subset of a large nationwide screening program and there was a high attendance rate in both the NBCSP (78%) [310] and the TMBC. Of all participants attending the NBCSP during the recruitment period of TMBC, 80% were eligible to attend TMBC and of these, 70% participated [311,312]. In order for this to lead to selection bias, the associations of SNPs and gene expression with MD must be systematically different among those who did not participate and those who did. We find this unlikely. The women participating in the MDG study have purposely been selected due to suspicious findings and can thus introduce bias in the associations and also result in poor generalizability. The degree of difference between the TMBC and the MDG study is uncertain and hence also the effect of the bias that is potentially introduced by the MDG study.

4.2.3 Assessment of menopausal status

Assessment of the menopausal status was performed in the MDG material which was analyzed in article II and III. Due to the importance of estrogen in both MD and BC development this evaluation is critical due to differences in hormone levels in pre and postmenopausal women. To get the menopausal status as accurate as possible the circulating levels of follicle stimulating hormone (FSH), Lutenizing Hormone (LH), FSH/LH ratio and estrogen in addition to age, were used to determine a women as pre- or postmenopausal. The assessment was performed by clinicians with a good knowledge of the field. Thus we conclude that the categorization is an improvement to the self-reported menopause status. For the TMBC study women were classified as postmenopausal if they were 56 years or older reporting no menstruation during the last 12 months, or had a FSH level above 20 IU/L.

4.2.4 MD measurement

The measure of MD is influenced by several factors as it is measured in two dimensions and variability may arise from positioning, compression and x-ray exposure, in addition to the subjectivity of the radiographer [313]. Additionally the reader assessing the mammography images may also introduce bias. However, in both studies the reader was blinded to the characteristics of the women and thus the reader was not influenced by pre-conceived notions. In the TMBC study the intra-reader agreement rate was 0.86, and the inter-reader agreement was 0.86 (both Pearson correlation coefficient) and for the MDG study the test-retest reliability was 0.99 for absolute density. The MD readings were performed by the same person in both studies, or someone trained by her (GU), and therefore differences in the quantification should be minimal. Hence, the results should not be notably influenced by bias introduced at the MD assessment level. In addition, the scanners used for the TMBC and the MDG differ which might also have introduced noise. By comparison the cobrascan used for scanning the images in the TMBC study produces darker images. Hence the absolute density measurements from the two studies may not be completely comparable.

4.2.5 Blood vs Tissue

The use of blood to detect common genetic variation is advantageous in that it does not matter when the sample is taken and may therefore be used as a prognostic tool. In addition blood is easier and less invasive to sample than tissue. Tissue on the other hand is advantageous in that the area of interest is sampled and the local environment can be analyzed. Tissue samples can thus elaborate on the local involvement of signaling pathways or hormone receptor status affecting our phenotype. Such tools are already in use in the diagnosis and treatment of BC based on receptor status of ER and PR. Analyzing healthy tissue from women with varying MD, may give clues as to what the initiating steps of tumorigenesis are and be used for early detection of women at risk of BC. In paper I we used blood to study SNPs, while in paper II and III we used both blood and tissue to study SNPs and gene expression respectively.

The downside of using blood for SNP analyses is the potential unspecificity it holds and due to the vast amount of SNPs in the genome many markers must be tested to elucidate association important for our phenotype. As a result, a large sample size that can accommodate all these tests is necessary. Tissue may provide more specific clues as to what pathways/ genes are at play

locally, and may therefore be used with SNP analysis to direct the focus to the disease causing genes. We applied such an approach in paper II and III where we selected SNPs with a potential effect on MD through mutual selection of SNPs from eQTL analysis with gene expression, and association analysis of SNPs with MD.

4.2.6 SNP quality control

The SNP quality control for the SNPs genotyped in the candidate gene study was performed at a genotyping facility. In addition to this we chose to exclude SNPs with a lower population frequency than 5% which meant that 105 SNPs was excluded from the analysis. The initial candidate SNPs were chosen to have a frequency of >1%, hence the number of SNPs analyzed was greatly reduced compared to the number of SNPs genotyped. It is important to acknowledge that this is not due to poor genotyping quality but rather an artifact of SNP quality control.

4.2.7 SNP analysis

Low penetrance genetic variants.

Most study groups are now aiming at collecting large sample sets for the detection of SNPs with low frequency and low penetrance. There is reason to believe that most of the low hanging fruits have been discovered (eg BRCA1/2, TP53), thus, we are now in need of new approaches to discover the low frequent, moderate frequent and common low penetrance genetic variations. Three classes of breast cancer susceptibility alleles with different risk and prevalence have emerged; rare high-penetrance alleles, rare-moderate-penetrance alleles and common low-penetrance alleles [314]. Using a candidate gene approach in combination with htSNPs we aimed to identify members belonging to the latter class in paper III. The challenge in studying such SNPs lies in the power to detect the small effects they generate. A study of large sample size may provide substantial support for their detection but this is often not obtained. In paper II and III we aimed at performing a variety of analyses to best search our data for loci associated with the level of MD.

Using GWAS to detect common low penetrant SNPs

The whole genome is estimated to encompass more than 11 million SNPs. The current GWAS arrays only cover a small fraction of that. In order to detect the effect of common low penetrant SNPs with GWAS a large sample set is required. Unfortunately, as the number of tests increase,

so does the probability of committing a type 1 error, hence reporting false positives. To tackle this issue a multiple testing correction is often applied. Such correction may involve adjusting the p-value level at which one determines the test as statistically significant. In these cases Bonferroni is commonly used. With the use of Bonferroni the common threshold of 0.05, becomes 4,6 $x10^{-7}$ when a 109 K SNP array is analyzed and the 10^{-7} threshold is often referred to as the GWAS significance threshold. The Bonferroni method is known as a conservative method, hence in paper II and III we chose to use the FDR method described in section 2.8.8. Still, FDR only tells you about the fraction of falsely rejected H_0 within all rejected H_0, i.e. the proportion of false positives among the correct positives. In paper II we analyzed only selected SNPs from the 109K array and hence the multiple testing problem is less prominent. Since we were searching for common low penetrant genetic variations we chose additional integrated analyses approaches, hoping that they in combination could strengthen our findings.

The candidate gene approach

The candidate gene approach is beneficial for studies where a known set of genes, genetic variants or as shown in paper III, a pathway of interest is tested. Compared with a genome wide approach a candidate gene approach may have more power to detect relatively small effects in addition to being more cost efficient and also provide flexibility in the choice of markers to be tested. The downside is that a candidate gene approach is heavily reliant on the *a priori* knowledge of the phenotype of interest and biological intuition. And although the right gene is selected it is still a question of testing the right SNP. The probability of selecting both the right gene and the right SNP is potentially very small [315]. To limit this problem we used the benefits of LD in selecting SNPs which not only represented itself but "tagged" other SNPs in the vicinity. At the most 65 SNPs were tested by one SNP. Additionally, the genes selected to be part of the estrogen pathway were many and thus included genes that were not solely in the estradiol pathway but branched into other pathways of potential interest. Another limitation prone in a candidate gene study is the software used. The technology is advancing fast and the tools are much better now than some years ago when these genes and SNPs were selected. However, there will always be a limitation to what information these databases hold because they are heavily reliant on being updated with current findings and on how the information is extracted and interpreted. In our candidate gene and SNP selection process we utilized several commercial and

free databases with the aim of detecting all known candidates possible at that time. For comparison, a GWAS approach is less affected and biased by such trends.

HapMap

The populations within the HapMap project are assumed to be representative of their respective ethnic groups, and were used in our selection of SNPs in paper III. However, we cannot underestimate the power of testing the causative allele. Only by sequencing the genome of all samples in a selected study population would it be possible to obtain true estimates of SNPs in LD and htSNPs for that exact population, but then the htSNPs would be redundant. Also, using the common LD threshold of 0.8, does not ensure LD across all samples in the study, thus the htSNP approach misses some samples and by that it is inaccurate. This approach is widely used and is currently one of the best options available. The issue of both power and transferability of htSNPs has been addressed by de Bakker [269,316], who concluded that the power was not compromised and that htSNPs could be used in populations with great diversity [316].

Discovery verification approach

The problems we face in GWA studies regarding sample size and multiple testing corrections also apply for small studies such as the candidate gene study performed in paper III. However, when looking for common low penetrant SNPs in smaller sample sets, the use of conventional multiple testing procedures is discouraging. We acknowledge that the sample sets analyzed are small, thus providing limited power to detect causative alleles. Thus in paper III we opted for a discovery/verification approach to test the validity of our results through verification. We used a p-value threshold of 0.1 in the discovery phase aiming at including potential causative SNPs while eliminating MD irrelevant SNPs from further testing. In addition this reduced the multiple testing problem somewhat. In the subsequent verification step we used the conventional p-value threshold 0.05, although several of the associations had p-values >0.01. Hence, in the discovery phase we chose to use an unconventional liberal p-value. However, the use of the p-value 0.05 is somewhat arbitrary and does not take the power of our experiment into account. We argue that by using the conventional p-value 0.05 in a small study with limited or unknown power it is easy to conclude that there are no significant results and therefore keep the H_0. This way of thinking is only applicable in experiments that are powerful enough and when the possibility of committing a type II error (retaining a wrong H_0) is unlikely [317]. The strength in our study lies in our

samples. The MDG with the biopsies from healthy women provide a rare opportunity to investigate the local environment in the breasts of women with a range of MD and holds great potential for shedding light on MD as an intermediated step to BC. TMBC is valuable in that it is a subset of a population screening and therefore represents the general population. However, our sample size is still small and to further empower our results we added gene expression data to our SNP data to facilitate an integrated analysis approach.

4.2.8 Integrated analysis

In paper II and III we use both expression and SNP data to elucidate the phenotype. As more markers are being tested more samples are needed. An integrated analysis approach may be the answer to this growing problem, the hypothesis being that one association is enforced or modified by another association at another molecular level. In the case of our SNP and expression analysis the theory is that variations in SNPs are mediated through gene expression, causing variation in our phenotype, MD. In paper II and III we analyzed the SNP profiles with the gene expression levels by performing eQTL analysis and separately these two were tested for association with the phenotype. Hence we tested the hypothesis that analysis on different genomic levels should point in the same direction when the same phenotype is analyzed. The current method involves a three-step analysis approach and the downside to this method is that even though the three results point towards the same association it might not be real. The abundance of SNPs and gene transcripts tested means that by chance we could pick up coherent associations. To deal with this issue we are currently developing a method that includes all variables in one analysis with the aim of excluding associations without relevance to the hypothesis.

5 Main conclusion and future perspectives

With a variety of analysis approaches we conclude that there is support for a role of genetic variation in the steroid hormone pathway and the IGF-pathway in MD and that these two combined may create steroid growth-factor crosstalk. The addition of associations with microenvironment specific factors supports the hypothesis that the environment in the breast could contribute to MD and ultimately BC. Increased research in the field of low-penetrant genetic variants is needed to shed light on complex diseases such as breast cancer. This work has already begun and as more groups collaborate to increase power to detect associations one could expect to see great advances in molecular medicine in the near future.

5.1 Integrated analyses

The future holds great promise for advances in systems biology with integrated analyses becoming a major part. As a result of the great advances being made in the field of technology, increasing information on the DNA, mRNA and protein level is being obtained. There are many modifying steps in going from DNA to mRNA and from mRNA to protein, hence an association seen on one level might be either silenced or amplified in the next. Therefore, the search for association of genetic variability with a given phenotype is likely to require information from each level. Additionally, interactions within a level are also of importance since they may reveal novel crosstalk signaling. Currently most studies are performed on either a single level or possibly two. The challenge and rate limiting step being the lack of methods and computer power needed to analyze all the levels at once in a systematic way in order to reveal biological interaction. The advantages of integrated analyses are many, the most prominent being the application in treatment of disease and personalized medicine.

5.2 Translation into the clinic

Currently the clinic use mammographic images solely for tumor detection. The vast amounts of information that lie within the images are thus not used unless they are used for research purposes such as the ones presented in this thesis. However knowledge concerning MD is growing fast due to its strong association with BC risk. The potential use of mammographic density in a clinical setting is not straight forward due to the complexity behind the association of MD with BC. Not everyone who has high mammographic density will develop BC suggesting that there are more factors to be considered. Genetic markers are important in this setting. Once the genetic markers of increased MD which may influence BC risk are established, and there are most probably many, they may be coupled with MD for determination of BC risk. The combination of such information may also be used to advise women with increased BC risk to take protective measures. The addition of such information into an already well-functioning screening program would be very beneficial. MD has been suggested added to the Gail model which is used to predict BC risk upon increasing prediction accuracy from 0.607 to 0.642 [318]. Interestingly, with the addition of seven SNPs associated with BC the concordance statistic increased to 0.632 [319].

5.3 Over diagnosis

With the introduction of mammographic screening programs there is an increased chance of detecting tumors that would have otherwise gone unnoticed due to lack of symptoms or health issues conferring a risk of a woman's life. Tumors that have previously been too small to be detected are being detected with x-ray technology. Some of these tumors are slow growing and would never be detected or cause disease. However, clinicians are unable to differentiate the slow-growing from the aggressive and thus treat all. Not only is this a great cost financially, it is a great personal cost for the woman diagnosed and treated for BC when it is not needed. The addition of genetic tests in the clinic that could differentiate the aggressive cases from the slow growing and less aggressive counterparts would greatly benefit the patient. Also, the improvement of screening devices could aid in a more precise diagnosis. The future holds great promise for the incorporation of molecular markers in screening and in treatment.

Abbreviations

ABDEN: Absolute density
ANOVA: analysis of variance
BC: Breast Cancer
BIRADS: Breast Imaging Reporting and Data System
BM: basement membrane
BMI: Body Mass Index
CC: craniocaudal
DHAE(S): Dehydroepiandrosterone (sulfate)
DNA: Deoxyribonucleic acid
E1: Estrone
E2: Estradiol
E3: Estriol
ECM: extracellular matrix
ELISA: Enzyme-Linked Immunosorbent Assay
EMT: epithelial-mesenchymal transition
eQTL: expression quantitative trait loci
ER: Estrogen receptor
FDR: False Discover Rate
FFTP: First full term pregnancy
FSH: Follicle Stimulating Hormone
FWER: family wise error rate
GH: Growth hormone
GWAS: genome wide association studies
HT: postmenopausal hormone therapy
HWE: Hardy Weinberg Equilibrium
IARC: International agency for research on cancer
IGF1: Insuline-Like Growth Factor 1
IGFBP3: Insulin-Like Growth Factor Binding Protein 3
LD: Linkage Disequilibrium
LH: Lutenizing Hormone
MD: Mammographic density
MDG: Mammographic Density and Genetics (study)
MLO: medio-lateral oblique
MMP: matrix metalloproteinase
NBCSP: Norwegian breast cancer screening program
PDEN: Percent density
PR: Progesterone receptor
ROI: region of interest
ROS: Reactive oxygen species
SAM: significant analysis of microarray
SHBG: Sex Hormone Binding Globulin
SNP: Single nucleotide polymorphism
TDLU: terminal duct lobular unit
TEB: Terminal end bud
TMBC: Tromsø Mammographic and Breast Cancer (study)

Reference List

1. Anderson E, Clarke RB, Howell A: **Estrogen responsiveness and control of normal human breast proliferation.** *J Mammary Gland Biol Neoplasia* 1998, **3:** 23-35.

2. Dawson EK: **A histological study of the normal mamma in realtion to tumor growth. I. Early development to maturity.** *Edinburgh Madical Journal* 1934, **41:** 653.

3. Sternlicht MD, Kouros-Mehr H, Lu P, Werb Z: **Hormonal and local control of mammary branching morphogenesis.** *Differentiation* 2006, **74:** 365-381.

4. Porter JC: **Proceedings: Hormonal regulation of breast development and activity.** *J Invest Dermatol* 1974, **63:** 85-92.

5. Cardiff RD: **Are the TDLU of the human the same as the LA of mice?** *J Mammary Gland Biol Neoplasia* 1998, **3:** 3-5.

6. Cardiff RD, Wellings SR: **The comparative pathology of human and mouse mammary glands.** *J Mammary Gland Biol Neoplasia* 1999, **4:** 105-122.

7. Russo J, Rivera R, Russo IH: **Influence of age and parity on the development of the human breast.** *Breast Cancer Res Treat* 1992, **23:** 211-218.

8. Ovalle WK, Nahirney PC: *Netter's Essential Histology*, 1st edn. Saunders Elsevier; 2008.

9. Harris Jay R., Lippman Marc E., Morrow Monica, Osborne C.Kent: *Diseases of the breast*, 2nd edn. Lippincott Williams and Wilkins; 2000.

10. Russo J, Russo: **Development of the human mammary gland.** In *The Mammary Gland: development, regulation, and function*. Edited by Margaret C Neville, Charles W Daniels. New York: Plenum Press; 1987:67-93.

11. Deugnier MA, Teuliere J, Faraldo MM, Thiery JP, Glukhova MA: **The importance of being a myoepithelial cell.** *Breast Cancer Res* 2002, **4:** 224-230.

12. Lakhani SR, O'Hare MJ: **The mammary myoepithelial cell--Cinderella or ugly sister?** *Breast Cancer Res* 2001, **3:** 1-4.

13. Polyak K, Kalluri R: **The role of the microenvironment in mammary gland development and cancer.** *Cold Spring Harb Perspect Biol* 2010, **2:** a003244.

14. Wheater Paul R, Burkitt HG, Daniels VG: **Female reproductive system.** In *Funtional Histology: a text and colour atlas*. New York: Churchill Livingstone; 1979.

15. McCave EJ, Cass CA, Burg KJ, Booth BW: **The normal microenvironment directs mammary gland development.** *J Mammary Gland Biol Neoplasia* 2010, **15:** 291-299.

16. Dulbecco R, Henahan M, Armstrong B: **Cell types and morphogenesis in the mammary gland.** *Proc Natl Acad Sci U S A* 1982, **79:** 7346-7350.

17. Rudland PS: **Histochemical organization and cellular composition of ductal buds in developing human breast: evidence of cytochemical intermediates between epithelial and myoepithelial cells.** *J Histochem Cytochem* 1991, **39:** 1471-1484.

18. Villadsen R, Fridriksdottir AJ, Ronnov-Jessen L, Gudjonsson T, Rank F, LaBarge MA *et al*.: **Evidence for a stem cell hierarchy in the adult human breast.** *J Cell Biol* 2007, **177:** 87-101.

19. Smalley M, Ashworth A: **Stem cells and breast cancer: A field in transit.** *Nat Rev Cancer* 2003, **3:** 832-844.

20. Visvader JE: **Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis.** *Genes Dev* 2009, **23:** 2563-2577.

21. Stingl J, Eaves CJ, Zandieh I, Emerman JT: **Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue.** *Breast Cancer Res Treat* 2001, **67:** 93-109.

22. Williams JM, Daniel CW: **Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis.** *Dev Biol* 1983, **97:** 274-290.

23. Humphreys RC: **Programmed cell death in the terminal endbud.** *J Mammary Gland Biol Neoplasia* 1999, **4:** 213-220.

24. Polyak K, Hu M: **Do myoepithelial cells hold the key for breast tumor progression?** *J Mammary Gland Biol Neoplasia* 2005, **10:** 231-247.

25. Narod SA: **Early-onset breast cancer: what do we know about the risk factors?: A Countercurrents Series.** *Curr Oncol* 2011, **18:** 204-205.

26. Sternlicht MD: **Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis.** *Breast Cancer Res* 2006, **8:** 201.

27. Gonzalez-Angulo AM, Hennessy BT, Mills GB: **Future of personalized medicine in oncology: a systems biology approach.** *J Clin Oncol* 2010, **28:** 2777-2783.

28. Arendt LM, Rudnick JA, Keller PJ, Kuperwasser C: **Stroma in breast development and disease.** *Semin Cell Dev Biol* 2010, **21:** 11-18.

29. Vuorio E, de CB: **The family of collagen genes.** *Annu Rev Biochem* 1990, **59:837-72.:** 837-872.

30. LaMarca HL, Rosen JM: **Minireview: hormones and mammary cell fate--what will I become when I grow up?** *Endocrinology* 2008, **149:** 4317-4321.

31. McManus MJ, Welsch CW: **The effect of estrogen, progesterone, thyroxine, and human placental lactogen on DNA synthesis of human breast ductal epithelium maintained in athymic nude mice.** *Cancer* 1984, **54:** 1920-1927.

32. Hughes ES: **The Development of the Mammary Gland: Arris and Gale Lecture, delivered at the Royal College of Surgeons of England on 25th October, 1949.** *Ann R Coll Surg Engl* 1950, **6:** 99-119.

33. Russo J, Russo IH: **Development of the human breast.** *Maturitas* 2004, **49:** 2-15.

34. Hennighausen L, Robinson GW: **Information networks in the mammary gland.** *Nat Rev Mol Cell Biol* 2005, **6:** 715-725.

35. Hennighausen L, Robinson GW: **Signaling pathways in mammary gland development.** *Dev Cell* 2001, **1:** 467-475.

36. Bocchinfuso WP, Lindzey JK, Hewitt SC, Clark JA, Myers PH, Cooper R *et al.*: **Induction of mammary gland development in estrogen receptor-alpha knockout mice.** *Endocrinology* 2000, **141:** 2982-2994.

37. Neville MC: **Anatomy and physiology of lactation.** *Pediatr Clin North Am* 2001, **48:** 13-34.

38. Russo IH, Russo J: **Role of hormones in mammary cancer initiation and progression.** *J Mammary Gland Biol Neoplasia* 1998, **3:** 49-61.

39. Bartow SA: **Use of the autopsy to study ontogeny and expression of the estrogen receptor gene in human breast.** *J Mammary Gland Biol Neoplasia* 1998, **3:** 37-48.

40. Marti A, Lazar H, Ritter P, Jaggi R: **Transcription factor activities and gene expression during mouse mammary gland involution.** *J Mammary Gland Biol Neoplasia* 1999, **4:** 145-152.

41. Howard BA, Gusterson BA: **Human breast development.** *J Mammary Gland Biol Neoplasia* 2000, **5:** 119-137.

42. Edwards RG, Howles CM, Macnamee C: **Clinical endocrinology of reproduction.** In *Hormones: from molecules to disease*. Edited by Baulieu EE, Kelly PA. New York: Chapman and Hall; 1990:457.

43. Russo J, Balogh GA, Chen J, Fernandez SV, Fernbaugh R, Heulings R *et al.*: **The concept of stem cell in the mammary gland and its implication in morphogenesis, cancer and prevention.** *Front Biosci* 2006, **11:151-72.:** 151-172.

44. Greenspan FS, Gardner DG: *Basic and clinical endocrinology*, 7th edn. Lange Medical Books/McGraw-Hill; 2004.

45. GeneCards. http://www.genecards.org . 2008.

46. Estrogen Receptor Structures & Functions. Biological Sciences Department, Carnegie Mellon University.Pittsburgh, PA . 2006. 13-11-2006.

47. Koehler KF, Helguero LA, Haldosen LA, Warner M, Gustafsson JA: **Reflections on the discovery and significance of estrogen receptor beta.** *Endocr Rev* 2005, **26:** 465-478.

48. Mulac-Jericevic B, Conneely OM: **Reproductive tissue selective actions of progesterone receptors.** *Reproduction* 2004, **128:** 139-146.

49. Martin MB, Stoica A: **Insulin-like growth factor-I and estrogen interactions in breast cancer.** *J Nutr* 2002, **132:** 3799S-3801S.

50. Adesanya OO, Zhou J, Samathanam C, Powell-Braxton L, Bondy CA: **Insulin-like growth factor 1 is required for G2 progression in the estradiol-induced mitotic cycle.** *Proc Natl Acad Sci U S A* 1999, **96:** 3287-3291.

51. Boisclair YR, Rhoads RP, Ueki I, Wang J, Ooi GT: **The acid-labile subunit (ALS) of the 150 kDa IGF-binding protein complex: an important but forgotten component of the circulating IGF system.** *J Endocrinol* 2001, **170:** 63-70.

52. Ruan W, Catanese V, Wieczorek R, Feldman M, Kleinberg DL: **Estradiol enhances the stimulatory effect of insulin-like growth factor-I (IGF-I) on mammary development and growth hormone-induced IGF-I messenger ribonucleic acid.** *Endocrinology* 1995, **136:** 1296-1302.

53. Ruan W, Monaco ME, Kleinberg DL: **Progesterone stimulates mammary gland ductal morphogenesis by synergizing with and enhancing insulin-like growth factor-I action.** *Endocrinology* 2005, **146:** 1170-1178.

54. Kleinberg DL, Barcellos-Hoff MH: **The pivotal role of insulin-like growth factor I in normal mammary development.** *Endocrinol Metab Clin North Am* 2011, **40:** 461-71, vii.

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

56. Yee D: **The insulin-like growth factor system as a target in breast cancer.** *Breast Cancer Res Treat* 1994, **32:** 85-95.

57. Lee AV, Weng CN, Jackson JG, Yee D: **Activation of estrogen receptor-mediated gene transcription by IGF-I in human breast cancer cells.** *J Endocrinol* 1997, **152:** 39-47.

58. Allen NE, Roddam AW, Allen DS, Fentiman IS, dos SS, I, Peto J *et al*.: **A prospective study of serum insulin-like growth factor-I (IGF-I), IGF-II, IGF-binding protein-3 and breast cancer risk.** *Br J Cancer* 2005, **92:** 1283-1287.

59. Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J: **An estimate of unique DNA sequence heterozygosity in the human genome.** *Hum Genet* 1985, **69:** 201-205.

60. Syvanen AC: **Accessing genetic variation: genotyping single nucleotide polymorphisms.** *Nat Rev Genet* 2001, **2:** 930-942.

61. Collins FS, Guyer MS, Charkravarti A: **Variations on a theme: cataloging human DNA sequence variation.** *Science* 1997, **278:** 1580-1581.

62. Zhang K, Calabrese P, Nordborg M, Sun F: **Haplotype block structure and its applications to association studies: power and study designs.** *Am J Hum Genet* 2002, **71:** 1386-1394.

63. Slatkin M: **Linkage disequilibrium--understanding the evolutionary past and mapping the medical future.** *Nat Rev Genet* 2008, **9:** 477-485.

64. Hapmap. http://www.hapmap.org/index.html.en . 2006.

65. **The International HapMap Project.** *Nature* 2003, **426:** 789-796.

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

67. de Moura Ramos EH, Martinelli S, Silva I, Nazario A, Facina G, Costa A *et al.*: **Association between estrogen receptor gene polymorphisms and breast density in postmenopausal women.** *Climacteric* 2009, **12:** 490-501.

68. Mackenzie PI, Gregory PA, Lewinsky RH, Yasmin SN, Height T, McKinnon RA *et al.*: **Polymorphic variations in the expression of the chemical detoxifying UDP glucuronosyltransferases.** *Toxicol Appl Pharmacol* 2005, **207:** 77-83.

69. Yong M, Schwartz SM, Atkinson C, Makar KW, Thomas SS, Newton KM *et al.*: **Associations between polymorphisms in glucuronidation and sulfation enzymes and mammographic breast density in premenopausal women in the United States.** *Cancer Epidemiol Biomarkers Prev* 2010, **19:** 537-546.

70. Haiman CA, Bernstein L, Berg D, Ingles SA, Salane M, Ursin G: **Genetic determinants of mammographic density.** *Breast Cancer Res* 2002, **4:** R5.

71. Stone J, Gurrin LC, Byrnes GB, Schroen CJ, Treloar SA, Padilla EJ *et al.*: **Mammographic density and candidate gene variants: a twins and sisters study.** *Cancer Epidemiol Biomarkers Prev* 2007, **16:** 1479-1484.

72. Howell A, Sims AH, Ong KR, Harvie MN, Evans DG, Clarke RB: **Mechanisms of Disease: prediction and prevention of breast cancer--cellular and molecular interactions.** *Nat Clin Pract Oncol* 2005, **2:** 635-646.

73. Dvorak HF: **Tumors: wounds that do not heal.** *The new england journal of medicine* 1986.

74. Beatson GT: **On the treatment of inoperable cases of carcinoma of the mamma: Suggestions for a new method of treatment, with illustrative cases.** *Lancet* 1986, **2:** 104-107.

75. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, **100:** 57-70.

76. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No.10. [Internet] . 2010. Lyon, France: International Agency for Research on Cancer. 20-9-2011.

77. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics.** *CA Cancer J Clin* 2011, **61:** 69-90.

78. Cancer registry of Norway. Cancer in Norway 2008 - Cancer incidence, mortality, survival and prevalence in Norway. 2009. Oslo: Cancer registry of Norway.

79. Jemal A, Center MM, DeSantis C, Ward EM: **Global patterns of cancer incidence and mortality rates and trends.** *Cancer Epidemiol Biomarkers Prev* 2010, **19:** 1893-1907.

80. Kaplan SS: **Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue.** *Radiology* 2001, **221:** 641-649.

81. Kelsey JL: **A review of the epidemiology of human breast cancer.** *Epidemiol Rev* 1979, **1:74-109.:** 74-109.

82. Hulka BS, Moorman PG: **Breast cancer: hormones and other risk factors.** *Maturitas* 2001, **38:** 103-113.

83. Wolfe JN: **Risk for breast cancer development determined by mammographic parenchymal pattern.** *Cancer* 1976, **37:** 2486-2492.

84. Boyd NF, O'Sullivan B, Campbell JE, Fishell E, Simor I, Cooke G *et al.*: **Mammographic signs as risk factors for breast cancer.** *Br J Cancer* 1982, **45:** 185-193.

85. Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB *et al.*: **Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study.** *J Natl Cancer Inst* 1995, **87:** 670-675.

86. Ma H, Luo J, Press MF, Wang Y, Bernstein L, Ursin G: **Is there a difference in the association between percent mammographic density and subtypes of breast cancer? Luminal A and triple-negative breast cancer.** *Cancer Epidemiol Biomarkers Prev* 2009, **18:** 479-485.

87. Tamimi RM, Byrne C, Colditz GA, Hankinson SE: **Endogenous hormone levels, mammographic density, and subsequent risk of breast cancer in postmenopausal women.** *J Natl Cancer Inst* 2007, **99:** 1178-1187.

88. Weinberg RA: **How cancer arises.** *Sci Am* 1996, **275:** 62-70.

89. Anderson E, REED SC, HUSEBY RA, OLIVER CP: **Possible relationship between menopause and age at onset of breast cancer.** *Cancer* 1950, **3:** 410-414.

90. LILIENFELD AM, JOHNSON EA: **The age distribution in female breast and genital cancers.** *Cancer* 1955, **8:** 875-882.

91. Pike MC, Krailo MD, Henderson BE, Casagrande JT, Hoel DG: **"Hormonal" risk factors, "breast tissue age" and the age-incidence of breast cancer.** *Nature* 1983, **303:** 767-770.

92. Opdahl S, Alsaker MD, Janszky I, Romundstad PR, Vatten LJ: **Joint effects of nulliparity and other breast cancer risk factors.** *Br J Cancer* 2011, **105:** 731-736.

93. MacMahon B, Cole P, Lin TM, Lowe CR, Mirra AP, Ravnihar B *et al*.: **Age at first birth and breast cancer risk.** *Bull World Health Organ* 1970, **43:** 209-221.

94. WYNDER EL, BROSS IJ, HIRAYAMA T: **A study of the epidemiology of cancer of the breast.** *Cancer* 1960, **13:559-601.:** 559-601.

95. Li CI, Malone KE, Daling JR, Potter JD, Bernstein L, Marchbanks PA *et al*.: **Timing of menarche and first full-term birth in relation to breast cancer risk.** *Am J Epidemiol* 2008, **167:** 230-239.

96. Lambe M, Hsieh CC, Chan HW, Ekbom A, Trichopoulos D, Adami HO: **Parity, age at first and last birth, and risk of breast cancer: a population-based study in Sweden.** *Breast Cancer Res Treat* 1996, **38:** 305-311.

97. Newcomb PA, Storer BE, Longnecker MP, Mittendorf R, Greenberg ER, Clapp RW *et al*.: **Lactation and a reduced risk of premenopausal breast cancer.** *N Engl J Med* 1994, **330:** 81-87.

98. Key TJ, Pike MC: **The role of oestrogens and progestagens in the epidemiology and prevention of breast cancer.** *Eur J Cancer Clin Oncol* 1988, **24:** 29-43.

99. Schedin P: **Pregnancy-associated breast cancer and metastasis.** *Nat Rev Cancer* 2006, **6:** 281-291.

100. Trichopoulos D, Hsieh CC, MacMahon B, Lin TM, Lowe CR, Mirra AP *et al*.: **Age at any birth and breast cancer risk.** *Int J Cancer* 1983, **31:** 701-704.

101. Hsieh C, Pavia M, Lambe M, Lan SJ, Colditz GA, Ekbom A *et al*.: **Dual effect of parity on breast cancer risk.** *Eur J Cancer* 1994, **30A:** 969-973.

102. Albrektsen G, Heuch I, Tretli S, Kvale G: **Breast cancer incidence before age 55 in relation to parity and age at first and last births: a prospective study of one million Norwegian women.** *Epidemiology* 1994, **5:** 604-611.

103. Press DJ, Pharoah P: **Risk factors for breast cancer: a reanalysis of two case-control studies from 1926 and 1931.** *Epidemiology* 2010, **21:** 566-572.

104. Abrams TJ, Guzman RC, Swanson SM, Thordarson G, Talamantes F, Nandi S: **Changes in the parous rat mammary gland environment are involved in parity-associated protection against mammary carcinogenesis.** *Anticancer Res* 1998, **18:** 4115-4121.

105. Russo J, Tay LK, Russo IH: **Differentiation of the mammary gland and susceptibility to carcinogenesis.** *Breast Cancer Res Treat* 1982, **2:** 5-73.

106. Britt K, Ashworth A, Smalley M: **Pregnancy and the risk of breast cancer.** *Endocr Relat Cancer* 2007, **14:** 907-933.

107. Vihko R, Apter D: **Endocrine characteristics of adolescent menstrual cycles: impact of early menarche.** *J Steroid Biochem* 1984, **20:** 231-236.

108. MacMahon B, Trichopoulos D, Brown J, Andersen AP, Cole P, deWaard F *et al*.: **Age at menarche, urine estrogens and breast cancer risk.** *Int J Cancer* 1982, **30:** 427-431.

109. Kelsey JL, Gammon MD, John EM: **Reproductive factors and breast cancer.** *Epidemiol Rev* 1993, **15:** 36-47.

110. **Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer.** *Lancet* 1997, **350:** 1047-1059.

111. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML *et al*.: **Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial.** *JAMA* 2002, **288:** 321-333.

112. ***IARC Monographs on the Evaluation of Carcinogenic Risks to Humans.Combined Estrogen-Progestogen Contraceptives and Combined Estrogen-Progestogen Menopausal Therapy.*** Lyon, France: International Agency for Research on Cancer; 2007.

113. **Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Collaborative Group on Hormonal Factors in Breast Cancer.** *Lancet* 1996, **347:** 1713-1727.

114. Key T, Appleby P, Barnes I, Reeves G: **Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies.** *J Natl Cancer Inst* 2002, **94:** 606-616.

115. Ross RK, Paganini-Hill A, Wan PC, Pike MC: **Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin.** *J Natl Cancer Inst* 2000, **92:** 328-332.

116. Chlebowski RT, Anderson GL, Gass M, Lane DS, Aragaki AK, Kuller LH *et al*.: **Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women.** *JAMA* 2010, **%20;304:** 1684-1692.

117. Beral V: **Breast cancer and hormone-replacement therapy in the Million Women Study.** *Lancet* 2003, **362:** 419-427.

118. Rosenberg LU, Granath F, Dickman PW, Einarsdottir K, Wedren S, Persson I *et al*.: **Menopausal hormone therapy in relation to breast cancer characteristics and prognosis: a cohort study.** *Breast Cancer Res* 2008, **10:** R78.

119. Tretli S: **Height and weight in relation to breast cancer morbidity and mortality. A prospective study of 570,000 women in Norway.** *Int J Cancer* 1989, **44:** 23-30.

120. Stephenson GD, Rose DP: **Breast cancer and obesity: an update.** *Nutr Cancer* 2003, **45:** 1-16.

121. Brody JG, Rudel RA, Michels KB, Moysich KB, Bernstein L, Attfield KR *et al*.: **Environmental pollutants, diet, physical activity, body size, and breast cancer: where do we stand in research to identify opportunities for prevention?** *Cancer* 2007, **109:** 2627-2634.

122. Colditz GA, Hankinson SE: **The Nurses' Health Study: lifestyle and health among women.** *Nat Rev Cancer* 2005, **5:** 388-396.

123. Brennan SF, Cantwell MM, Cardwell CR, Velentzis LS, Woodside JV: **Dietary patterns and breast cancer risk: a systematic review and meta-analysis.** *Am J Clin Nutr* 2010, **91:** 1294-1302.

124. Smith-Warner SA, Spiegelman D, Yaun SS, van den Brandt PA, Folsom AR, Goldbohm RA *et al*.: **Alcohol and breast cancer in women: a pooled analysis of cohort studies.** *JAMA* 1998, **279:** 535-540.

125. Michels KB, Mohllajee AP, Roset-Bahmanyar E, Beehler GP, Moysich KB: **Diet and breast cancer: a review of the prospective observational studies.** *Cancer* 2007, **109:** 2712-2749.

126. Kehrer JP: **Free radicals as mediators of tissue injury and disease.** *Crit Rev Toxicol* 1993, **23:** 21-48.

127. Gate L, Paul J, Ba GN, Tew KD, Tapiero H: **Oxidative stress induced in pathologies: the role of antioxidants.** *Biomed Pharmacother* 1999, **53:** 169-180.

128. Kang DH: **Oxidative stress, DNA damage, and breast cancer.** *AACN Clin Issues* 2002, **13:** 540-549.

129. Borek C: **Dietary antioxidants and human cancer.** *Integr Cancer Ther* 2004, **3:** 333-341.

130. Pan SY, Zhou J, Gibbons L, Morrison H, Wen SW: **Antioxidants and breast cancer risk- a population-based case-control study in Canada.** *BMC Cancer* 2011, **11:** 372.

131. Seibold P, Hein R, Schmezer P, Hall P, Liu J, Dahmen N *et al*.: **Polymorphisms in oxidative stress-related genes and postmenopausal breast cancer risk.** *Int J Cancer* 2011, **129:** 1467-1476.

132. Hankinson SE, Colditz GA, Willett WC: **Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones.** *Breast Cancer Res* 2004, **6:** 213-218.

133. **Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease.** *Lancet* 2001, **358:** 1389-1399.

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

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

136. FitzGerald MG, Marsh DJ, Wahrer D, Bell D, Caron S, Shannon KE *et al*.: **Germline mutations in PTEN are an infrequent cause of genetic predisposition to breast cancer.** *Oncogene* 1998, **17:** 727-731.

137. Levine AJ: **p53, the cellular gatekeeper for growth and division.** *Cell* 1997, **88:** 323-331.

138. Thompson D, Easton D: **The genetic epidemiology of breast cancer genes.** *J Mammary Gland Biol Neoplasia* 2004, **9:** 221-236.

139. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG *et al*.: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, **447:** 1087-1093.

140. Mitrunen K, Hirvonen A: **Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism.** *Mutat Res* 2003, **544:** 9-41.

141. **CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies.** *Am J Hum Genet* 2004, **74:** 1175-1182.

142. Nexo BA, Vogel U, Olsen A, Ketelsen T, Bukowy Z, Thomsen BL *et al*.: **A specific haplotype of single nucleotide polymorphisms on chromosome 19q13.2-3 encompassing the gene RAI is indicative of post-menopausal breast cancer before age 55.** *Carcinogenesis* 2003, **24:** 899-904.

143. Kammerer S, Roth RB, Reneland R, Marnellos G, Hoyal CR, Markward NJ *et al*.: **Large-scale association study identifies ICAM gene region as breast and prostate cancer susceptibility locus.** *Cancer Res* 2004, **64:** 8906-8910.

144. Benes P, Jurajda M, Zaloudik J, Izakovicova-Holla L, Vacha J: **C766T low-density lipoprotein receptor-related protein 1 (LRP1) gene polymorphism and susceptibility to breast cancer.** *Breast Cancer Res* 2003, **5:** R77-R81.

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

146. Shen J, Gammon MD, Terry MB, Wang L, Wang Q, Zhang F *et al*.: **Polymorphisms in XRCC1 modify the association between polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, dietary antioxidants, and breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 2005, **14:** 336-342.

147. Ziv E, Cauley J, Morin PA, Saiz R, Browner WS: **Association between the T29-->C polymorphism in the transforming growth factor beta1 gene and breast cancer among elderly white women: The Study of Osteoporotic Fractures.** *JAMA* 2001, **285:** 2859-2863.

148. Hoyal CR, Kammerer S, Roth RB, Reneland R, Marnellos G, Kiechle M *et al*.: **Genetic polymorphisms in DPF3 associated with risk of breast cancer and lymph node metastases.** *J Carcinog* 2005, **%19;4:13.:** 13.

149. Le ML, Donlon T, Kolonel LN, Henderson BE, Wilkens LR: **Estrogen metabolism-related genes and breast cancer risk: the multiethnic cohort study.** *Cancer Epidemiol Biomarkers Prev* 2005, **14:** 1998-2003.

150. Cai Q, Kataoka N, Li C, Wen W, Smith JR, Gao YT *et al*.: **Haplotype Analyses of CYP19A1 Gene Variants and Breast Cancer Risk: Results from the Shanghai Breast Cancer Study.** *Cancer Epidemiol Biomarkers Prev* 2008, **17:** 27-32.

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

152. Henningson M, Hietala M, Torngren T, Olsson H, Jernstrom H: **IGF1 htSNPs in relation to IGF-1 levels in young women from high-risk breast cancer families: implications for early-onset breast cancer.** *Fam Cancer* 2011, **10:** 173-185.

153. Cheng I, DeLellis HK, Haiman CA, Kolonel LN, Henderson BE, Freedman ML *et al*.: **Genetic determinants of circulating insulin-like growth factor (IGF)-I, IGF binding protein (BP)-1, and IGFBP-3 levels in a multiethnic population.** *J Clin Endocrinol Metab* 2007, **92:** 3660-3666.

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

155. Neuhausen SL, Brummel S, Ding YC, Singer CF, Pfeiler G, Lynch HT *et al*.: **Genetic variation in insulin-like growth factor signaling genes and breast cancer risk among BRCA1 and BRCA2 carriers.** *Breast Cancer Res* 2009, **11:** R76.

156. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA *et al*.: **Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer.** *Nat Genet* 2007, **39:** 865-869.

157. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K *et al*.: **Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study.** *J Natl Cancer Inst* 2011, **103:** 425-435.

158. Figueroa JD, Garcia-Closas M, Humphreys M, Platte R, Hopper JL, Southey MC *et al*.: **Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium.** *Hum Mol Genet* 2011, Dec 1;**20(23):4693-706**.

159. Milne RL, Goode EL, Garcia-Closas M, Couch FJ, Severi G, Hein R *et al*.: **Confirmation of 5p12 As a Susceptibility Locus for Progesterone-Receptor-Positive, Lower Grade Breast Cancer.** *Cancer Epidemiol Biomarkers Prev* 2011, **20:** 2222-2231.

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

161. Micalizzi DS, Farabaugh SM, Ford HL: **Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression.** *J Mammary Gland Biol Neoplasia* 2010, **15:** 117-134.

162. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M: **Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis.** *Lancet* 2004, **363:** 1346-1353.

163. Gronbaek H, Flyvbjerg A, Mellemkjaer L, Tjonneland A, Christensen J, Sorensen HT *et al*.: **Serum insulin-like growth factors, insulin-like growth factor binding proteins,**

and breast cancer risk in postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 2004, **13:** 1759-1764.

164. Shook D, Keller R: **Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development.** *Mech Dev* 2003, **120:** 1351-1383.

165. Choi SS, Diehl AM: **Epithelial-to-mesenchymal transitions in the liver.** *Hepatology* 2009, **50:** 2007-2013.

166. Fata JE, Werb Z, Bissell MJ: **Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes.** *Breast Cancer Res* 2004, **6:** 1-11.

167. Kristensen VN, Vaske CJ, Ursini-Siegel J, Van LP, Nordgard SH, Sachidanandam R *et al.*: **Integrated molecular profiles of invasive breast tumors and ductal carcinoma in situ (DCIS) reveal differential vascular and interleukin signaling.** *Proc Natl Acad Sci U S A* 2011.

168. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H *et al.*: **Stromal gene expression predicts clinical outcome in breast cancer.** *Nat Med* 2008, **14:** 518-527.

169. Byrne C: **Studying mammographic density: implications for understanding breast cancer.** *J Natl Cancer Inst* 1997, **89:** 531-533.

170. Wolfe JN: **Breast parenchymal patterns and their changes with age.** *Radiology* 1976, **121:** 545-552.

171. Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA *et al.*: **Mammographic features and breast cancer risk: effects with time, age, and menopause status.** *J Natl Cancer Inst* 1995, **87:** 1622-1629.

172. Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ *et al.*: **Mammographic breast density as an intermediate phenotype for breast cancer.** *Lancet Oncol* 2005, **6:** 798-808.

173. Ginsburg OM, Martin LJ, Boyd NF: **Mammographic density, lobular involution, and risk of breast cancer.** *Br J Cancer* 2008, **99:** 1369-1374.

174. Harvey JA, Bovbjerg VE: **Quantitative assessment of mammographic breast density: relationship with breast cancer risk.** *Radiology* 2004, **230:** 29-41.

175. McCormack VA, dos SS, I: **Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis.** *Cancer Epidemiol Biomarkers Prev* 2006, **15:** 1159-1169.

176. Yaghjyan L, Colditz GA, Collins LC, Schnitt SJ, Rosner B, Vachon C *et al.*: **Mammographic breast density and subsequent risk of breast cancer in**

postmenopausal women according to tumor characteristics. *J Natl Cancer Inst* 2011, **103:** 1179-1189.

177. Cil T, Fishell E, Hanna W, Sun P, Rawlinson E, Narod SA *et al.*: **Mammographic density and the risk of breast cancer recurrence after breast-conserving surgery.** *Cancer* 2009, **115:** 5780-5787.

178. Ma L, Fishell E, Wright B, Hanna W, Allan S, Boyd NF: **Case-control study of factors associated with failure to detect breast cancer by mammography.** *J Natl Cancer Inst* 1992, **%20;84:** 781-785.

179. Whitehead J, Carlile T, Kopecky KJ, Thompson DJ, Gilbert FI, Jr., Present AJ *et al.*: **Wolfe mammographic parenchymal patterns. A study of the masking hypothesis of Egan and Mosteller.** *Cancer* 1985, **56:** 1280-1286.

180. Li T, Sun L, Miller N, Nicklee T, Woo J, Hulse-Smith L *et al.*: **The association of measured breast tissue characteristics with mammographic density and other risk factors for breast cancer.** *Cancer Epidemiol Biomarkers Prev* 2005, **14:** 343-349.

181. Guo YP, Martin LJ, Hanna W, Banerjee D, Miller N, Fishell E *et al.*: **Growth factors and stromal matrix proteins associated with mammographic densities.** *Cancer Epidemiol Biomarkers Prev* 2001, **10:** 243-248.

182. Aiello EJ, Buist DS, White E, Porter PL: **Association between mammographic breast density and breast cancer tumor characteristics.** *Cancer Epidemiol Biomarkers Prev* 2005, **14:** 662-668.

183. Boyd NF, Connelly P, Byng J, Yaffe M, Draper H, Little L *et al.*: **Plasma lipids, lipoproteins, and mammographic densities.** *Cancer Epidemiol Biomarkers Prev* 1995, **4:** 727-733.

184. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al.*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci U S A* 2001, **98:** 10869-10874.

185. Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME: **Etiology of hormone receptor-defined breast cancer: a systematic review of the literature.** *Cancer Epidemiol Biomarkers Prev* 2004, **13:** 1558-1568.

186. Ziv E, Tice J, Smith-Bindman R, Shepherd J, Cummings S, Kerlikowske K: **Mammographic density and estrogen receptor status of breast cancer.** *Cancer Epidemiol Biomarkers Prev* 2004, **13:** 2090-2095.

187. Yang WT, Dryden M, Broglio K, Gilcrease M, Dawood S, Dempsey PJ *et al.*: **Mammographic features of triple receptor-negative primary breast cancers in young premenopausal women.** *Breast Cancer Res Treat* 2008, **111:** 405-410.

188. Conroy SM, Pagano I, Kolonel LN, Maskarinec G: **Mammographic density and hormone receptor expression in breast cancer: The Multiethnic Cohort Study.** *Cancer Epidemiol* 2011, **35:** 448-452.

189. Ding J, Warren R, Girling A, Thompson D, Easton D: **Mammographic density, estrogen receptor status and other breast cancer tumor characteristics.** *Breast J* 2010, **16:** 279-289.

190. Martin LJ, Boyd NF: **Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence.** *Breast Cancer Res* 2008, **10:** 201.

191. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ: **Mammographic densities and breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 1998, **7:** 1133-1144.

192. Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E: **Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer.** *Breast Cancer Res* 2005, **7:** R854-R861.

193. Muhimmah I., Oliver A., Denton E.R.E., Pont J., Pérez E., Zwiggelaar R.: **Comparison between wolfe, boyd, Bi-rads, and Tabar based mammographic risk assesment.** In *Digital Mammography: 8th Interantional workshop, IWDM 2006, Manchester, UK, June2006, Proceedings*. Edited by Astley S.M., Brady M., Zwiggelaar R. Berlin: Springer-Verlag; 2006:407-415.

194. Garrido-Estepa M, Ruiz-Perales F, Miranda J, Ascunce N, Gonzalez-Roman I, Sanchez-Contador C *et al*.: **Evaluation of mammographic density patterns: reproducibility and concordance among scales.** *BMC Cancer* 2010, **10:485.:** 485.

195. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S *et al*.: **Diagnostic performance of digital versus film mammography for breast-cancer screening.** *N Engl J Med* 2005, **353:** 1773-1783.

196. Ursin G., Qureshi SA: **Mammographic density - a useful biomarker for breast cancer risk in epidemiologic studies.** *Norsk Epidemiologi* 2009, **19:** 59-68.

197. Wolfe JN: **Breast patterns as an index of risk for developing breast cancer.** *AJR Am J Roentgenol* 1976, **126:** 1130-1137.

198. Gram IT, Funkhouser E, Tabar L: **The Tabar classification of mammographic parenchymal patterns.** *Eur J Radiol* 1997, **24:** 131-136.

199. American College of Radiology. American College of Radiology (ACR) Breast Imaging Reporting and Data System Atlas (BI-RADS® Atlas). 2003. Reston, Va.

200. Boyd NF, Jensen HM, Cooke G, Han HL: **Relationship between mammographic and histological risk factors for breast cancer.** *J Natl Cancer Inst* 1992, **84:** 1170-1179.

201. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ: **The quantitative analysis of mammographic densities.** *Phys Med Biol* 1994, **39:** 1629-1638.

202. Yaffe MJ, Boyd NF, Byng JW, Jong RA, Fishell E, Lockwood GA *et al*.: **Breast cancer risk and measured mammographic density.** *Eur J Cancer Prev* 1998, **7 Suppl 1:S47-55.:** S47-S55.

203. Ursin G, Astrahan MA, Salane M, Parisky YR, Pearce JG, Daniels JR *et al*.: **The detection of changes in mammographic densities.** *Cancer Epidemiol Biomarkers Prev* 1998, **7:** 43-47.

204. Khazen M, Warren RM, Boggis CR, Bryant EC, Reed S, Warsi I *et al*.: **A pilot study of compositional analysis of the breast and estimation of breast mammographic density using three-dimensional T1-weighted magnetic resonance imaging.** *Cancer Epidemiol Biomarkers Prev* 2008, **17:** 2268-2274.

205. Glide-Hurst CK, Duric N, Littrup P: **Volumetric breast density evaluation from ultrasound tomography images.** *Med Phys* 2008, **35:** 3988-3997.

206. Yaffe MJ: **Mammographic density. Measurement of mammographic density.** *Breast Cancer Res* 2008, **10:** 209.

207. Pearce MS, Tennant PW, Mann KD, Pollard TM, McLean L, Kaye B *et al*.: **Lifecourse predictors of mammographic density: the Newcastle Thousand Families cohort Study.** *Breast Cancer Res Treat* 2011.

208. Titus-Ernstoff L, Tosteson AN, Kasales C, Weiss J, Goodrich M, Hatch EE *et al*.: **Breast cancer risk factors in relation to breast density (United States).** *Cancer Causes Control* 2006, **17:** 1281-1290.

209. Greendale GA, Reboussin BA, Slone S, Wasilauskas C, Pike MC, Ursin G: **Postmenopausal hormone therapy and change in mammographic density.** *J Natl Cancer Inst* 2003, **95:** 30-37.

210. Marchesoni D, Driul L, Ianni A, Fabiani G, Della MM, Zuiani C *et al*.: **Postmenopausal hormone therapy and mammographic breast density.** *Maturitas* 2006, **53:** 59-64.

211. Tehranifar P, Reynolds D, Flom J, Fulton L, Liao Y, Kudadjie-Gyamfi E *et al*.: **Reproductive and menstrual factors and mammographic density in African American, Caribbean, and white women.** *Cancer Causes Control* 2011, **22:** 599-610.

212. Jeffreys M, Warren R, Gunnell D, McCarron P, Smith GD: **Life course breast cancer risk factors and adult breast density (United Kingdom).** *Cancer Causes Control* 2004, **15:** 947-955.

213. Tamimi RM, Eriksson L, Lagiou P, Czene K, Ekbom A, Hsieh CC *et al*.: **Birth weight and mammographic density among postmenopausal women in Sweden.** *Int J Cancer* 2010, **126:** 985-991.

214. Jeffreys M, Warren R, Highnam R, Davey SG: **Breast cancer risk factors and a novel measure of volumetric breast density: cross-sectional study.** *Br J Cancer* 2008, **98:** 210-216.

215. McCormack VA, dos SS, I, De Stavola BL, Perry N, Vinnicombe S, Swerdlow AJ *et al*.: **Life-course body size and perimenopausal mammographic parenchymal patterns in the MRC 1946 British birth cohort.** *Br J Cancer* 2003, **89:** 852-859.

216. Boyd NF, Martin LJ, Sun L, Guo H, Chiarelli A, Hislop G *et al*.: **Body size, mammographic density, and breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 2006, **15:** 2086-2092.

217. Key TJ, Appleby PN, Reeves GK, Roddam A, Dorgan JF, Longcope C *et al*.: **Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women.** *J Natl Cancer Inst* 2003, **%20;95:** 1218-1226.

218. Boyd NF, Stone J, Martin LJ, Jong R, Fishell E, Yaffe M *et al*.: **The association of breast mitogens with mammographic densities.** *Br J Cancer* 2002, **87:** 876-882.

219. Bremnes Y, Ursin G, Bjurstam N, Rinaldi S, Kaaks R, Gram IT: **Endogenous sex hormones, prolactin and mammographic density in postmenopausal Norwegian women.** *Int J Cancer* 2007, **121:** 2506-2511.

220. Tamimi RM, Hankinson SE, Colditz GA, Byrne C: **Endogenous sex hormone levels and mammographic density among postmenopausal women.** *Cancer Epidemiol Biomarkers Prev* 2005, **14:** 2641-2647.

221. Qureshi SA, Couto E, Hilsen M, Hofvind S, Wu AH, Ursin G: **Mammographic Density and Intake of Selected Nutrients and Vitamins in Norwegian Women.** *Nutr Cancer* 2011.

222. Brisson J, Verreault R, Morrison AS, Tennina S, Meyer F: **Diet, mammographic features of breast tissue, and breast cancer risk.** *Am J Epidemiol* 1989, **130:** 14-24.

223. Boyd NF, Greenberg C, Lockwood G, Little L, Martin L, Byng J *et al*.: **Effects at two years of a low-fat, high-carbohydrate diet on radiologic features of the breast: results from a randomized trial. Canadian Diet and Breast Cancer Prevention Study Group.** *J Natl Cancer Inst* 1997, **89:** 488-496.

224. Sala E, Warren R, Duffy S, Welch A, Luben R, Day N: **High risk mammographic parenchymal patterns and diet: a case-control study.** *Br J Cancer* 2000, **83:** 121-126.

225. Ursin G, Sun CL, Koh WP, Khoo KS, Gao F, Wu AH *et al*.: **Associations between soy, diet, reproductive factors, and mammographic density in Singapore Chinese women.** *Nutr Cancer* 2006, **56:** 128-135.

226. Cabanes A, Pastor-Barriuso R, Garcia-Lopez M, Pedraz-Pingarron C, Sanchez-Contador C, Vazquez Carrete JA *et al*.: **Alcohol, tobacco, and mammographic density: a population-based study.** *Breast Cancer Res Treat* 2011, **129:** 135-147.

227. Vachon CM, Kuni CC, Anderson K, Anderson VE, Sellers TA: **Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States).** *Cancer Causes Control* 2000, **11:** 653-662.

228. Qureshi SA, Couto E, Hofvind S, Wu AH, Ursin G: **Alcohol intake and mammographic density in postmenopausal Norwegian women.** *Breast Cancer Res Treat* 2011.

229. Cui Y, Rohan TE: **Vitamin D, calcium, and breast cancer risk: a review.** *Cancer Epidemiol Biomarkers Prev* 2006, **15:** 1427-1437.

230. Brisson J, Berube S, Diorio C, Sinotte M, Pollak M, Masse B: **Synchronized seasonal variations of mammographic breast density and plasma 25-hydroxyvitamin d.** *Cancer Epidemiol Biomarkers Prev* 2007, **16:** 929-933.

231. Knight JA, Vachon CM, Vierkant RA, Vieth R, Cerhan JR, Sellers TA: **No association between 25-hydroxyvitamin D and mammographic density.** *Cancer Epidemiol Biomarkers Prev* 2006, **15:** 1988-1992.

232. Green AK, Hankinson SE, Bertone-Johnson ER, Tamimi RM: **Mammographic density, plasma vitamin D levels and risk of breast cancer in postmenopausal women.** *Int J Cancer* 2010, **127:** 667-674.

233. Mishra G, McCormack V, Kuh D, Hardy R, Stephen A, dos SS, I: **Dietary calcium and vitamin D intakes in childhood and throughout adulthood and mammographic density in a British birth cohort.** *Br J Cancer* 2008, **99:** 1539-1543.

234. Ziv E, Shepherd J, Smith-Bindman R, Kerlikowske K: **Mammographic breast density and family history of breast cancer.** *J Natl Cancer Inst* 2003, **95:** 556-558.

235. Kataoka M, Antoniou A, Warren R, Leyland J, Brown J, Audley T *et al*.: **Genetic models for the familial aggregation of mammographic breast density.** *Cancer Epidemiol Biomarkers Prev* 2009, **18:** 1277-1284.

236. Stone J, Dite GS, Gunasekara A, English DR, McCredie MR, Giles GG *et al*.: **The heritability of mammographically dense and nondense breast tissue.** *Cancer Epidemiol Biomarkers Prev* 2006, **15:** 612-617.

237. Ursin G, Lillie EO, Lee E, Cockburn M, Schork NJ, Cozen W *et al*.: **The relative importance of genetics and environment on mammographic density.** *Cancer Epidemiol Biomarkers Prev* 2009, **18:** 102-112.

238. Stone J, Gurrin LC, Hayes VM, Southey MC, Hopper JL, Byrnes GB: **Sibship analysis of associations between SNP haplotypes and a continuous trait with application to mammographic density.** *Genet Epidemiol* 2010, **34:** 309-318.

239. Diorio C, Sinotte M, Brisson J, Berube S, Pollak M: **Vitamin D pathway polymorphisms in relation to mammographic breast density.** *Cancer Epidemiol Biomarkers Prev* 2008, **17:** 2505-2508.

240. Crandall CJ, Sehl ME, Crawford SL, Gold EB, Habel LA, Butler LM *et al*.: **Sex steroid metabolism polymorphisms and mammographic density in pre- and early perimenopausal women.** *Breast Cancer Res* 2009, **11:** R51.

241. Dumas I, Diorio C: **Polymorphisms in genes involved in the estrogen pathway and mammographic density.** *BMC Cancer* 2010, **10:** 636.

242. Maskarinec G, Lurie G, Williams AE, Le ML: **An investigation of mammographic density and gene variants in healthy women.** *Int J Cancer* 2004, **%20;112:** 683-688.

243. Haiman CA, Hankinson SE, De V, I, Guillemette C, Ishibe N, Hunter DJ *et al*.: **Polymorphisms in steroid hormone pathway genes and mammographic density.** *Breast Cancer Res Treat* 2003, **77:** 27-36.

244. Yong M, Schwartz SM, Atkinson C, Makar KW, Thomas SS, Newton KM *et al*.: **Associations between polymorphisms in glucuronidation and sulfation enzymes and mammographic breast density in premenopausal women in the United States.** *Cancer Epidemiol Biomarkers Prev* 2010, **19:** 537-546.

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

246. Li J, Eriksson L, Humphreys K, Czene K, Liu J, Tamimi RM *et al*.: **Genetic variation in the estrogen metabolic pathway and mammographic density as an intermediate phenotype of breast cancer.** *Breast Cancer Res* 2010, **12:** R19.

247. Taverne CW, Verheus M, McKay JD, Kaaks R, Canzian F, Grobbee DE *et al*.: **Common genetic variation of insulin-like growth factor-binding protein 1 (IGFBP-1), IGFBP-3, and acid labile subunit in relation to serum IGF-I levels and mammographic density.** *Breast Cancer Res Treat* 2010, **123:** 843-855.

248. Verheus M, McKay JD, Kaaks R, Canzian F, Biessy C, Johansson M *et al*.: **Common genetic variation in the IGF-1 gene, serum IGF-I levels and breast density.** *Breast Cancer Res Treat* 2007, **112:** 109-122.

249. Diorio C, Brisson J, Berube S, Pollak M: **Genetic Polymorphisms Involved in Insulin-like Growth Factor (IGF) Pathway in Relation to Mammographic Breast Density and IGF Levels.** *Cancer Epidemiol Biomarkers Prev* 2008, **17:** 880-888.

250. Biong M, Gram IT, Brill I, Johansen F, Solvang HK, Alnaes GI *et al*.: **Genotypes and haplotypes in the insulin-like growth factors, their receptors and binding proteins in relation to plasma metabolic levels and mammographic density.** *BMC Med Genomics* 2010, **%19;3:9.:** 9.

251. Lai JH, Vesprini D, Zhang W, Yaffe MJ, Pollak M, Narod SA: **A polymorphic locus in the promoter region of the IGFBP3 gene is related to mammographic breast density.** *Cancer Epidemiol Biomarkers Prev* 2004, **13:** 573-582.

252. dos SS, I, Johnson N, De SB, Torres-Mejia G, Fletcher O, Allen DS *et al.*: **The insulin-like growth factor system and mammographic features in premenopausal and postmenopausal women.** *Cancer Epidemiol Biomarkers Prev* 2006, **15:** 449-455.

253. Lindstrom S, Vachon CM, Li J, Varghese J, Thompson D, Warren R *et al.*: **Common variants in ZNF365 are associated with both mammographic density and breast cancer risk.** *Nat Genet* 2011, **43:** 185-187.

254. Odefrey F, Stone J, Gurrin LC, Byrnes GB, Apicella C, Dite GS *et al.*: **Common genetic variants associated with breast cancer and mammographic density measures that predict disease.** *Cancer Res* 2010, **70:** 1449-1458.

255. Woolcott CG, Maskarinec G, Haiman CA, Verheus M, Pagano IS, Le ML *et al.*: **Association between breast cancer susceptibility loci and mammographic density: the Multiethnic Cohort.** *Breast Cancer Res* 2009, **11:** R10.

256. Haakensen VD, Biong M, Lingjaerde OC, Holmen MM, Frantzen JO, Chen Y *et al.*: **Expression levels of uridine 5'-diphospho-glucuronosyltransferase genes in breast tissue from healthy women are associated with mammographic density.** *Breast Cancer Res* 2010, **12:** R65.

257. Engvall E, Perlmann P: **Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G.** *Immunochemistry* 1971, **8:** 871-874.

258. Berson SA, Yalow RS: **General principles of radioimmunoassay. 1968.** *Clin Chim Acta* 2006, **369:** 125-143.

259. NBCSP Quality Manual. Kreftregisteret . http://www.kreftregisteret.no/no/Forebyggende/Mammografiprogrammet/Kvalitetsmanual/ 15-12-2010. 22-9-2011.

260. Byng JW, Boyd NF, Little L, Lockwood G, Fishell E, Jong RA *et al.*: **Symmetry of projection in the quantitative analysis of mammographic images.** *Eur J Cancer Prev* 1996, **5:** 319-327.

261. Box GEP, COX DR: **An analysis of Transformations.** *Journal of the Royal Statistical Society Series B (Methodological),* 1964, **26:** 211-252.

262. CGAP. http://cgap.nci.nih.gov . 2006.

263. iPATH. http://escience.invitrogen.com/ipath/ . 2006.

264. Pathway Assist/Studio. http://www.ariadnegenomics.com/ . 2006.

265. Ensemble. http://www.ensembl.org/index.html . 1-8-2005.

266. CHIP. http://snpper.chip.org/ . 2006.

267. SIFT. http://sift.jcvi.org/ . 1-8-2011.

268. Ng PC, Henikoff S: **Predicting deleterious amino acid substitutions.** *Genome Res* 2001, **11:** 863-874.

269. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D: **Efficiency and power in genetic association studies.** *Nat Genet* 2005, **37:** 1217-1223.

270. Haploview. http://www.broad.mit.edu/mpg/haploview/ . 2006.

271. Botstein D, White RL, Skolnick M, Davis RW: **Construction of a genetic linkage map in man using restriction fragment length polymorphisms.** *Am J Hum Genet* 1980, **32:** 314-331.

272. Illumina. http://www.illumina.com/ . 2011.

273. Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB: **Bioinformatics challenges for personalized medicine.** *Bioinformatics* 2011, **27:** 1741-1748.

274. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS *et al.*: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.** *Proc Natl Acad Sci U S A* 2009, **106:** 9362-9367.

275. Heid CA, Stevens J, Livak KJ, Williams PM: **Real time quantitative PCR.** *Genome Res* 1996, **6:** 986-994.

276. Applied Biosystems. Essentials of Real time PCR: Guide. 2011.

277. Sequenom. http://www.sequenom.com/home/ . 2011.

278. Paul Oeth, Martin Beaulieu, Christopher Park, Dominik kosman, Guy del Mistro, Dirk van den Boom *et al.*. iPLEX Assay: Increased Plexing Efficiency and Flexibility for MassARRAY System Through Single Base Primer Extension wiht Mass-Modified Terminators. 28-4-2005.

279. Millis MP: **Medium-throughput SNP genotyping using mass spectrometry: multiplex SNP genotyping using the iPLEX(R) Gold assay.** *Methods Mol Biol* 2011, **700:61-76.:** 61-76.

280. Cannings C, Edwards AW: **Expected genotypic frequencies in a small sample: deviation from Hardy-Weinberg equilibrium.** *Am J Hum Genet* 1969, **21:** 245-247.

281. Gresham D, Dunham MJ, Botstein D: **Comparing whole genomes using DNA microarrays.** *Nat Rev Genet* 2008, **9:** 291-302.

282. Cleveland WS, Devlin SJ: **Locally Weighted REgression:An Approach to Regression Analysis by Local Fitting.** *Journal of the american statistical association* 1988, **83:** 596-610.

283. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS: **A genome-wide scalable SNP genotyping assay using microarray technology.** *Nat Genet* 2005, **37:** 549-554.

284. Douglas G.Altman: *Practical statistics for medical research*, 1 edn. Chapman & Hall; 1991.

285. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98:** 5116-5121.

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

287. Akey J, Jin L, Xiong M: **Haplotypes vs single marker linkage disequilibrium tests: what do we gain?** *Eur J Hum Genet* 2001, **9:** 291-300.

288. Excoffier L, Slatkin M: **Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population.** *Mol Biol Evol* 1995, **12:** 921-927.

289. Goeman JJ, van de Geer SA, de KF, van Houwelingen HC: **A global test for groups of genes: testing association with a clinical outcome.** *Bioinformatics* 2004, **20:** 93-99.

290. Ronald A.Fisher: *Statistical Methods For Research Workers*. Edinburgh: Oliver and Boyd; 1925.

291. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I: **Controlling the false discovery rate in behavior genetics research.** *Behav Brain Res* 2001, **125:** 279-284.

292. National Institute of Allergy and Infectious Disease. DAVID. NIH.http://david.abcc.ncifcrf.gov/ . 2011.

293. Huang dW, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37:** 1-13.

294. Huang dW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4:** 44-57.

295. Kanehisa Laboratories. KEGG. http://www.genome.jp/kegg/ . 2011.

296. Yao J, Weremowicz S, Feng B, Gentleman RC, Marks JR, Gelman R *et al.*: **Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression.** *Cancer Res* 2006, **66:** 4065-4078.

297. Verheus M, Maskarinec G, Erber E, Steude JS, Killeen J, Hernandez BY *et al*.: **Mammographic density and epithelial histopathologic markers.** *BMC Cancer* 2009, **9:182.:** 182.

298. Tamimi RM, Colditz GA, Wang Y, Collins LC, Hu R, Rosner B *et al*.: **Expression of IGF1R in normal breast tissue and subsequent risk of breast cancer.** *Breast Cancer Res Treat* 2011, **128:** 243-250.

299. NCBI. Entrez Gene. http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene . 2008.

300. Jones RA, Campbell CI, Gunther EJ, Chodosh LA, Petrik JJ, Khokha R *et al*.: **Transgenic overexpression of IGF-IR disrupts mammary ductal morphogenesis and induces tumor formation.** *Oncogene* 2007, **26:** 1636-1644.

301. Kleinberg DL, Wood TL, Furth PA, Lee AV: **Growth hormone and insulin-like growth factor-I in the transition from normal mammary development to preneoplastic mammary lesions.** *Endocr Rev* 2009, **30:** 51-74.

302. Dillon RL, White DE, Muller WJ: **The phosphatidyl inositol 3-kinase signaling network: implications for human breast cancer.** *Oncogene* 2007, **26:** 1338-1345.

303. Vogt PK, Hart JR, Gymnopoulos M, Jiang H, Kang S, Bader AG *et al*.: **Phosphatidylinositol 3-kinase: the oncoprotein.** *Curr Top Microbiol Immunol* 2010, **347:79-104.:** 79-104.

304. Stevens KN, Garcia-Closas M, Fredericksen Z, Kosel M, Pankratz VS, Hopper JL *et al*.: **Evaluation of variation in the phosphoinositide-3-kinase catalytic subunit alpha oncogene and breast cancer risk.** *Br J Cancer* 2011, 10.

305. Shupnik MA: **Crosstalk between steroid receptors and the c-Src-receptor tyrosine kinase pathways: implications for cell proliferation.** *Oncogene* 2004, **23:** 7979-7989.

306. Kelemen LE, Sellers TA, Vachon CM: **Can genes for mammographic density inform cancer aetiology?** *Nat Rev Cancer* 2008, **8:** 812-823.

307. Keely PJ: **Mechanisms by which the extracellular matrix and integrin signaling act to regulate the switch between tumor suppression and tumor promotion.** *J Mammary Gland Biol Neoplasia* 2011, **16:** 205-219.

308. Cantor CR: **The use of genetic SNPs as new diagnostic markers in preventive medicine.** *Ann N Y Acad Sci* 2005, **1055:48-57.:** 48-57.

309. Cordell HJ: **Detecting gene-gene interactions that underlie human diseases.** *Nat Rev Genet* 2009, **10:** 392-404.

310. Lynge E, Braaten T, Njor SH, Olsen AH, Kumle M, Waaseth M *et al*.: **Mammography activity in Norway 1983 to 2008.** *Acta Oncol* 2011, **50:** 1062-1067.

311. Yngve Bremnes: *Hormones, Smoking and Mammographic Density in Postmenopausal Women. The Tromsø Mammography and Breast Cancer Study.* 2007.

312. Gram IT, Lund E: **Breast cancer screening programme as setting for an adjunct research project: effect on programme attendance.** *J Med Screen* 2008, **15:** 44-45.

313. Schreer I: **Dense Breast Tissue as an Important Risk Factor for Breast Cancer and Implications for Early Detection.** *Breast Care (Basel)* 2009, **4:** 89-92.

314. Stratton MR, Rahman N: **The emerging landscape of breast cancer susceptibility.** *Nat Genet* 2008, **40:** 17-22.

315. Orr N, Chanock S: **Common genetic variation and human disease.** *Adv Genet* 2008, **62:1-32.:** 1-32.

316. de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA *et al.*: **Transferability of tag SNPs in genetic association studies in multiple populations.** *Nat Genet* 2006, **38:** 1298-1303.

317. Underwood AJ: *Expresiments in ecology. Their logical design and interpretation using analysis of variance*. Cambridge University Press; 1997.

318. Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C *et al.*: **Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density.** *J Natl Cancer Inst* 2006, **98:** 1215-1226.

319. Gail MH: **Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk.** *J Natl Cancer Inst* 2008, **100:** 1037-1041.

Errata

Reference 158:
Was originally:
Figueroa JD, Garcia-Closas M, Humphreys M, Platte R, Hopper JL, Southey MC *et al*.: **Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium.** *Hum Mol Genet* 2011, **%20.**.

Changed to:
Figueroa JD, Garcia-Closas M, Humphreys M, Platte R, Hopper JL, Southey MC *et al*.: **Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium.** *Hum Mol Genet* 2011, Dec 1;**20(23):4693-706**.

Figure 8:
Was originally:
Figure 1: Age-standardized breast cancer incidence rate per 100,000 among women worldwide [77].

Changed to:
Figure 2: Age-standardized breast cancer incidence rate per 100,000 among women worldwide in 2010 [77].

Page 29, para 1, last sentence:
Was originally:
Additionally, cyclic changes in MD are seen in premenopausal women due to the production and release of estrogen and progesterone from the ovaries [1]

Changed to:
Additionally, cyclic changes in the proliferation rate of breast epithelial cells are seen in premenopausal women due to the production and release of estrogen and progesterone from the ovaries [1]. Breast epithelial and stromal cells attenuate X-rays, and increased proliferation is positively associated with MD [180,191]. Hence, MD may be seen to change according to the menstrual cycle.

Page 29, para 2, line 2:
Was originally:
"…and while BMI is associated with breast cancer risk…"

Changed to:
"…and while BMI is positively associated with breast cancer risk…"

Page 59, section 4.2.3:

Was originally:
For the TMBC study however this was not an issue since being postmenopausal was a prerequisite to be eligible to participate, hence excluding pre- or perimenopausal women.

Changed to:
For the TMBC study women were classified as postmenopausal if they were 56 years or older reporting no menstruation during the last 12 months, or had a FSH level above 20 IU/L.

Additions to abbreviations list:
ANOVA: analysis of variance
BIRADS: Breast Imaging Reporting and Data System
BM: basement membrane
BMI: Body Mass Index
CC: craniocaudal
DNA: Deoxyribonucleic acid
E1: Estrone
E2: Estradiol
E3: Estriol
ECM: extracellular matrix
EMT: epithelial-mesenchymal transition
eQTL: expression quantitative trait loci
ER: Estrogen receptor
FFTP: First full term pregnancy
FWER: family wise error rate
GH: Growth hormone
GWAS: genome wide association studies
HT: postmenopausal hormone therapy
IARC: International agency for research on cancer
MLO: medio-lateral oblique
MMP: matrix metalloproteinase
NBCSP: Norwegian breast cancer screening program
PR: Progesterone receptor
ROI: region of interest
ROS: Reactive oxygen species
SAM: significant analysis of microarray
TDLU: terminal duct lobular unit
TEB: Terminal end bud

Table 2
Was originally:

Material name	N (healthy)	N (breast cancer)	Description	Paper
Tromsø Mammography and Breast Cancer study (TMBC)	433		Healthy norwegian postmenopausal women in the age group 55-71 years. Information on diet, menstrual and reproductive factors are obtained through an extensive questionnaire . All women have negative mammograms with no sign of cancer.	I & III
Mammographic Density and Genetics (MDG)	120	66	Norwegian pre- and postmenopausal women in the age group 22-87 years, with and without breast cancer. Information obtained on menstrual and reproductive factors, including serum hormone levels and mammographic density.	II& III

Changed to:

Material name	N (healthy)	N (breast cancer)	Description	Paper
Tromsø Mammography and Breast Cancer study (TMBC)	964*/ 433**		Healthy norwegian postmenopausal women in the age group 55-71 years. Information on diet, menstrual and reproductive factors are obtained through an extensive questionnaire . All women have negative mammograms with no sign of cancer.	I* & III**
Mammographic Density and Genetics (MDG)	120	66	Norwegian pre- and postmenopausal women in the age group 22-87 years, with and without breast cancer. Information obtained on menstrual and reproductive factors, including serum hormone levels and mammographic density.	II& III

I

Paper I

Margarethe Biong, Inger Torhild Gram, Ilene Brill, Fredrik Johansen, Grethe I.G. Alnaes, Toril Fagerheim, Yngve Bremnes, Stephen J. Chanock, Laurie Burdett, Meredith Yaeger, Giske Ursin, Vessela N.Kristensen.

Genotypes and haplotypes in the insulin-like growth factors, their receptors and binding proteins in relation to plasma metabolic levels and mammographic density.

BMC Medical Genomics 2010 Mar 19;3:9.

Biong et al. BMC Medical Genomics 2010, 3:9
http://www.biomedcentral.com/1755-8794/3/9

BMC Medical Genomics

RESEARCH ARTICLE Open Access

Genotypes and haplotypes in the insulin-like growth factors, their receptors and binding proteins in relation to plasma metabolic levels and mammographic density

Margarethe Biong[1], Inger T Gram[2], Ilene Brill[3], Fredrik Johansen[1], Hiroko K Solvang[1,4], Grethe IG Alnaes[1], Toril Fagerheim[5], Yngve Bremnes[2], Stephen J Chanock[6,7], Laurie Burdett[7], Meredith Yeager[7], Giske Ursin[8,9], Vessela N Kristensen[1,10*]

Abstract

Background: Increased mammographic density is one of the strongest independent risk factors for breast cancer. It is believed that one third of breast cancers are derived from breasts with more than 50% density. Mammographic density is affected by age, BMI, parity, and genetic predisposition. It is also greatly influenced by hormonal and growth factor changes in a woman's life cycle, spanning from puberty through adult to menopause. Genetic variations in genes coding for hormones and growth factors involved in development of the breast are therefore of great interest. The associations between genetic polymorphisms in genes from the IGF pathway on mammographic density and circulating levels of IGF1, its binding protein IGFBP3, and their ratio in postmenopausal women are reported here.

Methods: Samples from 964 postmenopausal Norwegian women aged 55-71 years were collected as a part of the Tromsø Mammography and Breast Cancer Study. All samples were genotyped for 25 SNPs in IGF1, IGF2, IGF1R, IGF2R, IGFALS and IGFBP3 using Taqman (ABI). The main statistical analyses were conducted with the PROC HAPLOTYPE procedure within SAS/GENETICS™ (SAS 9.1.3).

Results: The haplotype analysis revealed six haploblocks within the studied genes. Of those, four had significant associations with circulating levels of IGF1 or IGFBP3 and/or mammographic density. One haplotype variant in the IGF1 gene was found to be associated with mammographic density. Within the IGF2 gene one haplotype variant was associated with levels of both IGF1 and IGFBP3. Two haplotype variants in the IGF2R were associated with the level of IGF1. Both variants of the IGFBP3 haplotype were associated with the IGFBP3 level and indicate regulation in cis.

Conclusion: Polymorphisms within the IGF1 gene and related genes were associated with plasma levels of IGF1, IGFBP3 and mammographic density in this study of postmenopausal women.

Background

Increased mammographic density is one of the strongest independent risk factors for breast cancer [1-8]. The risk of developing breast cancer is four to six times higher in women with dense breast tissue compared to women with less dense tissue[2]. It has been estimated that breasts with more than 50% mammographic density give rise to one third of breast cancer cases[2]. Mammographic density is thus a stronger cancer risk factor than the most traditional risk factors such as nulliparity[1], age at first birth, age at menarche, age at menopause, use of postmenopausal hormone therapy (HT) and alcohol consumption. Mammographic density is influenced by age[2], body mass index (BMI)[2], parity[2], menopause status[2], HT[1,9], IGF1[10,11] and genetics[2].

* Correspondence: vessela.kristensen@medisin.uio.no
[1]Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Montebello 0310, Oslo, Norway

Exposure to endogenous and exogenous steroid hormones and growth factors has been linked to both increased mammographic density and breast cancer risk.

It has been estimated from twin studies that as much as 65% of the variation in mammographic density may be due to hereditary factors[12,13]. The genetic factors that influence mammographic density might be the same as the ones involved in the development of breast cancer[14,15]. The identification of the genes coding for these factors may therefore provide a better understanding of the genetics and the biology of the breast. Of particular interest are the insulin-like growth factor 1 and 2 (*IGF1/2*), which both have the ability to stimulate cell proliferation and inhibit cell death in many tissue types [16]. IGF1 is a mitogen predicted to be involved in the development of several human cancers, including breast cancer. In addition some studies have shown an association between circulating levels of IGF1 and increased levels of mammographic density [17-21]. In the bloodstream IGF1 binds to several IGF binding proteins (IGFBPs) which prolong its half-life and contribute to its delivery to target tissues[22]. IGFBP3, the principal carrier of IGF1 and IGF2[23], is mainly regulated by growth hormone (GH). IGFBP3 itself has the ability to promote apoptosis[24], thus it is sometimes referred to as an anticancer protein. The levels of IGF1 and IGFBP3 combined may be associated with breast cancer by stimulating proliferation of breast epithelial cells[22]. IGF1 bound to IGFBP3, may bind an acid-labile subunit (ALS) to form ternary complexes[25]. ALS is synthesized in the liver upon regulation of growth factors such as GH. By forming a ternary complex, ALS prolongs the half-lives of circulating IGFs from 10 minutes (free form), and 30-39 minutes (binary complex) to more than 12 hours. As a result, the reservoir of serum IGF1 levels in human adults can reach ~1000 fold that of insulin[26]. IGF1 and IGF2 both bind to the IGF1 receptor (IGF1R) to exert their growth promoting effects[27]. IGF2 may also bind to the IGF2 receptor (IGF2R) upon which it is internalized and degraded. Together, environmental and genetic factors determine the circulating levels of IGFs and their binding proteins[28]. Previous studies on polymorphisms/haplotypes in *IGFI*, *IGFBP3* and *IGFALS* and their association to breast cancer susceptibility [25,29-31] and circulating levels of both IGFI [32] and IGFBP3[31] have been reported making them interesting targets in mammographic density studies. Tamimi *et al*. found positive correlation between common genetic variants in *IGF1* and mammographic density[11]. Since variation in specific genes affects the levels of IGF1 and IGFBP3 and may influence mammographic density as well as breast cancer susceptibility, we set out to analyze 24 SNPs in *IGF1*, *IGF1R*, *IGF2*, *IGF2R*, *IGFBP3* and *IGFALS* and their association with mammographic density and plasma levels of IGF1 and IGFBP3 among 964 postmenopausal women.

Methods

Study Population

The Tromsø Mammography and Breast Cancer Study [21,33-37] was conducted among postmenopausal women, ages 55 to 71 years, residing in the municipality of Tromsø, Norway, and attending the population-based Norwegian Breast Cancer Screening Program at the University Hospital of North Norway. The women were recruited in the spring of 2001 and 2002. After the women had undergone their mammographic screening, they were interviewed by a trained research nurse about reproductive and menstrual details, previous history of cancer, smoking status, and use of postmenopausal hormone therapy and other medications. The participants had their height measured to the nearest centimeter and their weight measured to the nearest half kilogram. The women had blood samples drawn and were given a questionnaire to be completed at home. The questionnaire elicited information on demographics, additional menstrual and reproductive factors, as well as lifestyle and dietary factors. All women signed an informed consent. The National Data Inspection Board and the Regional Committee for Medical Research Ethics approved the study. Altogether, 1,041 women were included in the study. This accounted for 70.1% of the women attending the Norwegian Breast Cancer Screening Program during the recruitment period[37].

We excluded 22 women because of a previously (n = 16) or newly (n = 6) diagnosed breast cancer and one woman because of ongoing chemotherapy treatment. Women who were 56 years or older or who reported no menstruation during the last 12 months, or whose serum follicle-stimulating hormone level was above 20 IU/L, were classified as postmenopausal. By these criteria, three women were equivocal for menopausal status and excluded. We further excluded 11 women whose mammograms were unreadable for technical reasons. Seventeen women with missing blood samples and 23 women due to missing SNP analysis values were also excluded, leaving 964 women for analysis. More details are described elsewhere. http://uit.no/density

Mammographic classifications

Left craniocaudal mammograms were digitized using a Cobrascan CX-812 scanner (Radiographic Digital Imaging, Torrance, CA) at a resolution of 150 pixels/in. Percent and absolute mammographic densities were determined using the Madena computer-based threshold method, developed at the University of Southern California[38]. Briefly, the method was as follows: The digitized mammographic image was viewed on a computer

screen. A reader (trained by GU) outlined the total area of the breast using a computerized tool, and the software then counted the number of pixels within the outline. Mammographic density was then assessed (by GU) by first identifying a region of interest that incorporated all dense areas except those representing the pectoralis muscle and other scanning artifacts, and then applying a yellow tint to all pixels within the region of interest shaded at or above a threshold intensity of gray. The software then counted the tinted pixels, which represent the area of absolute density (ABDEN). Percent density (PDEN) equals the amount of absolute density divided by the total breast area.

The reader of the mammograms did not have any information of the characteristics of the study participants. More details are given elsewhere[33].

Peptide Assays

Nonfasting venous blood samples were obtained from the study participants at the day of mammographic screening. After centrifugation, plasma samples were stored at -70°C.

IGF1 and IGFBP3 levels were measured in ng/ml with the use of ELISA from Diagnostic Systems Laboratories, Inc. (Webster, TX). The IGF1 assays included an acid-ethanol precipitation to extract IGF1 from its binding proteins. Measurements were performed on previously never-thawed plasma samples.

All IGF1 and IGFBP3 assays were conducted at the laboratory for hormone analyses (Nutrition and Cancer Group, IARC, Lyon, France). The mean intrabatch coefficients of variation were 5.1% for IGF1 and 6.1% for IGFBP3. The interbatch coefficients of variation were 10.6% for IGF1 and 9% for IGFBP3.

The IGF1/IGFBP3 molar ratio was calculated as a possible indicator of IGF1 bioavailability. More details are described elsewhere[21].

DNA extraction

Peripheral blood from healthy women was collected in EDTA-tubes. The DNA was isolated by phenol/chloroform extraction followed by ethanol precipitation using the Applied Biosystems Model 340A Nucleic Acid Extractor and stored in TE-buffer at 2-8°C. The sample concentrations were measured by UV/Vis spectrophotometer (Nanodrop ND-1000).

Genotyping methods

5 ng of lyophilized sample DNA was used to perform a 5 ul Taqman (5' nuclease assay) reaction. Reactions were set up using 2.5 ul of the 2× Universal Master Mix (Applied Biosystems, Foster City, CA) and assay-specific concentrations of primers and probes. All reactions were set up in a 384 (96*4) well plate and heat-sealed using an ABgene ALPS 300 heat sealer and clear heat sealing film (ABgene, Rochester, NY). Reaction plates were thermocycled, and endpoint reads were conducted on the ABI 7900HT sequence detection system. Cluster Analysis was conducted on the scatter plot of Allele 1 Rn versus Allele 2 Rn. Genotypic segregation was displayed in the allelic plot, containing four distinct clusters, which represent the NTCs (no template controls) and three possible genotypes clusters along the horizontal, vertical and diagonal axes, which represent the Allele 1, Allele 2 and Allele 1/Allele 2 respectively. The data were exported in text format for further analysis. The sequences of the respective probes are given at http://snp500cancer.nci.nih.gov upon search for each SNP (rs number).

Statistical analysis

Simple descriptive statistics and other analyses were performed on the final study population with the use of SAS® 9.1.3 software. For each SNP within the haplotype regions evaluated, we calculated the allele and genotype frequencies using programming algorithms written in Base SAS®.

The haplotype analyses were performed in SAS/GENETICS using the PROC HAPLOTYPE procedure. This procedure utilizes the Expectation Maximization (EM) algorithm to predict the maximum likelihood estimates of the haplotype frequencies assuming Hardy-Weinberg equilibrium. The standard errors and the confidence intervals are estimated, by default, under a binomial assumption for each haplotype frequency estimate. In addition, the linkage disequilibrium (LD) option in PROC HAPLOTYPE was specified which generated a likelihood ratio test for linkage disequilibrium testing a null hypothesis of no association between the SNPs within a given haplotype region. The null hypothesis was rejected for all haplotype regions evaluated in this paper; the haplotype frequency estimates from the alternative hypothesis are reported here. The haplotype frequency threshold was set to 0.05, and haplotypes with a lower frequency were not included in the subsequent haplotype-trend regression analysis done in SAS®.

We applied the programming algorithms for haplotype-trend regression as developed by SAS/GENETICS; these methods also utilize the SAS/STAT procedure PROC REG for the regression models[39,40]. Haplotype-trend regression models [41] were used to assess whether haplotypes were associated with absolute mammographic density and percent mammographic density in square centimeters, in addition to the levels of IGF1, IGFBP3 measured in ng/ml and their molar ratio IGF1/IGFBP3. Box-Cox transformations were applied to absolute mammographic density and percent density to more closely approximate a normal distribution in this

study population [42]. The optimum power parameters of the Box-Cox transformation were obtained by the Dynamic programming called the Symplex method[43] and the code was written in R. All analyses were adjusted for age and BMI. The haplotypes within a gene were included individually in a model for comparison to the other haplotypes within the same gene [11].

For each haplotype region a model with haplotypes with frequencies >= 0.05 and the covariates age and BMI was compared to a model with age and BMI alone; the procedure computed the F statistic and the two-tailed significance probability for these models. In addition, all of the above-specified models were stratified by current HT users versus never and past HT users.

We also conducted additional analyses stratified by age and BMI and analyzed with general linear models using Box-Cox transformed values of mammographic density as the outcome.

We conducted further analyses of the association between individual SNPs in the haplotype regions and each of the outcome measurements. We assessed least square mean mammographic density by genotype adjusting for age and BMI using generalized linear models (PROC GLM). We estimated trend statistics by running separate models including genotype as an ordinal variable. For all the above-mentioned analyses we defined the results as statistically significant if the p-values were below 0.05.

Table 1 Selected characteristics of the 964 participants.

	All
Mean	
Age at screening, y	61.4 (± 4.6)
Age at menarche, y	13.3 (± 1.4)
Age at first birth*, y	22.9 (± 3.7)
Number of children*	2.9 (± 1.3)
Education, y	9.8 (± 3.4)
Age at menopause, y	48.6 (± 5.1)
BMI, kg/m^2	27.3 (± 4.8)
Frequency (%)	
Ever Oral Contraceptive use	51.1
Parous	92.6
Ever postmenopausal hormone therapy use	43.4
Current postmenopausal hormone therapy use	26.0
Past postmenopausal hormone therapy use	17.3
Never postmenopausal hormone therapy use	56.6
Breast cancer in first degree relative	8.3
Median	
Percent mammographic density, %	9.6 (0-69.2)
Absolute mammographic density, cm^2	14.8 (0-155.2)
Non-dense mammographic area, cm^2	147.0 (21.6-448.9)
IGF1 (ng/ml)	28.7 (2.2-104.6)
IGFBP3 (ng/ml)	151.2 (52.17-241.5)

*Among parous women only.

Data collected as part of the Tromsø Mammography and Breast Cancer(TMBC) study in Tromsø, Norway in 2001-2002, presented as mean (± standard deviation), frequency (%), and median (range).

Results

Characteristics of the Tromsø Mammographic and Breast Cancer Study

Shown in Table 1 are the selected characteristics of the participants. The median percent mammographic density was 9.6% (0.0-69.2%) with a mean of 12.7% (± 12.2%). The median absolute mammographic density was 14.8 cm^2 (0.0-155.2 cm^2), with mean of 19.3 cm^2 (± 20.3 cm^2).

Haplotype analysis

The allele and genotype frequencies for all studied SNPs in the genes IGF1, IGF1R, IGFBP3, IGF2, IGF2R and IGFALS are provided in Table 2.

The estimated common haplotypes (frequency higher than 5%) for the 6 genes and their association to the parameters studied are shown in Table 3. Haplotypes in four of the six studied genes were found significantly associated with the plasma levels and/or mammographic density (Table 3). One haplotype variant TG in *IGF1* was significantly associated with an increase in absolute mammographic density (p = 0.0334). However, upon stratification by current and past/never use of HT (See Additional files 1 and 2) this association was no longer significant (p = 0.1439 and p = 0.1271, respectively). In *IGF2* the rare common haplotype variant CA was associated with increased levels of both IGF1 (p = 0.0014) and IGFBP3 (p = 0.0181), with significant global association for these parameters p = 0.0138, p = 0.0408, respectively. After stratification by HT, the association between this haplotype and IGF1 levels was still statistically significant (p = 0.0011), and the association with IGFBP3 levels was borderline statistically significant (p = 0.0730), for women currently taking HT. For women who were past or never users of HT the association was no longer significant (p = 0.1261 and 0.2194 respectively). Haplotypes 1 and 5 in the *IGF2 receptor* (*IGF2R*) were found associated with decreased levels of circulating IGF1, p = 0.0397 and 0.0455, respectively. In addition, the global p-value of 0.0524 indicated an overall borderline association of all seven listed haplotypes in *IGF2R* with the IGF1 level. Stratification by HT status also revealed a significant association between haplotype 1 and IGF1 levels in never/past HT users (p = 0.0353), but not in current users (p = 0.7907). On the other hand, the association between haplotype 5 and IGF1 levels was no longer statistically significant in never/past HT users (p = 0.2031) but borderline significant in current HT users (p = 0.0637).

Table 2 Allele and genotype frequencies of the 964 participants.

Gene		Gene ID	SNP ID	Allele 1 frequency Allele 2 frequency	A1/A1 N = (%) A1/A2 N = (%) A2/A2 N = (%)
IGF1	1	IGF1-02	rs6220	T = 0.66703 C = 0.33297	C/C = 104 (11.2) T/C = 410 (44.2) T/T = 414 (44.6)
	2	IGF1-04	rs2162679	G = 0.16772 A = 0.83228	G/G = 22 (2.3) G/A = 275 (28.9) A/A = 654 (68.8)
IGF2	1	IGF2-02	rs734351	T = 0.62393 C = 0.37607	C/C = 129 (13.8) T/C = 443 (47.5) T/T = 360 (38.6)
	2	IGF2-03	rs3213216	G = 0.61645 A = 0.38355	A/A = 131 (14.0) G/A = 456 (48.7) G/G = 349 (37.3)
IGF1R	1	IGF1R-05	rs2137680	G = 0.70768 A = 0.29232	A/A = 92 (9.7) G/A = 372 (39.1) G/G = 487 (51.2)
	2	IGF1R-18	rs2175795	G = 0.70952 A = 0.29048	A/A = 90 (9.5) G/A = 369 (39.0) G/G = 486 (51.4)
	3	IGF1R-06	rs907806	G = 0.09201 A = 0.90799	G/G = 15 (1.6) G/A = 145 (15.2) A/A = 791 (83.2)
	4	IGF1R-04	rs3743258	G = 0.72569 A = 0.27431	A/A = 71 (7.5) G/A = 377 (39.9) G/G = 498 (52.6)
	5	IGF1R-26	rs3743259	G = 0.30765 A = 0.69235	G/G = 86 (9.1) G/A = 407 (43.3) A/A = 448 (47.6)
	6	IGF1R-03	rs2272037	C = 0.57354 T = 0.42646	T/T = 171 (18.1) C/T = 464 (49.1) C/C = 310 (32.8)
	7	IGF1R-01	rs2229765	G = 0.56019 A = 0.43981	A/A = 184 (19.4) G/A = 465 (49.1) G/G = 298 (31.5)
	8	IGF1R-07	rs2016347	T = 0.52784 G = 0.47216	G/G = 202 (21.2) T/G = 495 (52.0) T/T = 255 (26.8)
IGF2R	1	IGF2R-05	rs1570070	A = 0.63097 G = 0.36903	G/G = 132 (14.0) A/G = 432 (45.8) A/A = 379 (40.2)
	2	IGF2R-01	rs894817	G = 0.68873 A = 0.31127	A/A = 97 (10.6) G/A = 375 (41.0) G/G = 442 (48.4)
	3	IGF2R-02	rs998075	G = 0.49840 A = 0.50160	G/G = 235 (25) G/A = 467 (49.7) A/A = 238 (25.3)
	4	IGF2R-11	rs998074	C = 0.50211 T = 0.49789	T/T = 235 (24.8) C/T = 474 (50) C/C = 239 (25.2)
	5	IGF2R-04	rs629849	G = 0.86456 A = 0.13544	A/A = 21 (2.2) G/A = 211 (22.6) G/G = 702 (75.2)
	6	IGF2R-07	rs2282140	C = 0.89504 T = 0.10496	T/T = 6 (0.6) C/T = 187 (19.7) C/C = 755 (79.6)
	7	IGF2R-03	rs1803989	T = 0.09746 C = 0.90254	T/T = 10 (1.1) T/C = 164 (17.4) C/C = 770 (81.6)

Table 2: Allele and genotype frequencies of the 964 participants. *(Continued)*

IGFALS	1	IGFALS-05	rs9282731	T = 0.00105 C = 0.99895	T/T = 0 (0) T/C = 2 (0.2) C/C = 946 (99.8)
	2	IGFALS-01	rs17559	T = 0.09057 C = 0.90943	T/T = 13 (1.4) T/C = 145 (15.4) C/C = 786 (83.3)
	3	IGFALS-02	rs3751893	T = 0.21186 C = 0.78814	T/T = 41 (4.3) T/C = 318 (33.7) C/C = 585 (62.0)
IGFBP3	1	IGFBP3-05	rs9282734	C = 0.00317 A = 0.99683	C/C = 0 (0) C/A = 6 (0.6) A/A = 941 (99.4)
	2	IGFBP3-04	rs2471551	G = 0.17766 C = 0.82234	G/G = 27 (2.9) G/C = 280 (29.8) C/C = 633 (67.3)

Both haplotype variants identified in *IGFBP3* were significantly associated with IGFBP3 plasma levels with a global p-value of 0.0009. Haplotype variant 1 was associated with lower IGFBP3 levels (p = 0.0009) whilst haplotype 2 was associated with increased IGFBP3 levels (p = 0.0008). After stratification by HT, these associations were still significant in the group that were never/past HT users (p = 0.004 and 0.003 for haplotypes 1 and 2 respectively) but not among current HT users (p = 0.3833 and 0.4220).

None of the common haplotypes within the *IGF1 receptor* (*IGF1R*) haplotype or *IGFALS* haplotype were found significantly associated with any of the parameters studied. There was a borderline association between *IGFALS* haplotype 1 and 3 and IGFBP3 levels (p = 0.0582 and 0.0769 respectively). After stratification by HT, haplotype 1 was significantly associated with the IGFBP3 level (p = 0.0309), while haplotype 3 was still borderline significant (p = 0.0639) in the never/past HT group.

Stratified analysis by age and BMI

We examined if the significant association found between *IGF1* haplotype 4 and absolute mammographic density was specific to groups of age or BMI, and stratified the analysis by tertiles of age and tertiles of BMI (see Additional file 3). Women within the age groups > = 59 to < 64, and > = 64 years (Tertile 2 and 3) carrying the haplotype 4 variant had a trend towards higher levels of mammographic density (p = 0.0976, p = 0.0879). Stratification by BMI, revealed a trend towards higher levels of mammographic density (p = 0.0989) for the women in BMI tertile 2.

Single SNP analysis

In the six abovementioned genes we had 24 SNPs that were analyzed for association to the levels of IGF1, IGFBP3 and mammographic density (see Additional file 4). Eight of the 24 SNPs were found to be significantly associated with one or more of the parameters studied. In *IGF1R*, rs907806 was found to be significantly associated with the levels of IGFBP3 (p-trend = 0.0111). SNP rs3743259 was found to be significantly associated with mammographic density measured as both percent (p-trend = 0.0328) and absolute (p-trend = 0.0389) mammographic density. SNP rs2229765 and rs2016347 were significantly associated with mammographic density measured as both percent and absolute mammographic density, p-trend = 0.0265 and 0.0100, and, p-trend = 0.0434 and 0.0160 respectively. In *IGF2R*, rs998075 and rs998074 were found significantly associated with IGF1 (p-trend = 0.0072, p-trend = 0.0083) and IGFBP3 levels (p-trend = 0.0359, p-trend = 0.0320).

In *IGFALS*, rs9282731 was found significantly associated with the levels of IGFBP3 (p-trend = 0.0205). In *IGFBP3*, rs2471551 was found significantly associated with the levels of IGFBP3 (p-trend = 0.0009), denoting an association in cis.

Discussion

This population-based cross-sectional study shows an association between a common genetic haplotype in *IGF1* and absolute mammographic density in postmenopausal women after adjustment for age and BMI. Although not statistically significant, stratification by age and BMI revealed that the upper age tertiles and middle BMI tertile increased the mammographic density level. One haplotype in *IGF2* was associated with the levels of both IGF1 and IGFBP3, while two haplotypes in the *IGF2R* gene were associated with the levels of IGF1. Within the *IGFBP3* gene, two haplotypes were found associated with the IGFBP3 level indicating a regulation in cis.

The strength of our study is the large sample size and the fact that the samples were collected as part of a population-based screening project with high attendance

Table 3 Associations of the common haplotypes with IGF1, IGFBP3 and mammographic density levels.

		IGF1 level	IGFBP3 level	IGFratio	PDEN[a]	ABDEN[b]
IGF1						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. CG	*0,1185*	0.5553	0.9493	0.6462	0.2337	0.2693
2. CA	*0,21486*	0.1424	0.5075	0.4549	0.5147	0.6701
3. TA	*0,61718*	0.7733	0.7103	0.9016	0.4486	0.8462
4. TG	*0,04947*	0.1683	0.7237	0.2713	0.1737	***0.0334***
Global association		0.3342	0.924	0.6641	0.3069	0.1167
IGF2						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. TG	*0,34537*	0.2769	0.9845	0.1648	0.2297	0.4051
2. CG	*0,27102*	0.6093	0.1440	0.3058	0.5776	0.7190
3. TA	*0,27821*	0.7660	0.9920	0.5860	0.5254	0.6165
4. CA	*0,10541*	***0.0014***	***0.0181***	0.0974	0.8518	0.9403
Global association		***0.0138***	***0.0408***	0.2712	0.6353	0.8313
IGF1R						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. GGAGATGT	*0,09532*	0.8216	0.6761	0.7512	0.8689	0.9008
2. GGAGATAG	*0,07847*	0.1245	0.0885	0.5697	0.1153	0.1678
3. GGAGACGT	*0,12829*	0.8324	0.9646	0.9343	0.4403	0.6585
4. GGAAGCAG	*0,05028*	0.4479	0.6971	0.4485	0.7200	0.8879
5. GGAGACAG	*0,10502*	0.8550	0.9531	0.8820	0.1567	0.2125
Global association		0.5384	0.5162	0.8893	0.3016	0.4638
IGF2R						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. GAATACC	*0,06242*	***0.0397***	0.1228	0.2853	0.2140	0.0784
2. AGGCGCC	*0,27961*	0.2376	0.3324	0.5912	0.7018	0.6786
3. GAGCGTC	*0,06251*	0.7378	0.8651	0.9925	0.9465	0.7463
4. GAATGCC	*0,11233*	0.4057	0.4787	0.9129	0.7414	0.5293
5. AGATGCC	*0,2451*	***0.0455***	0.0631	0.4429	0.7105	0.8569
6. AGATACC	*0,07038*	0.4484	0.3361	0.8486	0.9180	0.8191
7. GAGCGCT	*0,06327*	0.5430	0.4985	0.6552	0.0575	0.0622
Global association		0.0524	0.1619	0.8720	0.5036	0.2659
IGFals						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. CTC	*0,08951*	0.8055	0.0582	0.2993	0.4069	0.5648
2. CCC	*0,69765*	0.2290	0.6156	0.3095	0.8880	0.9099
3. CCT	*0,21171*	0.1412	0.0769	0.6082	0.6291	0.7605
Global association		0.3387	0.0523	0.4678	0.6624	0.8261
IGFBP3						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. AG	*0,17767*	0.1630	***0.0009***	0.1595	0.5689	0.2204
2. AC	*0,81917*	0.1635	***0.0008***	0.1531	0.5551	0.2302
Global association		0.1630	***0.0009***	0.1595	0.5689	0.2204

All associations are adjusted by age and BMI. Significant p-values (< 0,05) are marked in bold italics.

[a]PDEN: Percent Density.

[b]ABDEN: Absolute Density, measured in cm^2.

rate[37]. Highly experienced personnel that were blinded to the characteristics of the women performed the reading and measurements of both the mammographic density and hormone levels. Also, we had information on age, BMI and HT use and were able to both adjust and stratify for these variables when necessary.

The limitation of our study is that the associations made with the polymorphisms within *IGF2R* were difficult to interpret due to the lack of measurements of its ligand IGF2. The women in the study were all postmenopausal, and some were taking HT, which could influence the circulating levels of IGF1 and IGFBP3. Furthermore, HT is demonstrated to have an impact on mammographic measurements, increasing the density. However, as HT cannot have influenced genotype, it is not technically a confounder, and as in most other analyses of mammographic density, we adjusted for age and BMI. We did however, stratify for HT, predominantly because of the possibility that HT use could be an effect modifier, i.e. have modified the effect of genotype on mammographic density.

Mammographic density is reduced by successive pregnancies and menopause, as well as with advancing age. Furthermore, mammographic density may reflect the cumulative exposure to hormones and growth factors that stimulate cell division and growth in the breast. Pike and colleagues proposed a model, stating that the effects of hormone exposure throughout life and the accumulation of genetic damage may cause an increased probability of breast cancer later in life [44]. The age-specific absolute risk of breast cancer caused by mammographic density is not yet determined, and it is unknown whether interventions that reduce cumulative exposure to density will reduce risk of breast cancer [2]. Because of the role of the IGF pathway in breast development and cellular proliferation, genetic variation within this pathway is of interest. Similar to mammographic density, IGF1 levels are related to age [22], BMI [45], and menopause status[21] and young women tend to have higher IGF1 levels than women in their postmenopausal years[22].

A positive association of IGF1 and IGFBP3 levels in relation to mammographic density in premenopausal women was found in most[17-20,32]but not all [46-48] previous studies. In postmenopausal women however the results are less consistent [18-21,47]. Data already published from this study were positive[21], suggesting an association between IGF1 and mammographic density. Our finding of a common haplotype in *IGF1* associated with mammographic density is in agreement with previously published findings of an association with IGF1 levels and density. Associations of genetic variants in *IGF1* and mammographic density in postmenopausal women can potentially better reflect the lifetime exposure of IGF1[11] compared to the IGF1 level measured at a certain time point, at a late stage, in a woman's life.

Other studies [18-20,46,47] such as the one of Dos Santos Silva *et al*. found no association between postmenopausal mammographic density and levels of IGF1, IGF2 or IGFBP3, nor the ratio of IGF1/IGFBP3, although, there was an association between the mammographic lucent area and IGFBP3 serum levels[47]. In summary, many studies have looked at variations in the IGF genes and their relationship with IGF plasma levels and mammographic density, but the results remain inconclusive, emphasizing the need for more studies.

The incidence of breast cancer has been associated with levels of IGF1 and IGFBP3 in premenopausal women in most [22,49-53], but not all studies [54,55]. Hankinson *et al*. performed a nested case-control study and found that IGF1 levels were higher among premenopausal women who developed breast cancer before age 50 than among age matched women who remained cancer free. For the postmenopausal women in the study no such association was established[22]. In postmenopausal women the findings are less clear and positive association of either IGF1, IGFBP3 or both with breast cancer [49,50,56] has been reported while other studies are negative[22,54,57]. The use of HT by postmenopausal women is known to lower both IGF1 and IGFBP3 levels significantly; thus the IGF1-associated increase in mammographic density seen in the non-HT users, may be difficult to observe in the HT users[21]. Nevertheless, one study reported an increased risk of breast cancer with increasing IGF1 levels also for postmenopausal HT-users (>55 years)[54]. Despite lowering the IGF1 levels, HT increases the mammographic density for most women and the age related decrease in mammographic density around the age of 55-64 does not commence in these women [58]. These findings support an emerging model of crosstalk between IGF1 and estrogens, suggesting that estrogens act through their receptor (ER) and affect the IGF1 expression[59].

IGF1 haplotypes

Variation in mammographic density due to polymorphisms in the *IGF1* gene has been reported in both pre- and postmenopausal women, but the association between genetic variants in *IGF1* and mammographic density in breast tissue in postmenopausal women has been inconclusive [11,17,46].

Among the common haplotypes analyzed in *IGF1* the least frequent haplotype was statistically significantly associated with an increase in ABDEN levels. This haplotype consists of the major allele of rs6220 and the minor allele of rs2162679. Other studies have reported an association between the minor allele of rs6220 and

mammographic density in premenopausal women[17,60] but to our knowledge, no other study has looked at this SNP in relation to postmenopausal mammographic density. Separately, the SNPs comprising this haplotype have been reported to be associated with IGF1 levels and breast cancer risk. The polymorphism rs6220 has also been correlated with elevated IGF1 levels whereas homozygosity G/G of rs2162679, has been associated with reduced breast cancer risk as well as reduced levels of IGFBP3[25]. This is in agreement with our observation of increased mammographic density for haplotype 4, given that low levels of IGFBP3 and high levels of IGF1 have been reported to increase mammographic density[19].

It is surprising to find the *IGF1* haplotype associated with increased mammographic density most strongly in women with higher BMI at an older age. However, since postmenopausal production of estrogens takes place predominantly in the adipose tissue, an increase in BMI would hypothetically result in increased estrogen levels. In turn estrogens may increase cellular IGF1 through crosstalk, and IGF1 may up-regulate the receptor response to estrogens[61]. This haplotype has a frequency of 0.04947 in the population studied which equal to 45 women and could be said to have little power. However, similar results have been reported by Muti *et al*. who reported that heavier postmenopausal women (BMI>26) had IGF1 levels associated with breast cancer risk [61]. Analysis of the SNPs in *IGF1* did not reveal any significant associations with any of the parameters studied (IGF1, IGFBP3, IGFratio or mammographic density), and thus we were unable to verify the previous findings regarding these SNPs and association to IGF1 levels and breast cancer risk. However, this is an indication that the aforementioned association of the *IGF1* haplotype 4 with mammographic density is dependent on the co-occurrence of these two SNPs.

IGFBP3 haplotypes

Several associations of *IGFBP3* polymorphisms and levels of IGFBP3 have been reported [25,46,47,62,63] one example is the -202(rs2854744) polymorphism associated with increased levels of IGFBP3 [46,47,62]. The -202 SNP has been associated with levels of IGF1 [60], IGFBP3 [46,47,60] and premenopausal mammographic density [46] but not with postmenopausal mammographic density [46,47]. The present study examined SNPs in the surrounding area of the -202 polymorphism which is an area suggested to be in strong LD[25,64].

The two *IGFBP3* haplotypes analyzed here were found to be significantly associated with the levels of IGFBP3 suggesting a putative regulatory effect *in cis*. The C allele of SNP rs2471551 of the *IGFBP3* haplotype has previously been associated with increased levels of IGFBP3 in combination with surrounding SNPs[25]. Upon single SNP analysis we found a significant trend of SNP rs2471551 with the level of IGFBP3 indicating that having two copies of the frequent allele C increase the least square mean of IGFBP3 compared to having two copies of the rare allele G, confirming the finding of Canzian *et al*. 2006. In the haplotype analysis we found that the AC haplotype was associated with higher levels of IGFBP3 than the AG haplotype and can thus confirm the result published by Canzian *et al*. that SNP rs2471551 is associated with increased IGFBP3 levels [25]. In addition, these haplotypes signify a trend of association with low levels (AG) and high levels (AC) of mammographic density (Table 3). These findings are consistent with the review of Fletcher *et al*. in which most reports agree on increased breast cancer risk with high levels of IGFBP3[65]. This may most strongly apply to premenopausal women, in whom the IGF levels are higher than in postmenopausal[22], and where the IGF axis is postulated to have an increased role because of involvement of the sex hormones[52]. The emerging belief that higher levels of IGFBP3 may decrease mammographic density and may decrease the risk of cancer [22,25,66] through low IGF1/IGFBP3 ratio must be further substantiated.

IGF2 haplotypes

Even though it is known that the circulating IGF2 concentration is much higher than that of IGF1, there is limited evidence on its mitogenic activity in relation to breast cancer and disease[47,49,56], thus implications of IGF2 on breast cancer risk are inconclusive. Studies on genetic variants of *IGF2* in relation to breast disease are few, and to our knowledge, the present study is the first to look at *IGF2* polymorphisms in relation to levels of mammographic density, IGF1 and IGFBP3. The finding of a common haplotype (4, Table 3) significantly associated with higher levels of both IGF1 and IGFBP3 may be explained by a decrease in clearance of IGF1 due to potentially lower levels of IGF2. IGFBP3 is the principal carrier of both IGF1 and IGF2 and the possibility of a regulatory feedback of IGF1 and IGFBP3 through polymorphisms in *IGF2* cannot be excluded. However, we did not have measurements of IGF2 and could therefore not test this. The association with haplotype 4 was no longer seen in analyses of the single *IGF2* SNPs and thus there is reason to believe that the association is dependent on the combination of the two SNPs comprising this haplotype. After stratifying the analysis by HT the association was still significant for the women currently taking HT, implying that HT could be an effect modifier of this association. The sex hormones play an important role in the IGF axis and could be the reason why the association is stronger in these women.

IGF1R haplotypes

The IGF1 receptor (IGF1R) sets off a complex cascade of signals upon binding of its ligands, IGF1 and IGF2 [67]. IGF1R functions as an anti-apoptotic agent by enhancing cell survival, and has been found expressed in most breast cancer cell lines[68] and highly over-expressed in most malignant tissues[69]. In a study on genetic variation and breast cancer survival, Deming *et al*. [64] found SNP rs951715 within the *IGF1R* gene associated with breast cancer survival in postmenopausal women, whereas SNP rs2229765 included in the present study, was not[64]. SNP rs2229765 results in a silent mutation, and has thus far not been found associated with any epidemiological traits. In the haplotype analysis none of the *IGF1R* haplotypes were found significantly associated with any of the studied parameters. The single SNP analysis revealed significant association of SNP rs2229765 with both percent and absolute mammographic density, increased numbers of the G allele increased the least squares means of mammographic density. In addition the SNPs rs3743259 and rs2016347 were also significantly associated with percent and absolute mammographic density, and in both cases increased number of the most frequent allele increased the least square mean of mammographic density. Functional studies are needed to investigate if these SNPs influence the affinity to IGF1 and IGF2 increasing their growth promoting effects and possibly mammographic density.

IGF2R haplotypes

IGF1 and IGF2 send their mitogenic and antiapoptotic signals through a common thyrosine kinase receptor, the IGF1R. Modulation of the mitogenic pathway occurs in part via the M6P/IGF2R, which functions in the internalization and degradation of IGF2[27]. IGF2R is also important in the activation process of TGFβ, which amongst other properties has the ability to inhibit cell growth. Loss of heterozygosity (LOH) of the M6P/IGF2R has been linked to liver and breast cancers, whereas somatic mutations of the M6P/IGF2R have been found in cancers of the prostate, lung, endometrium, brain, stomach and colorectum[27]. Chen *et al*. found that decreased ribosomal expression of the receptor leads to increased proliferation of MCF7 cells by a IGF2 related mechanism, mediated through IGF1R [27]. These findings have led to the suggestion that *IGF2R* is a tumor suppressor gene. Our results show that two of the *IGF2R* haplotypes are significantly associated to decreased levels of IGF1. For haplotype 1 the association was still significant after stratification for HT, in women that are never or past users of HT. Postmenopausal women that are not under the influence of hormones potentially have lower IGF1 levels. Analysis performed on the individual SNPs supports this finding, with a significant association of two *IGF2R* SNPs and the levels of IGF1, in addition they are also significantly associated with the levels of IGFBP3. IGF1 which is produced in the liver is influenced by several factors such as growth hormone and insulin, and its bioavailability is regulated by IGF2, IGFBPs and Als (acid-labile protein subunit) [65]. It is well known that IGF2 can act through IGF1R, in contrast IGF1 does not act through IGF2R, and to our knowledge no association between *IGF2R* and IGF1 levels have been described. IGF2R is able to degrade IGF2 and thereby regulates the circulating concentration of IGF2, in turn IGF2 clearance has the ability to regulate the level of IGF1. Thus, a possible explanation for the association of the two haplotypes and the SNP within *IGF2R* with IGF1 levels could be a change in clearance of IGF2 levels leading to decreased production of IGF1 through a regulative feedback loop. Whether or not such an interaction is present between IGF2 and IGF1 levels is impossible to confirm, due to lacking measurements of circulating IGF2.

IGFALS haplotypes

Despite being an important member in IGF regulation, few studies have looked at this protein and variations within it in regards to breast cancer[25,64]. Canzian *et al*. [25] studied three SNPs within exon 2 of *IGFALS* in regards to breast cancer risk, two of which are included in the *IGFALS* haplotype of our study (rs3751893, rs17559), and found that homozygous carriers of SNP rs3751893 were associated with reduced circulating levels of IGF1. Deming *et al*. conducted a study on *IGFALS* promoter SNPs in relation to menopausal status but found no association [64].

The *IGFALS* haplotypes in this study were not significantly associated with neither the levels of IGF1, IGFBP3, their ratio nor mammographic density. Stratification by HT of the haplotype analysis revealed a significant association of the never/past HT group with the level of IGFBP3, in addition the SNP analysis revealed significant association of SNP rs9282731 with the level of IGFBP3. Increased number of the C allele increases the IGFBP3 level compared to the rare allele G. One hypothesis could be that the C allele modifies the IGFALS and reduces either its affinity or reduces its level causing increased level of free IGFBP3. Although the SNP analysis is based on low frequencies, functional studies could to be done to verify such a hypothesis. Further investigation into the role of this protein is needed to establish its involvement in the development of mammographic density and breast cancer.

Conclusion

In conclusion, haplotypes were defined for each of the six genes from the *IGF* pathway studied here. Four

genes had common haplotype variants (>5%) significantly associated with the metabolic levels of the gene products and mammographic density. One haplotype variant in *IGF1* was found associated with mammographic density. In *IGF2* one haplotype variant was associated with the level of both IGF1 and IGFBP3. Two haplotype variants and two SNPs in *IGF2R* were associated with the levels of IGF1. Both variants of the *IGFBP3* haplotype and one SNP were associated with IGFBP3 level, indicating a regulatory function in *cis*.

Additional file 1: Haplotype analysis stratified by current HT use. Associations of the common haplotypes with IGF1, IGFBP3 and mammographic density levels stratified by current HT use.

Additional file 2: Haplotype analysis stratified by never/past HT use. Associations of the common haplotypes with IGF1, IGFBP3 and mammographic density levels stratified by never past HT use.

Additional file 3: IGF haplotype association stratified by age and BMI. Significant association of IGF1 haplotype 4 with ABDEN stratified by age and BMI tertiles.

Additional file 4: Associations of single SNPs with IGF1, IGFBP3 levels, IGFratio, and mammographic density. PDEN-Percent Density, ABDEN-Absolute density. LSMEANS-Least squares Means. Measurements of mammographic density are boxcox transformed.

Abbreviations

SNP: Single Nucleotide Polymorphism; ABI: Applied Biosystems; HT: hormone therapy; GH: growth hormone; ALS: Acid Labile Subunit; TMBC: Tromsø Mammography and Breast Cancer study; ELISA: Enzyme-linked immunosorbent assay; IARC: International Agency for Research on Cancer; EDTA: ethylenediaminetetraacetic acid; DNA: deoxyribonucleic acid; UV: Ultra Violet light; NTC: No template control; EM: Expectation Maximization; LD: Linkage Disequilibrium; BMI: Body Mass Index; ABDEN: Absolute Breast Density; PDEN: Percent Breast Density.

Acknowledgements

We specially thank the women who participated in the Tromsø Mammography and Breast Cancer Study. The study was conducted in collaboration with the Department of Clinical Research and the Department of Radiology, Center for Breast Imaging, University Hospital of North Norway, the Cancer Registry of Norway and the Norwegian Women and Cancer Study (PI Professor Eiliv Lund). The study was supported by grants from the EXTRA funds from Norwegian Foundation for Health and Rehabilitation 2003/2/0068, the University of Tromsø; the Norwegian Cancer Society, the Aakre Foundation, the Norwegian Women's Public Health Association, the National Cancer Institute grant R03CA105948 (PI Professor Gertraud Maskarinec), Northern Norway Regional Health Authority. MB was supported by a PhD fellowship grant from The Research Council of Norway, grant 175240/S10. FJ was supported by the Norwegian Cancer Society. We would also like to thank Tonje Braaten for statistical advice.

Author details

[1]Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Montebello 0310, Oslo, Norway. [2]Institute of Community Medicine, University of Tromsø, Breivika, Norway. [3]University of Alabama at Birmingham, School of Public Health, Department of Epidemiology, Birmingham, Alabama, USA. [4]Department of Biostatistics, Institute of Basic Medical Science, University of Oslo, Oslo, Norway. [5]Department of Genetics, University Hospital of North Norway, Tromsø; Norway. [6]NCI, NIH, Pediatric Branch, Bethesda, USA. [7]Core Genotyping Facility, Advanced Technology Center, National Cancer Institute, Bethesda, MD, 20892, USA. [8]Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway. [9]Department of Preventive Medicine University of Southern California Keck School of Medicine, Los Angeles, CA, USA. [10]Faculty of Medicine, (Faculty Division Ahus), UiO, Oslo, Norway.

Authors' contributions

MB analyzed the results and wrote the manuscript, IB carried out the statistical analysis in SAS in collaboration with TF and ITG. FJ contributed to the statistical analysis in Haploview, HS contributed to the interpretation of the statistical analyses. GIGA performed the DNA isolation and sample preparation. YB was involved in the epidemiological part of the studies SC at the NCI. NIH provided laboratory analyses, where MY designed the assays and LB was instrumental for the genotyping. GU read the mammograms and supervised the statistical analysis, ITG is the principle investigator of the Tromsø Mammography and Breast Cancer Study and VNK is the principal investigator of the molecular part of the present study. VNK established the concept, designed and organized this study. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 7 August 2009 Accepted: 19 March 2010
Published: 19 March 2010

References

1. Harvey JA, Bovbjerg VE: **Quantitative assessment of mammographic breast density: relationship with breast cancer risk.** *Radiology* 2004, **230**:29-41.
2. Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ, *et al*: **Mammographic breast density as an intermediate phenotype for breast cancer.** *Lancet Oncol* 2005, **6**:798-808.
3. Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, *et al*: **Mammographic features and breast cancer risk: effects with time, age, and menopause status.** *J Natl Cancer Inst* 1995, **87**:1622-1629.
4. Byrne C, Schairer C, Brinton LA, Wolfe J, Parekh N, Salane M, *et al*: **Effects of mammographic density and benign breast disease on breast cancer risk (United States).** *Cancer Causes Control* 2001, **12**:103-110.
5. Byrne C: **Studying mammographic density: implications for understanding breast cancer.** *J Natl Cancer Inst* 1997, **89**:531-533.
6. Ursin G, Hovanessian-Larsen L, Parisky YR, Pike MC, Wu AH: **Greatly increased occurrence of breast cancers in areas of mammographically dense tissue.** *Breast Cancer Res* 2005, **7**:R605-R608.
7. Russo J, Mailo D, Hu YF, Balogh G, Sheriff F, Russo IH: **Breast differentiation and its implication in cancer prevention.** *Clin Cancer Res* 2005, **11**:931s-936s.
8. Ziv E, Shepherd J, Smith-Bindman R, Kerlikowske K: **Mammographic breast density and family history of breast cancer.** *J Natl Cancer Inst* 2003, **95**:556-558.
9. Greendale GA, Reboussin BA, Slone S, Wasilauskas C, Pike MC, Ursin G: **Postmenopausal hormone therapy and change in mammographic density.** *J Natl Cancer Inst* 2003, **95**:30-37.
10. Haiman CA, Hankinson SE, De VI, Guillemette C, Ishibe N, Hunter DJ, *et al*: **Polymorphisms in steroid hormone pathway genes and mammographic density.** *Breast Cancer Res Treat* 2003, **77**:27-36.
11. Tamimi RM, Cox DG, Kraft P, Pollak MN, Haiman CA, Cheng I, *et al*: **Common genetic variation in IGF1, IGFBP-1, and IGFBP-3 in relation to mammographic density: a cross-sectional study.** *Breast Cancer Res* 2007, **9**:R18.
12. Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MR, *et al*: **Heritability of mammographic density, a risk factor for breast cancer.** *N Engl J Med* 2002, **19;347**:886-894.
13. Stone J, Dite GS, Gunasekara A, English DR, McCredie MR, Giles GG, *et al*: **The heritability of mammographically dense and nondense breast tissue.** *Cancer Epidemiol Biomarkers Prev* 2006, **15**:612-617.
14. Kelemen LE, Sellers TA, Vachon CM: **Can genes for mammographic density inform cancer aetiology?** *Nat Rev Cancer* 2008, **8**:812-823.
15. Martin LJ, Boyd NF: **Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence.** *Breast Cancer Res* 2008, **10**:201.
16. Pollak M: **Insulin-like growth factor physiology and cancer risk.** *Eur J Cancer* 2000, **36**:1224-1228.

17. Verheus M, McKay JD, Kaaks R, Canzian F, Biessy C, Johansson M, *et al*: **Common genetic variation in the IGF-1 gene, serum IGF-I levels and breast density.** *Breast Cancer Res Treat* 2007, **112**:109-122.
18. Boyd NF, Stone J, Martin LJ, Jong R, Fishell E, Yaffe M, *et al*: **The association of breast mitogens with mammographic densities.** *Br J Cancer* 2002, **87**:876-882.
19. Diorio C, Pollak M, Byrne C, Masse B, Hebert-Croteau N, Yaffe M, *et al*: **Insulin-like growth factor-I, IGF-binding protein-3, and mammographic breast density.** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:1065-1073.
20. Byrne C, Colditz GA, Willett WC, Speizer FE, Pollak M, Hankinson SE: **Plasma insulin-like growth factor (IGF) I, IGF-binding protein 3, and mammographic density.** *Cancer Res* 2000, **60**:3744-3748.
21. Bremnes Y, Ursin G, Bjurstam N, Rinaldi S, Kaaks R, Gram IT: **Insulin-like Growth Factor and Mammographic Density in Postmenopausal Norwegian Women.** *Cancer Epidemiol Biomarkers Prev* 2007, **16**:57-62.
22. Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B, *et al*: **Circulating concentrations of insulin-like growth factor-I and risk of breast cancer.** *Lancet* 1998, **351**:1393-1396.
23. OMIM: **NCBI.** 2008, Ref Type: Electronic Citation.
24. Diorio C, Berube S, Byrne C, Masse B, Hebert-Croteau N, Yaffe M, *et al*: **Influence of insulin-like growth factors on the strength of the relation of vitamin D and calcium intakes to mammographic breast density.** *Cancer Res* 2006, **66**:588-597.
25. Canzian F, McKay JD, Cleveland RJ, Dossus L, Biessy C, Rinaldi S, *et al*: **Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study.** *Br J Cancer* 2006, **94**:299-307.
26. Boisclair YR, Rhoads RP, Ueki I, Wang J, Ooi GT: **The acid-labile subunit (ALS) of the 150 kDa IGF-binding protein complex: an important but forgotten component of the circulating IGF system.** *J Endocrinol* 2001, **170**:63-70.
27. Chen Z, Ge Y, Landman N, Kang JX: **Decreased expression of the mannose 6-phosphate/insulin-like growth factor-II receptor promotes growth of human breast cancer cells.** *BMC Cancer* 2002, **2**:**18**:18.
28. Harrela M, Koistinen H, Kaprio J, Lehtovirta M, Tuomilehto J, Eriksson J, *et al*: **Genetic and environmental components of interindividual variation in circulating levels of IGF-I, IGF-II, IGFBP-1, and IGFBP-3.** *J Clin Invest* 1996, **98**:2612-2615.
29. Cheng I, Penney KL, Stram DO, Le ML, Giorgi E, Haiman CA, *et al*: **Haplotype-based association studies of IGFBP1 and IGFBP3 with prostate and breast cancer risk: the multiethnic cohort 60.** *Cancer Epidemiol Biomarkers Prev* 2006, **15**:1993-1997.
30. Ren Z, Cai Q, Shu XO, Cai H, Li C, Yu H, *et al*: **Genetic polymorphisms in the IGFBP3 gene: association with breast cancer risk and blood IGFBP-3 protein levels among Chinese women 61.** *Cancer Epidemiol Biomarkers Prev* 2004, **13**:1290-1295.
31. Al-Zahrani A, Sandhu MS, Luben RN, Thompson D, Baynes C, Pooley KA, *et al*: **IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer 62.** *Hum Mol Genet* 2006, **15**:1-10.
32. Johansson H, Gandini S, Bonanni B, Mariette F, Guerrieri-Gonzaga A, Serrano D, *et al*: **Relationships between circulating hormone levels, mammographic percent density and breast cancer risk factors in postmenopausal women.** *Breast Cancer Res Treat* 2007, **108**:57-67.
33. Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E: **Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer.** *Breast Cancer Res* 2005, **7**: R854-R861.
34. Bremnes Y, Ursin G, Bjurstam N, Rinaldi S, Kaaks R, Gram IT: **Endogenous sex hormones, prolactin and mammographic density in postmenopausal Norwegian women.** *Int J Cancer* 2007, **121**:2506-2511.
35. Bremnes Y, Ursin G, Bjurstam N, Lund E, Gram IT: **Different types of postmenopausal hormone therapy and mammographic density in Norwegian women.** *Int J Cancer* 2007, **120**:880-884.
36. Bremnes Y, Ursin G, Bjurstam N, Gram IT: **Different measures of smoking exposure and mammographic density in postmenopausal Norwegian women: a cross-sectional study.** *Breast Cancer Res* 2007, **9**:R73.
37. Gram IT, Lund E: **Breast cancer screening programme as setting for an adjunct research project: effect on programme attendance.** *J Med Screen* 2008, **15**:44-45.
38. Ursin G, Astrahan MA, Salane M, Parisky YR, Pearce JG, Daniels JR, *et al*: **The detection of changes in mammographic densities.** *Cancer Epidemiol Biomarkers Prev* 1998, **7**:43-47.
39. Czika W, Yu X: *SAS/Genetics PROC HAPLOTYPE SAS Help Documentation* 2003, Ref Type: Computer Program.
40. Czika W, Yu X: **Gene Frequencies and Linkage Disequilibrium.** ***Genetic Analysis of Complex Traits Using SAS.*** NC: SAS PublishingArnold Myron Saxton 2004.
41. Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG: **Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals.** *Hum Hered* 2002, **53**:79-91.
42. Box GEP, Cox DR: **An analysis of transformations.** *Journal of the Royal Statistical Society* 1964, **26**:211-522.
43. Dantzig GB, Thapa MN: *Linear Programming:Theory and Extensions* Princeton, Princeton University Press 1963.
44. Pike MC, Krailo MD, Henderson BE, Casagrande JT, Hoel DG: **"Hormonal" risk factors, "breast tissue age" and the age-incidence of breast cancer.** *Nature* 1983, **303**:767-770.
45. Gram IT, Norat T, Rinaldi S, Dossus L, Lukanova A, Tehard B, *et al*: **Body mass index, waist circumference and waist-hip ratio and serum levels of IGF-I and IGFBP-3 in European women.** *Int J Obes (Lond)* 2006, **30**:1623-1631.
46. Lai JH, Vesprini D, Zhang W, Yaffe MJ, Pollak M, Narod SA: **A polymorphic locus in the promoter region of the IGFBP3 gene is related to mammographic breast density.** *Cancer Epidemiol Biomarkers Prev* 2004, **13**:573-582.
47. dos SSI, Johnson N, De SB, Torres-Mejia G, Fletcher O, Allen DS, *et al*: **The insulin-like growth factor system and mammographic features in premenopausal and postmenopausal women.** *Cancer Epidemiol Biomarkers Prev* 2006, **15**:449-455.
48. Maskarinec G, Takata Y, Chen Z, Gram IT, Nagata C, Pagano I, *et al*: **IGF-I and mammographic density in four geographic locations: a pooled analysis.** *Int J Cancer* 2007, **121**:1786-1792.
49. Allen NE, Roddam AW, Allen DS, Fentiman IS, dos SSI, Peto J, *et al*: **A prospective study of serum insulin-like growth factor-I (IGF-I), IGF-II, IGF-binding protein-3 and breast cancer risk.** *Br J Cancer* 2005, **92**:1283-1287.
50. Toniolo P, Bruning PF, Akhmedkhanov A, Bonfrer JM, Koenig KL, Lukanova A, *et al*: **Serum insulin-like growth factor-I and breast cancer.** *Int J Cancer* 2000, **88**:828-832.
51. Krajcik RA, Borofsky ND, Massardo S, Orentreich N: **Insulin-like growth factor I (IGF-I), IGF-binding proteins, and breast cancer.** *Cancer Epidemiol Biomarkers Prev* 2002, **11**:1566-1573.
52. Yu H, Shu XO, Li BD, Dai Q, Gao YT, Jin F, *et al*: **Joint effect of insulin-like growth factors and sex steroids on breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 2003, **12**:1067-1073.
53. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M: **Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis.** *Lancet* 2004, **363**:1346-1353.
54. Kaaks R, Lundin E, Rinaldi S, Manjer J, Biessy C, Soderberg S, *et al*: **Prospective study of IGF-I, IGF-binding proteins, and breast cancer risk, in northern and southern Sweden.** *Cancer Causes Control* 2002, **13**:307-316.
55. Petridou E, Papadiamantis Y, Markopoulos C, Spanos E, Dessypris N, Trichopoulos D: **Leptin and insulin growth factor I in relation to breast cancer (Greece).** *Cancer Causes Control* 2000, **11**:383-388.
56. Gronbaek H, Flyvbjerg A, Mellemkjaer L, Tjonneland A, Christensen J, Sorensen HT, *et al*: **Serum insulin-like growth factors, insulin-like growth factor binding proteins, and breast cancer risk in postmenopausal women.** *Cancer Epidemiol Biomarkers Prev* 2004, **13**:1759-1764.
57. Kahan Z, Gardi J, Nyari T, Foldesi I, Hajnal-Papp R, Ormandi K, *et al*: **Elevated levels of circulating insulin-like growth factor-I, IGF-binding globulin-3 and testosterone predict hormone-dependent breast cancer in postmenopausal women: a case-control study 7.** *Int J Oncol* 2006, **29**:193-200.
58. Sterns EE, Zee B: **Mammographic density changes in perimenopausal and postmenopausal women: is effect of hormone replacement therapy predictable?** *Breast Cancer Res Treat* 2000, **59**:125-132.
59. Martin MB, Stoica A: **Insulin-like growth factor-I and estrogen interactions in breast cancer.** *J Nutr* 2002, **132**:3799S-3801S.

60. Diorio C, Brisson J, Berube S, Pollak M: **Genetic Polymorphisms Involved in Insulin-like Growth Factor (IGF) Pathway in Relation to Mammographic Breast Density and IGF Levels.** *Cancer Epidemiol Biomarkers Prev* 2008, **17**:880-888.
61. Muti P, Quattrin T, Grant BJ, Krogh V, Micheli A, Schunemann HJ, *et al*: **Fasting glucose is a risk factor for breast cancer: a prospective study.** *Cancer Epidemiol Biomarkers Prev* 2002, **11**:1361-1368.
62. Deal C, Ma J, Wilkin F, Paquette J, Rozen F, Ge B, *et al*: **Novel promoter polymorphism in insulin-like growth factor-binding protein-3: correlation with serum levels and interaction with known regulators.** *J Clin Endocrinol Metab* 2001, **86**:1274-1280.
63. Patel AV, Cheng I, Canzian F, Le ML, Thun MJ, Berg CD, *et al*: **IGF-1, IGFBP-1, and IGFBP-3 polymorphisms predict circulating IGF levels but not breast cancer risk: findings from the Breast and Prostate Cancer Cohort Consortium (BPC3).** *PLoS ONE* 2008, **3**:e2578.
64. Deming SL, Ren Z, Wen W, Shu XO, Cai Q, Gao YT, *et al*: **Genetic variation in IGF1, IGF-1R, IGFALS, and IGFBP3 in breast cancer survival among Chinese women: a report from the Shanghai Breast Cancer Study.** *Breast Cancer Res Treat* 2007, **104**:309-319.
65. Fletcher O, Gibson L, Johnson N, Altmann DR, Holly JM, Ashworth A, *et al*: **Polymorphisms and circulating levels in the insulin-like growth factor system and risk of breast cancer: a systematic review.** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:2-19.
66. Bohlke K, Cramer DW, Trichopoulos D, Mantzoros CS: **Insulin-like growth factor-I in relation to premenopausal ductal carcinoma in situ of the breast.** *Epidemiology* 1998, **9**:570-573.
67. Ellis MJ, Jenkins S, Hanfelt J, Redington ME, Taylor M, Leek R, *et al*: **Insulin-like growth factors in human breast cancer.** *Breast Cancer Res Treat* 1998, **52**:175-184.
68. Peyrat JP, Bonneterre J: **Type 1 IGF receptor in human breast diseases.** *Breast Cancer Res Treat* 1992, **22**:59-67.
69. Entrez Gene: **NCBI.** 2008, Ref Type: Electronic Citation.

Pre-publication history
The pre-publication history for this paper can be accessed here:http://www.biomedcentral.com/1755-8794/3/9/prepub

doi:10.1186/1755-8794-3-9
Cite this article as: Biong *et al.*: Genotypes and haplotypes in the insulin-like growth factors, their receptors and binding proteins in relation to plasma metabolic levels and mammographic density. *BMC Medical Genomics* 2010 **3**:9.

Additional file 1: Haplotype analysis stratified by current HT use. Association of the common haplotypes with IGF1, IGFBP3 and mammographic density levels stratified by current HT use.

Current HT users		IGF1 level	IGFBP3 level	IGFratio	PDENboxcox	ABDENboxcox
IGF1						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. CG	*0.09490*	0.1219	0.3873	0.1913	0.6427	0.9555
2. CA	*0.22268*	0.4679	0.0968	0.1782	0.8978	0.9776
3. TA	*0.63014*	0.5387	0.7018	0.0565	0.9660	0.5053
4. TG	*0.05228*	0.4944	0.2504	0.9214	0.4535	0.1439
Global association		0.3976	0.2876	0.2518	0.8522	0.5117
IGF2						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. TG	*0.33323*	***0.0112***	0.8501	***0.0052***	0.9663	0.8901
2. CG	*0.28817*	0.7290	0.2706	0.2614	0.5640	0.3282
3. TA	*0.30190*	0.2844	0.7782	0.5182	0.4749	0.2534
4. CA	*0.07670*	***0.0011***	0.0730	0.0648	0.7246	0.5542
Global association		***0.0048***	0.2282	***0.0394***	0.8848	0.6262
IGF1R						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. GGAGATGT	*0.11012*	0.6563	0.8098	0.5256	0.4780	0.4310
2. GGAGATAG	*0.10101*	0.0772	0.6412	0.0936	0.4562	0.6239
3. GGAGACGT	*0.11925*	0.7626	0.6978	0.5733	0.5906	0.6480
4. GGAAGCAG	*0.05133*	0.4913	0.7193	0.3297	0.8369	0.7753
5. GGAGACAG	*0.08880*	0.8123	0.4810	0.3479	0.2996	0.5989
Global association		0.3637	0.9029	0.2485	0.7580	0.9147
IGF2R						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. GAATACC	*0.07013*	0.7907	0.3831	0.2320	0.5556	0.4664
2. AGGCGCC	*0.27946*	0.1284	0.8579	0.0649	0.8623	0.8078
3. GAGCGTC	*0.06144*	0.8951	0.4026	0.6172	0.9766	0.7003
4. GAATGCC	*0.11360*	0.1035	0.0917	0.5339	0.4646	0.6433
5. AGATGCC	*0.25907*	0.0637	0.3975	0.1730	0.4472	0.4238
6. AGATACC	*0.07700*	***0.0468***	0.0746	0.7336	0.6399	0.9008
7. GAGCGCT	*0.05188*	0.2023	0.7381	0.3635	0.8605	0.9104
Global association		0.0931	0.3578	0.5098	0.9606	0.9672
IGFals						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. CTC	*0.07513*	0.9466	0.8964	0.9862	0.9928	0.8924
2. CCC	*0.70523*	0.1870	0.5946	0.2548	0.3173	0.4770
3. CCT	*0.21738*	0.1734	0.7706	0.1383	0.2099	0.3217
Global association		0.3857	0.9452	0.3275	0.4515	0.6125
IGFBP3						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. AG	*0.17080*	0.9717	0.3833	0.5014	0.2985	0.3816
2. AC	*0.82306*	0.9656	0.4220	0.5814	0.4115	0.5423
Global association		0.9717	0.3833	0.5014	0.2985	0.3816

Additional file 2: Haplotype analysis stratified by never/past HT use. Associations of the common haplotypes with IGF1, IGFBP3 and mammographic density levels stratified by past HT use.

Never/past HT users		IGF1 level	IGFBP3 level	IGFratio	PDENboxcox	ABDENboxcox
IGF1						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. CG	*0.12673*	0.7721	0.8803	0.9686	0.3266	0.2583
2. CA	*0.21224*	0.1749	0.8104	0.0935	0.4752	0.6031
3. TA	*0.61245*	0.6206	0.9348	0.3911	0.4133	0.5658
4. TG	*0.04858*	0.2537	0.6457	0.1912	0.2820	0.1271
Global association		0.4683	0.9705	0.2971	0.4656	0.2818
IGF2						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. TG	*0.35104*	0.8659	0.9345	0.9613	0.1648	0.3001
2. CG	*0.26352*	0.6899	0.3343	0.6472	0.3728	0.3659
3. TA	*0.26845*	0.4284	0.9350	0.3122	0.9405	0.7504
4. CA	*0.11698*	0.1261	0.2194	0.4518	0.3689	0.3842
Global association		0.2786	0.4505	0.6651	0.5398	0.6419
IGF1R						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. GGAGATGT	*0.09271*	0.5184	0.3785	0.9830	0.4638	0.5384
2. GGAGATAG	*0.07590*	0.3968	***0.0660***	0.7267	0.3592	0.4056
3. GGAGACGT	*0.13436*	0.9310	0.7015	0.6464	0.7911	0.9179
4. GGAAGCAG	*0.04894*	0.9523	0.5311	0.8849	0.6729	0.8562
5. GGAGACAG	*0.10686*	0.9469	0.9057	0.8972	0.3545	0.2963
Global association		0.8757	0.3048	0.9970	0.5721	0.6319
IGF2R						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. GAATACC	0.05947	***0.0353***	***0.0162***	0.6714	0.2321	0.0852
2. AGGCGCC	0.27986	0.6724	0.1855	0.5560	0.7252	0.7431
3. GAGCGTC	0.06317	0.8761	0.6086	0.8140	0.8361	0.7588
4. GAATGCC	0.11287	0.8820	0.6460	0.6785	0.5531	0.4353
5. AGATGCC	0.24032	0.2031	0.1104	0.8861	0.3359	0.7428
6. AGATACC	0.06749	0.8584	0.9716	0.6940	0.7871	0.9058
7. GAGCGCT	0.06651	0.9600	0.5990	0.9363	***0.0192***	***0.0266***
Global association		0.2256	0.0641	0.9940	0.2558	0.1752
IGFals						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. CTC	*0.09452*	0.9784	***0.0309***	0.1889	0.3988	0.6116
2. CCC	*0.69535*	0.4304	0.7848	0.5236	0.4818	0.6116
3. CCT	*0.20935*	0.3723	0.0639	0.8156	0.8464	0.8129
Global association		0.6680	***0.0291***	0.4218	0.6651	0.8364
IGFBP3						
Haplotype	*Frequency*	p-value	p-value	p-value	p-value	p-value
1. AG	0.18006	0.0925	***0.0004***	0.2294	0.9920	0.3924
2. AC	0.81781	0.1027	***0.0003***	0.1880	0.8599	0.3203
Global association		0.0925	***0.0004***	0.2294	0.9920	0.3924

Additional file 3: IGF1 haplotype association stratified by age and BMI.

Significant association of IGF1 haplotype 4with ABDEN stratified by age and BMI tertiles.

IGF1			ABDEN				
Haplotype	rs6220	rs2162679	Tertile	Age	p-value	BMI	p-value
4	T	G	1	<59	0.8158	<24.9158	0.9698
			2	>=59 to <64	0.0976	>=24.9158 to <28.7805	0.0989
			3	>=64	0.0879	>=28.7805	0.1073

Additional file 4: Associations of single SNPs with IGF1 and IGFBP3 levels, IGFratio and mammographic density. PDEN-Percent Density, ABDEN-Absolute density. LSMEANS-least squares means

Gene	Order of occurence on chromosome	Gene ID	SNP ID	Allele 1 frequency Allele 2 frequency	A1/A1 N= (%) A1/A2 N= (%) A2/A2 N= (%)	IGF1 LSMEANS	P-value for genotype association with IGF1 levels	IGF1 trend p-value	IGFBP3 LSMEANS	P-value for genotype association with IGFBP3 levels	IGFBP3 trend p-value	IGFratio LSMEANS	P-value for genotype association with IGFratio levels	IGFratio trend p-value
IGF1	1	IGF1-02	rs6220	T= 0.66703 C= 0.33297	C/C = 104 (11.2) T/C = 410 (44.2) T/T = 414 (44.6)	228.794538 224.083569 221.376975	C/C = 0.3651 T/C = 0.6034 T/T = -	0.3568	4409.208498 4332.192603 4325.345624	C/C = 0.4413 T/C = 0.9210 T/T = -	0.5341	0.193880 0.194584 0.192954	C/C = 0.8802 T/C = 0.6758 T/T = -	0.7569
	2	IGF1-04	rs2162679	G= 0.16772 A= 0.83228	G/G = 22 (2.3) G/A = 275 (28.9) A/A = 654 (68.8)	215.353065 218.686234 224.992534	G/G = 0.5488 G/A = 0.2384 A/A = -	0.2056	4102.464140 4357.195733 4330.723549	G/G = 0.2855 G/A = 0.7096 A/A = -	0.8211	0.202565 0.188851 0.195301	G/G = 0.5458 G/A = 0.1072 A/A = -	0.3076
IGF2	1	IGF2-02	rs734351	T= 0.62393 C= 0.37607	C/C = 129 (13.8) T/C = 443 (47.5) T/T = 360 (38.6)	231.700386 223.298586 220.944296	C/C = 0.1619 T/C = 0.6561 T/T = -	0.2015	4318.783581 4348.113044 4328.873081	C/C = 0.9214 T/C = 0.7844 T/T = -	0.9800	0.201855 0.193980 0.191098	C/C = 0.0621 T/C = 0.4659 T/T = -	0.0795
	2	IGF2-03	rs3213216	G= 0.61645 A= 0.38355	A/A = 131 (14.0) G/A = 456 (48.7) G/G = 349 (37.3)	228.553393 224.028072 218.226897	A/A = 0.1747 G/A = 0.2709 G/G = -	0.1376	4409.657416 4338.007912 4282.942655	A/A = 0.2130 G/A = 0.4348 G/G = -	0.2008	0.194120 0.194376 0.192333	A/A = 0.7554 G/A = 0.6074 G/G = -	0.6629
IGF1R	1	IGF1R-05	rs2137680	G= 0.70768 A= 0.29232	A/A = 92 (9.7) G/A = 372 (39.1) G/G = 487 (51.2)	217.992383 228.592738 220.459622	A/A = 0.7713 G/A = 0.1119 G/G = -	0.5321	4225.606145 4373.490458 4327.649819	A/A = 0.3667 G/A = 0.5014 G/G = -	0.7633	0.195805 0.196470 0.191600	A/A = 0.5080 G/A = 0.2040 G/G = -	0.2515
	2	IGF1R-18	rs2175795	G= 0.70952 A= 0.29048	A/A = 90 (9.5) G/A = 369 (39.0) G/G = 486 (51.4)	218.001326 228.362361 220.200928	A/A = 0.7974 G/A = 0.1117 G/G = -	0.5099	4183.674205 4370.167977 4327.356000	A/A = 0.2071 G/A = 0.5301 G/G = -	0.5635	0.197411 0.196514 0.191356	A/A = 0.3459 G/A = 0.1803 G/G = -	0.1642
	3	IGF1R-06	rs907806	G= 0.09201 A= 0.90799	G/G = 15 (1.6) G/A = 145 (15.2) A/A = 791 (83.2)	243.079740 226.843648 222.239103	G/G = 0.2832 G/A = 0.4954 A/A = -	0.2561	4829.481627 4458.529866 4301.457745	G/G = **0.0403** G/A = 0.0785 A/A = -	**0.0111**	0.186989 0.190931 0.194466	G/G = 0.6057 G/A = 0.4831 A/A = -	0.3954
	4	IGF1R-04	rs3743258	G= 0.72569 A= 0.27431	A/A = 71 (7.5) G/A = 377 (39.9) G/G = 498 (52.6)	227.298447 224.057247 221.808059	A/A = 0.5639 G/A = 0.6590 G/G = -	0.5133	4440.351866 4339.652717 4308.088331	A/A = 0.2938 G/A = 0.6400 G/G = -	0.3235	0.192306 0.194692 0.193425	A/A = 0.8750 G/A = 0.7399 G/G = -	0.9252
	5	IGF1R-26	rs3743259	G= 0.30765 A= 0.69235	G/G = 86 (9.1) G/A = 407 (43.3) A/A = 448 (47.6)	227.233377 222.137698 223.132243	G/G = 0.6414 G/A = 0.8455 A/A = -	0.8240	4426.887944 4322.784657 4324.955586	G/G = 0.3832 G/A = 0.9745 A/A = -	0.5503	0.194257 0.193374 0.193759	G/G = 0.9396 G/A = 0.9198 A/A = -	0.9991
	6	IGF1R-03	rs2272037	C= 0.57354 T= 0.42646	T/T = 171 (18.1) C/T = 464 (49.1) C/C = 310 (32.8)	220.814249 224.474830 223.816195	T/T = 0.6740 C/T = 0.9046 C/C = -	0.7349	4302.416981 4335.252957 4366.154643	T/T = 0.4961 C/T = 0.6691 C/C = -	0.4887	0.193755 0.193717 0.193797	T/T = 0.9937 C/T = 0.9844 C/C = -	0.9909
	7	IGF1R-01	rs2229765	G= 0.56019 A= 0.43981	A/A = 184 (19.4) G/A = 465 (49.1) G/G = 298 (31.5)	213.809051 227.912175 222.199595	A/A = 0.2290 G/A = 0.3016 G/G = -	0.3802	4260.644802 4356.513705 4342.373925	A/A = 0.3795 G/A = 0.8476 G/G = -	0.4546	0.188564 0.196926 0.192602	A/A = 0.4391 G/A = 0.2954 G/G = -	0.6248
	8	IGF1R-07	rs2016347	T= 0.52784 G= 0.47216	G/G = 202 (21.2) T/G = 495 (52.0) T/T = 255 (26.8)	221.899176 223.837216 222.745263	G/G = 0.9041 T/G = 0.8498 T/T = -	0.9237	4361.390334 4354.449126 4270.875418	G/G = 0.3313 T/G = 0.2749 T/T = -	0.3061	0.192148 0.193088 0.196213	G/G = 0.4379 T/G = 0.4679 T/T = -	0.4220

Gene	Order of occurence on chromosome	Gene ID	SNP ID	Allele 1 frequency Allele 2 frequency	A1/A1 N= (%) A1/A2 N= (%) A2/A2 N= (%)	IGF1 LSMEANS	P-value for genotype association with IGF1 levels	IGF1 trend p-value	IGFBP3 LSMEANS	P-value for genotype association with IGFBP3 levels	IGFBP3 trend p-value	IGFratio LSMEANS	P-value for genotype association with IGFratio levels	IGFratio trend p-value
IGF2R	1	IGF2R-05	rs1570070	A= 0.63097 G= 0.36903	G/G = 132 (14.0) A/G = 432 (45.8) A/A = 379 (40.2)	222.638281 223.916504 222.325485	G/G = 0.9668 A/G = 0.7617 A/A = -	0.8827	4408.668272 4299.501522 4339.172580	G/G = 0.4856 A/G = 0.5687 A/A = -	0.7325	0.190130 0.196489 0.192061	G/G = 0.7310 A/G = 0.2591 A/A = -	0.8805
	2	IGF2R-01	rs894817	G= 0.68873 A= 0.31127	A/A = 97 (10.6) G/A = 375 (41.0) G/G = 442 (48.4)	218.180110 220.044936 225.033016	A/A = 0.4073 G/A = 0.3364 G/G = -	0.2773	4319.705592 4302.097350 4360.740914	A/A = 0.7093 G/A = 0.3963 G/G = -	0.4886	0.190243 0.192900 0.193636	A/A = 0.5838 G/A = 0.8500 G/G = -	0.6173
	3	IGF2R-02	rs998075	G= 0.49840 A= 0.50160	G/G = 235 (25) G/A = 467 (49.7) A/A = 238 (25.3)	233.365157 222.718986 214.958679	G/G = **0.0072** G/A = 0.1901 A/A = -	**0.0072**	4443.124966 4324.092379 4252.918600	G/G = **0.0357** G/A = 0.3633 A/A = -	**0.0359**	0.196427 0.194707 0.189408	G/G = 0.1713 G/A = 0.2325 A/A = -	0.1698
	4	IGF2R-11	rs998074	C= 0.50211 T= 0.49789	T/T = 235 (24.8) C/T = 474 (50) C/C = 239 (25.2)	215.345824 222.169514 233.377994	T/T =**0.0083** C/T = 0.0578 C/C = -	**0.0083**	4255.853620 4313.974239 4450.471132	T/T =**0.0321** C/T = 0.0824 C/C = -	**0.0320**	0.189646 0.194807 0.196046	T/T =0.2113 C/T = 0.7798 C/C = -	0.2115
	5	IGF2R-04	rs629849	G= 0.86456 A= 0.13544	A/A = 21 (2.2) G/A = 211 (22.6) G/G = 702 (75.2)	223.201710 218.969538 224.156740	A/A = 0.9539 G/A = 0.3737 G/G = -	0.4555	4314.456970 4322.687004 4338.752836	A/A = 0.9121 G/A = 0.8364 G/G = -	0.8222	0.191607 0.190543 0.194598	A/A = 0.8093 G/A = 0.3548 G/G = -	0.3831
	6	IGF2R-07	rs2282140	C= 0.89504 T= 0.10496	T/T = 6 (0.6) C/T = 187 (19.7) C/C = 755 (79.6)	250.250342 227.048876 222.072193	T/T = 0.3573 C/T = 0.4157 C/C = -	0.2833	4537.366780 4375.895743 4320.964217	T/T = 0.5936 C/T = 0.4986 C/C = -	0.4167	0.200427 0.194327 0.193641	T/T = 0.7666 C/T = 0.8809 C/C = -	0.8090
	7	IGF2R-03	rs1803989	T= 0.09746 C= 0.90254	T/T = 10 (1.1) T/C = 164 (17.4) C/C = 770 (81.6)	235.262271 227.875231 221.685117	T/T = 0.5663 T/C = 0.3361 C/C = -	0.2725	4448.622277 4355.191482 4319.448456	T/T = 0.6808 T/C = 0.6746 C/C = -	0.5820	0.198773 0.197948 0.192863	T/T = 0.7385 T/C = 0.2909 C/C = -	0.2833
IGFALS	1	IGFALS-05	rs9282731	T= 0.00105 C= 0.99895	T/T = 0 (0) T/C = 2 (0.2) C/C = 946 (99.8)	- 161.921568 223.381716	T/T = - T/C = 0.2442 C/C = -	0.2442	- 2716.928657 4337.313740	T/T = - T/C = **0.0205** C/C = -	**0.0205**	- 0.240761 0.193711	T/T = - T/C =0.2327 C/C = -	0.2327
	2	IGFALS-01	rs17559	T= 0.09057 C= 0.90943	T/T = 13 (1.4) T/C = 145 (15.4) C/C = 786 (83.3)	210.471319 225.943286 222.570006	T/T = 0.5598 T/C = 0.6156 C/C = -	0.9062	4328.093414 4477.276578 4306.494200	T/T = 0.9376 T/C = 0.0567 C/C = -	0.1040	0.181530 0.191264 0.194119	T/T = 0.4144 T/C = 0.5687 C/C = -	0.3704
	3	IGFALS-02	rs3751893	T= 0.21186 C= 0.78814	T/T = 41 (4.3) T/C = 318 (33.7) C/C = 585 (62.0)	203.772346 222.070321 225.263215	T/T = 0.0740 T/C = 0.5389 C/C = -	0.1339	4107.027517 4292.543092 4373.803696	T/T = 0.0928 T/C = 0.2367 C/C = -	0.0648	0.186262 0.194250 0.193986	T/T = 0.3915 T/C = 0.9460 C/C = -	0.6467
IGFBP3	1	IGFBP3-05	rs9282734	C= 0.00317 A= 0.99683	C/C = 0 (0) C/A = 6 (0.6) A/A = 941 (99.4)	- 222.807859 223.276054	C/C = - C/A = 0.9878 A/A = -	0.9878	- 4203.965916 4335.411104	C/C = - C/A = 0.7450 A/A = -	0.7450	- 0.197841 0.193751	C/C = - C/A = 0.8578 A/A = -	0.8578
	2	IGFBP3-04	rs2471551	G= 0.17766 C= 0.82234	G/G = 27 (2.9) G/C = 280 (29.8) C/C = 633 (67.3)	216.143027 218.087485 225.290873	G/G = 0.5319 G/C = 0.1789 C/C = -	0.1633	4078.69610 4184.61063 4403.86764	G/G = 0.0989 G/C = **0.0020** C/C = -	**0.0009**	0.200480 0.197045 0.191992	G/G = 0.4464 G/C = 0.2076 C/C = -	0.1608

Gene	Order of occurence on chr.	Gene ID	SNP ID	Allele 1 frequency Allele 2 frequency	A1/A1 N= (%) A1/A2 N= (%) A2/A2 N= (%)	PDEN LSMEANS	P-value for genotype association with PDEN	PDEN trend p-value	ABDEN LSMEANS	P-value for genotype association with ABDEN	ABDEN trend p-value
IGF1	1	IGF1-02	rs6220	T= 0.66703 C= 0.33297	C/C = 104 (11.2) T/C = 410 (44.2) T/T = 414 (44.6)	8.420803 9.897281 9.818354	C/C = 0.0856 T/C = 0.8784 T/T = -	0.2251	14.816061 17.041429 16.792477	C/C = 0.1314 T/C = 0.7647 T/T = -	0.3236
	2	IGF1-04	rs2162679	G= 0.16772 A= 0.83228	G/G = 22 (2.3) G/A = 275 (28.9) A/A = 654 (68.8)	9.387797 9.627498 9.711560	G/G = 0.8405 G/A = 0.8749 A/A = -	0.8187	17.206517 16.708536 16.570944	G/G = 0.8062 G/A = 0.8728 A/A = -	0.8001
IGF2	1	IGF2-02	rs734351	T= 0.62393 C= 0.37607	C/C = 129 (13.8) T/C = 443 (47.5) T/T = 360 (38.6)	9.435882 10.028727 9.335732	C/C = 0.8957 T/C = 0.1885 T/T = -	0.5498	16.130520 17.140852 16.203810	C/C = 0.9527 T/C = 0.2709 T/T = -	0.7140
	2	IGF2-03	rs3213216	G= 0.61645 A= 0.38355	A/A = 131 (14.0) G/A = 456 (48.7) G/G = 349 (37.3)	10.263199 9.569421 9.616533	A/A = 0.3972 G/A = 0.9291 G/G = -	0.5193	17.457489 16.350082 16.585875	A/A = 0.4780 G/A = 0.7820 G/G = -	0.6407
IGF1R	1	IGF1R-05	rs2137680	G= 0.70768 A= 0.29232	A/A = 92 (9.7) G/A = 372 (39.1) G/G = 487 (51.2)	9.646736 9.548271 9.850173	A/A =0.8094 G/A =0.5539 G/G = -	0.6278	16.435828 16.245701 17.021855	A/A = 0.6657 G/A = 0.3442 G/G = -	0.4189
	2	IGF1R-18	rs2175795	G= 0.70952 A= 0.29048	A/A = 90 (9.5) G/A = 369 (39.0) G/G = 486 (51.4)	9.692912 9.494576 9.831203	A/A = 0.8711 G/A = 0.5107 G/G = -	0.6421	16.568264 16.163719 17.005126	A/A = 0.7501 G/A = 0.3073 G/G = -	0.4416
	3	IGF1R-06	rs907806	G= 0.09201 A= 0.90799	G/G = 15 (1.6) G/A = 145 (15.2) A/A = 791 (83.2)	9.783674 9.906908 9.682057	G/G = 0.9581 G/A = 0.7375 A/A = -	0.7633	14.736675 17.186902 16.635540	G/G = 0.5423 G/A = 0.6098 A/A = -	0.9319
	4	IGF1R-04	rs3743258	G= 0.72569 A= 0.27431	A/A = 71 (7.5) G/A = 377 (39.9) G/G = 498 (52.6)	10.438523 10.090925 9.397717	A/A = 0.2681 G/A = 0.1706 G/G = -	0.1170	17.438157 17.270490 16.217141	A/A = 0.4203 G/A = 0.1965 G/G = -	0.1878
	5	IGF1R-26	rs3743259	G= 0.30765 A= 0.69235	G/G = 86 (9.1) G/A = 407 (43.3) A/A = 448 (47.6)	10.805789 10.000545 9.207663	G/G = 0.0671 G/A = 0.1186 A/A = -	**0.0328**	18.254669 17.206038 15.877919	G/G = 0.0914 G/A = 0.1052 A/A = -	**0.0389**
	6	IGF1R-03	rs2272037	C= 0.57354 T= 0.42646	T/T = 171 (18.1) C/T = 464 (49.1) C/C = 310 (32.8)	9.194752 10.112010 9.320621	T/T = 0.8591 C/T = 0.1478 C/C = -	0.8562	16.207064 17.219119 15.994279	T/T = 0.8524 C/T = 0.1648 C/C = -	0.6281
	7	IGF1R-01	rs2229765	G= 0.56019 A= 0.43981	A/A = 184 (19.4) G/A = 465 (49.1) G/G = 298 (31.5)	9.314492 9.276820 10.663908	A/A = 0.0519 G/A = **0.0117** G/G = -	**0.0265**	15.914559 16.148680 17.957342	A/A = 0.0682 G/A = **0.0415** G/G = -	**0.0434**
	8	IGF1R-07	rs2016347	T= 0.52784 G= 0.47216	G/G = 202 (21.2) T/G = 495 (52.0) T/T = 255 (26.8)	8.971330 9.503061 10.718326	G/G = **0.0121** T/G = **0.0332** T/T = -	**0.0100**	15.435012 16.446818 18.091805	G/G = **0.0179** T/G = **0.0738** T/T = -	**0.0160**

Gene	Order of occurence on chr.	Gene ID	SNP ID	Allele 1 frequency Allele 2 frequency	A1/A1 N= (%) A1/A2 N= (%) A2/A2 N= (%)	PDEN LSMEANS	P-value for genotype association with PDEN	PDEN trend p-value	ABDEN LSMEANS	P-value for genotype association with ABDEN	ABDEN trend p-value
IGF2R	1	IGF2R-05	rs1570070	A= 0.63097 G= 0.36903	G/G = 132 (14.0) A/G = 432 (45.8) A/A = 379 (40.2)	9.486512 9.598251 9.806614	G/G = 0.6701 A/G = 0.6909 A/A = -	0.6241	16.415877 16.498150 16.788923	G/G = 0.7581 A/G = 0.7307 A/A = -	0.7071
	2	IGF2R-01	rs894817	G= 0.68873 A= 0.31127	A/A = 97 (10.6) G/A = 375 (41.0) G/G = 442 (48.4)	10.486480 9.683566 9.664397	A/A = 0.3249 G/A = 0.9708 G/G = -	0.4544	17.931381 16.726862 16.611531	A/A = 0.3278 G/A = 0.8914 G/G = -	0.4280
	3	IGF2R-02	rs998075	G= 0.49840 A= 0.50160	G/G = 235 (25) G/A = 467 (49.7) A/A = 238 (25.3)	9.568132 10.120840 9.069487	G/G = 0.4643 G/A = 0.0751 A/A = -	0.4590	16.073387 17.442138 15.792463	G/G = 0.7983 G/A = 0.0834 A/A = -	0.7901
	4	IGF2R-11	rs998074	C= 0.50211 T= 0.49789	T/T = 235 (24.8) C/T = 474 (50) C/C = 239 (25.2)	9.046190 10.153476 9.429398	T/T =0.5741 C/T = 0.2195 C/C = -	0.5808	15.725245 17.509541 15.950126	T/T =0.8382 C/T = 0.1018 C/C = -	0.8465
	5	IGF2R-04	rs629849	G= 0.86456 A= 0.13544	A/A = 21 (2.2) G/A = 211 (22.6) G/G = 702 (75.2)	11.711571 8.813396 9.841260	A/A = 0.2581 G/A = 0.0782 G/G = -	0.4153	19.218450 15.173202 16.951443	A/A = 0.3951 G/A = 0.0587 G/G = -	0.2858
	6	IGF2R-07	rs2282140	C= 0.89504 T= 0.10496	T/T = 6 (0.6) C/T = 187 (19.7) C/C = 755 (79.6)	10.063397 9.800851 9.693318	T/T = 0.9039 C/T = 0.8605 C/C = -	0.8377	17.232055 17.192037 16.578049	T/T = 0.8947 C/T = 0.5338 C/C = -	0.5332
	7	IGF2R-03	rs1803989	T= 0.09746 C= 0.90254	T/T = 10 (1.1) T/C = 164 (17.4) C/C = 770 (81.6)	8.516442 10.292788 9.572028	T/T = 0.6551 T/C = 0.2593 C/C = -	0.4296	13.488473 17.881770 16.437358	T/T =0.4388 T/C = 0.1609 C/C = -	0.3764
IGFALS	1	IGFALS-05	rs9282731	T= 0.00105 C= 0.99895	T/T = 0 (0) T/C = 2 (0.2) C/C = 946 (99.8)	- 6.913271 9.694368	T/T = - T/C = 0.5970 C/C = -	0.5970	- 14.898518 16.654675	T/T = - T/C = 0.8360 C/C = -	0.8360
	2	IGFALS-01	rs17559	T= 0.09057 C= 0.90943	T/T = 13 (1.4) T/C = 145 (15.4) C/C = 786 (83.3)	9.756571 9.099188 9.812087	T/T = 0.9787 T/C = 0.2891 C/C = -	0.3698	16.824194 16.012004 16.804497	T/T = 0.9953 T/C = 0.4651 C/C = -	0.5448
	3	IGFALS-02	rs3751893	T= 0.21186 C= 0.78814	T/T = 41 (4.3) T/C = 318 (33.7) C/C = 585 (62.0)	9.819025 9.880650 9.618368	T/T = 0.8674 T/C = 0.6132 C/C = -	0.6430	16.821725 16.858507 16.601189	T/T = 0.9094 T/C = 0.7584 C/C = -	0.7724
IGFBP3	1	IGFBP3-05	rs9282734	C= 0.00317 A= 0.99683	C/C = 0 (0) C/A = 6 (0.6) A/A = 941 (99.4)	- 9.212126 9.686634	C/C = - C/A = 0.8760 A/A = -	0.8760	- 17.326668 16.652327	C/C = - C/A = 0.8906 A/A = -	0.8906
	2	IGFBP3-04	rs2471551	G= 0.17766 C= 0.82234	G/G = 27 (2.9) G/C = 280 (29.8) C/C = 633 (67.3)	10.027088 9.345877 9.790371	G/G = 0.8711 G/C = 0.4040 C/C = -	0.5667	16.033637 15.850826 16.958474	G/G = 0.6939 G/C = 0.1969 C/C = -	0.2192

II

Paper II

Vilde D. Haakensen, Margarethe Biong, Ole Christian Lingjærde, Marit Muri Holmen, Jan Ole Frantzen, Ying Chen, Dina Navjord, Linda Romundstad, Torben Lüders, Ida K. Bukholm, Hiroko K Solvang, Vessela N. Kristensen, Giske Ursin, Anne-Lise Børresen-Dale, Åslaug Helland,

Expression levels of uridine 5'diphosphoglucoronosyltransferase genes in breast tissue from healthy women are associated with mammographic density.

Breast Cancer Research 2010 ;12(4):R65

Breast Cancer RESEARCH

RESEARCH ARTICLE Open Access

Expression levels of uridine 5′-diphospho-glucuronosyltransferase genes in breast tissue from healthy women are associated with mammographic density

Vilde D Haakensen[1,2], Margarethe Biong[1], Ole Christian Lingjærde[3,4], Marit Muri Holmen[5], Jan Ole Frantzen[6], Ying Chen[7], Dina Navjord[8], Linda Romundstad[9], Torben Lüders[10], Ida K Bukholm[2,11], Hiroko K Solvang[1], Vessela N Kristensen[1,2,10], Giske Ursin[12,13], Anne-Lise Børresen-Dale[1,2], Åslaug Helland[1,2,14*]

Abstract

Introduction: Mammographic density (MD), as assessed from film screen mammograms, is determined by the relative content of adipose, connective and epithelial tissue in the female breast. In epidemiological studies, a high percentage of MD confers a four to six fold risk elevation of developing breast cancer, even after adjustment for other known breast cancer risk factors. However, the biologic correlates of density are little known.

Methods: Gene expression analysis using whole genome arrays was performed on breast biopsies from 143 women; 79 women with no malignancy (healthy women) and 64 newly diagnosed breast cancer patients, both included from mammographic centres. Percent MD was determined using a previously validated, computerized method on scanned mammograms. Significance analysis of microarrays (SAM) was performed to identify genes influencing MD and a linear regression model was used to assess the independent contribution from different variables to MD.

Results: SAM-analysis identified 24 genes differentially expressed between samples from breasts with high and low MD. These genes included three uridine 5′-diphospho-glucuronosyltransferase (*UGT*) genes and the oestrogen receptor gene (*ESR1*). These genes were down-regulated in samples with high MD compared to those with low MD. The UGT gene products, which are known to inactivate oestrogen metabolites, were also down-regulated in tumour samples compared to samples from healthy individuals. Several single nucleotide polymorphisms (SNPs) in the *UGT* genes associated with the expression of *UGT* and other genes in their vicinity were identified.

Conclusions: Three UGT enzymes were lower expressed both in breast tissue biopsies from healthy women with high MD and in biopsies from newly diagnosed breast cancers. The association was strongest amongst young women and women using hormonal therapy. UGT2B10 predicts MD independently of age, hormone therapy and parity. Our results indicate that down-regulation of *UGT* genes in women exposed to female sex hormones is associated with high MD and might increase the risk of breast cancer.

* Correspondence: Aslaug.Helland@rr-research.no
[1]Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Montebello, NO-0310, Norway
Full list of author information is available at the end of the article

Introduction

Breast cancer is a common disease in women. Knowledge about the first steps in tumour initiation is important for early detection. However, the exact mechanisms of tumour initiation are still unknown.

Mammographic density (MD), captured on film screen mammograms, refers to the content and architectural structure of the adipose, connective and epithelial tissues in the female breast [1]. In epidemiological studies, a high percentage of MD confers a four to six fold elevated risk of developing breast cancer [1-3] and has been proposed as a possible surrogate marker for the disease [4]. The relative risk associated with MDs remains at this magnitude even after adjustment for all other known breast cancer risk factors. Breasts with high MD have greater tissue cellularity and more tissue collagen [5]. Still, little is known as to how MD confers the increased breast cancer risk. MD is to a large degree an inherited trait, although it is also influenced by environmental factors, hormone therapy being an evident example [6]. The genetic factors determining the inheritability are largely unknown.

In order to elucidate how MD increases the risk of breast cancer; we searched for the biological correlates to MD. Gene expression analysis on biopsies from breasts of healthy women with varying degrees of MD was performed. The gene expression profiles represent the gene activity of the different cell types in the biopsy, producing a fingerprint of the breast tissue within the biopsy of that particular woman.

The breast is an oestrogen-sensitive organ. MD varies with levels of female hormones, and is reduced after menopause. The uridine 5'-diphospho-glucuronosyltransferase (*UGT*) genes encode enzymes inactivating several endogenous and exogenous compounds, including sex hormones (Figure 1) [7]. UGT1A1 is known to be responsible for the glucuronidation of bilirubin, but is also shown to glucuronidate catechol oestrogens [8,9]. Polymorphisms in this gene have previously been linked to MD in premenopausal women [10]. UGT2B7 is known to conjugate oestrone, one of the active oestradiol metabolites. This enzyme has previously been found to be down-regulated in tumour tissue compared with non-malignant tissue, leading to the conclusion that UGT expression could lead to the promotion of carcinogenesis [11] but there are no reports on this gene in relation to MD in the literature. Less is known about the other *UGT2B* genes, although there is extensive structural homology. We will use the *UGT* genes as a term describing three *UGT2B* genes significantly down-regulated in our analyses (*UGT2B7*, *UGT2B10* and *UGT2B11*). Other *UGT* genes are specified in the text. In this study we analysed biopsies from breasts of healthy women and found genes whose expression is associated with MD.

Materials and methods

Subjects

The women included in this study had all attended one of six breast diagnostic centres in Norway that are part of the governmentally funded National Breast Cancer Screening Program between 2002 and 2007 [12]. Women were eligible if they did not currently use

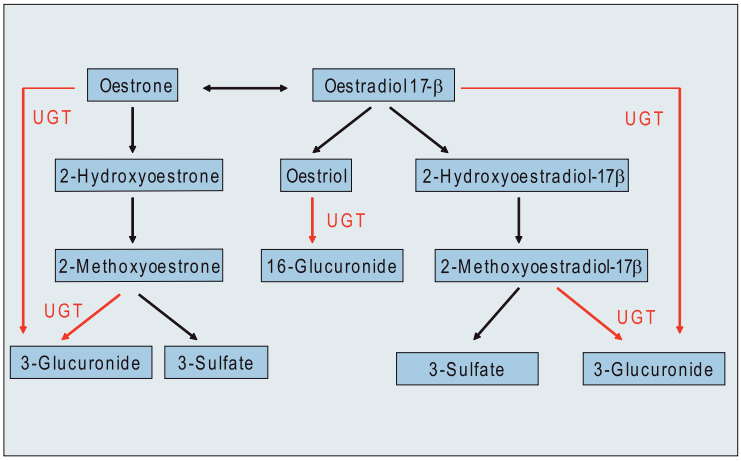


Figure 1 UGTs conjugate oestrogen-substrates into biologically inactive oestrogen glucuronides. The figure gives a schematic view with focus on glucuronidation and not a complete picture of oestradiol metabolism. Androgens are also inactivated by uridine 5'-diphospho-glucuronosyltransferases (UGTs), but are not included in this illustration.

anticoagulants, did not have breast implants and were not currently pregnant or lactating. A total of 186 women were recruited to the study; 120 healthy women with no malignant disease but some visible density in the mammograms, referred to here as healthy women, and 66 women with a newly diagnosed breast cancer. Of these, quality tested expression data were obtained from biopsies from 79 healthy women and 64 breast cancer patients.

The women were either referred to a breast diagnostic centre for a second look due to some irregularity of the initial screening mammogram (n = 69) or due to clinical findings (n = 83). For 34 women the type of referral was unknown.

The women provided information about height, weight, parity, hormone therapy use and family history of breast cancer. Two breast biopsies and three blood samples were collected from each woman. All women provided signed informed consent. The study was approved by the local ethical committee and local authorities (IRB approval no S-02036).

Core biopsies

Two breast biopsies were obtained from each woman with a 14 gauge needle, for RNA- and DNA-extraction. In healthy women, the biopsies were taken from an area with no visible pathology, but with some MD to ensure that the biopsies did not contain only fatty tissue, which yields little RNA. The sampling was guided by ultrasound. At one hospital, six of the biopsies from breasts of healthy women were collected from a benign lesion (mostly fibroadenomas). For the cancer patients, all biopsies were taken from the tumour. The tissue was either fresh-frozen at -80°C or soaked in ethanol and RNAlater (Ambion, Austin, TX, USA), transported and subsequently stored at -20°C.

Pathology

The haematoxylin eosinofil sections from the tumours of the breast cancer patients were evaluated locally and then re-evaluated by one pathologist (YC). Information about tumour size, histological grade and type, oestrogen and progesterone receptor status, human epidermal growth factor receptor (HER) 2 status and sentinel node status was recorded and entered into a database managed by the Office for Clinical Research at Oslo University Hospital, Radiumhospitalet. Pathology evaluations were not available for the biopsies from breasts of healthy women.

RNA-expression analysis

Homogenisation, cell lysis and RNA extraction were performed using the RNeasy Mini Protocol (Qiagen, Valencia, CA, USA). RNA quality was controlled by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and concentration was determined using NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). A total of 40 samples, mostly from normal breast tissue, were excluded from further analyses due to a low RNA amount (< 10 ng) or poor RNA quality. RNA was then amplified and labelled using the Agilent Low RNA input Fluorescent Linear Amplification Kit Protocol. Amplified tumour RNA was labelled by Cy5 (Amersham Biosciences, Little Chalfont, England, UK) and amplified RNA from Universal Human total RNA (Stratagene, La Jolla, CA, USA) was labelled by Cy3 (Amersham Biosciences, Little Chalfont, England, UK). RNA from the remaining 146 biopsies was further hybridised on Agilent Human Whole Genome Oligo Microarrays (G4110A) (Agilent Technologies, Santa Clara, CA, USA). Three arrays had to be excluded due to poor quality leaving data from 143 subjects (79 healthy individuals and 64 breast cancer patients) for further analysis. Of the 79 biopsies from healthy women, 5 had been obtained from a benign lesion. By ultrasound and mammography these 5 were described as fibroadenoma (n = 4) or microcalcification (n = 1).

RNA-data processing

The microarrays were scanned by an Agilent scanner (Agilent Technologies, Santa Clara, CA, USA) and processed in Feature Extraction 9.1.3.1 (Agilent Technologies, Santa Clara, CA, USA). Locally weighted scatterplot smoothing (lowess) was used to normalise the data. The normalised and log2-transformed data were stored in the Stanford Microarray Database [13] and retrieved from the database for further statistical analyses. Flagged spots were treated as missing values. The dataset now counted 40,791 probes. Clone IDs with 20% or more missing values were excluded. Gene filtering was performed to include only probes with variation across samples, so that probes with less than three arrays being at least 1.6 standard deviations from the mean were excluded. For the 79 healthy women, this probe filtration resulted in an expression dataset of 9,767 probes and 79 arrays each representing one individual. For the breast cancer women, a dataset of 64 arrays and 10,153 probes were obtained after filtration, and for both groups combined, a dataset of 143 arrays and 13,699 probes were obtained. Missing values were imputed in R using the method impute.knn in the library impute [14].

Genotyping

Blood DNA was extracted by phenol/chloroform extraction followed by ethanol precipitation (Nuclear Acid Extractor 340A; Applied Biosystems, Foster City, CA,

USA) according to standard procedures. *UGT* genotype data was retrieved from two sources: genome wide association studies (GWAS) using the Human-1 109K BeadChip (Illumina Inc, San Diego, CA, USA) and candidate gene-based study using iPlex, Sequenom. For the GWAS, each sample was subject to whole genome amplification using Illumina proprietary reagents [15]. The amplified DNA was fragmented and hybridised according to the protocol. The BeadArray reader (Illumina Inc, San Diego, CA, USA) with the BeadScan software (Illumina Inc, San Diego, CA, USA) was used to image the beadchips. Non-polymorphic probes and probes with more than 20% missing values and were excluded and data processed as described previously [16]. The candidate gene single nucleotide polymorphism (SNP) analyses were performed using the iPLEX assay in conjunction with the Sequenom MassARRAY platform. Multiplexing was performed in 384 plates using 1 ul DNA per well with one well containing up to 29 reactions. The technology is described in detail on the sequenom web-page [17].

Mammograms

Routine descriptions of mammograms by local radiologists were collected. Craniocaudal mammograms of both breasts were digitised using a high-resolution Kodak Lumisys 85 scanner (Kodak, Rochester, NY, USA). Density was quantified using the University of Southern California Madena assessment method [18]. In brief, the method works as follows: a reader (trained by GU) outlines the total area of the breast using a computerised tool, the software then counts the number of pixels. This represents the total breast area. MD is assessed (by GU), first by identifying a region of interest that incorporates all dense areas except those representing the pectoralis muscle and scanning artifacts, and then by applying a yellow tint to all pixels within the region of interest shaded at or above a threshold intensity of gray. The software then counts the tinted pixels, which represents the area of absolute density. The percent density is the absolute density area divided by the total breast area and is the value used for these analyses. Test-retest reliability was 0.99 for absolute density.

Statistical analysis

Clustering was performed using MatLab (version R2007b) (The MathWorks Inc., Natick, MA, USA) with Ward linkage and Euclidean distances. Before clustering, the data were gene centred, that is, for every probe the mean expression across all samples was calculated and was subtracted from the log2-ratios for that gene. This was performed for visualisation purposes only, clustering with uncentred data returns the same clusters. Significance analysis of microarrays (SAM, Stanford University, CA, USA) (version 3.02) [19,20] for Excel (Microsoft, Redmond, WA, USA) was used for analysis of differentially expressed genes between two groups of data. The data were not gene centred for the SAM analysis. A total of 500 permutations were used. Quantitative SAM analysis was used to identify genes differentially expressed according to MD as a continuous variable. Statistical significance tests and regression analysis were performed in R 2.9.0 [21]. To test for difference in the mean of phenotypic variables (MD, age, body mass index (BMI)) in different clusters of women, we used two-sided t-tests (assuming equal variance in the groups) and analysis of variance (ANOVA) for continuous variables and chi-squared/Fisher's exact tests for categorical variables [22]. To investigate the similarities of distributions of *UGT* genes between tumour samples and normal samples with low MD and high MD respectively, Kullback-Leibler distances between normalised distributions of the histograms of the data were calculated by use of MatLab (The MathWorks Inc., Natick, MA, USA). The cancer samples in our study were grouped into subtypes and assigned a risk group using the PAM50 gene list published by Parker et al [23]. SNP-analysis was performed using R 2.9.0 [21]. The association between gene expression and SNPs was assessed using expression quantitative trait loci (eQTL) [24]*in cis* (10^6 bp on each side of the gene) using the R package eMap v1.1 [25]. Comparing the akaike information criterion for different models predicting MD, the lower criterion singled out a linear regression model as the model fitting the distribution of the data best. A linear regression model was fitted in R 2.9.0 with MD as a continuous response variable and the following covariates: UGT2B7, two probes for UGT2B10, UGT2B11, ESR1, age, BMI, current hormone therapy, age at first birth and parity. Gene expression, age, age at first birth and BMI were entered into the model as continuous variables. Stepwise variable selection was performed, starting with all variables included in the model. For every step, the variable with the highest P value was rejected from the model and the model was refitted. This was repeated until all variables included in the model had a P value less than 0.05. To correct for the influence of age, this variable was forced to stay in the model. A sensitivity analysis was performed excluding extreme ages (30 years or younger) to check the robustness of the data. We also fitted linear regression stratified on age (younger or older than 50 years of age) and current use of hormone therapy. Gene ontology analysis was performed by the use of DAVID Bioinformatics Resources 2008 from the National Institute of Allergy and Infectious Diseases, NIH [26]. Functional annotation clustering was applied and the following gene ontology categories were selected: biological

processes (all), molecular function (all) and the KEGG pathway database. We included gene ontology terms with a *P* value (false discovery rate (FDR)-corrected) of less than 0.01 containing between 5 and 500 genes.

The normalised, log2-transformed data are available in Gene Expression Omnibus with accession number [GEO:GSE18672]. The data are not gene centered or gene filtered.

Results

Gene expression and mammographic density

To identify genes differentially expressed according to MD we performed quantitative SAM with MD as a continuous variable using gene expression data from the normal biopsies. Of 9,767 probes, only 25 probes, representing 24 genes, were differentially expressed according to MD, with reduced expression associated with higher MD (FDR < 25%; Table 1) [see Additional file 1]. Gene ontology analysis revealed no significant terms and we found no pathway associated with this gene set. The *UGT* genes and oestrogen receptor gene (*ESR1*) were among the genes significantly down-regulated in breasts with high MD. The percentage of samples with low UGT expression was higher in tumour samples than in normal samples with low MD, whereas the percentage was more similar between tumour samples and normal samples with high MD [see Figure S1 in Additional file 2]. The function of UGT-enzymes in oestradiol metabolism is illustrated in Figure 1. In healthy women, the expression of the different *UGT* genes was highly correlated with each other and the four probes clustered together [see Figures S2 and S3 and Table S1 in Additional file 2].

Table 1 Genes differentially expressed according to mammographic density in non-cancer samples

Gene symbol	Agilent ID	Cytogenetic band
729641	A_24_P932736	8p21.1
FLJ10404	A_23_P427472	5q35.3
VPS18	A_24_P18802	15q15.1
UGT2B10	A_23_P7342	4q13.2
CABP7	A_24_P177236	22q12.2
CD86	A_24_P131589	3q13.33
UGT2B11	A_23_P212968	4q13.2
580687	A_23_P152570	17p11.2
DIAPH2::RPA4	A_23_P254212	Xq21.33
LMOD1	A_32_P199824	1q32.1
UGT2B10	A_24_P521559	4q13.2
PIK3R5	A_23_P66543	17p13.1
ATG7	A_32_P107994	3p25.2
LRRC2	A_23_P155463	3p21.31
RBL1	A_23_P28733	20q11.23
NPY1R	A_23_P69699	4q32.2
810781	A_23_P144244	3q13.33
593535	A_32_P80016	15q26.1
H2AFJ	A_23_P204277	12p12.3
666399	A_32_P35668	20p12.3
Transcribed	A_24_P640617	2p25.2
Transcribed	A_32_P20997	20q13.13
UGT2B7	A_23_P136671	4q13
ESR1	A_23_P309739	6q25.1
SAPS1	A_23_P119448	19q13.42

MD was lower in women with BMI of 25 or more compared with those with BMI of less than 25 ($P = 0.01$), but unrelated to other epidemiological variables. UGT expression was not significantly associated with age, BMI, age at first birth or current hormone therapy use in the healthy women [see Table S2 in Additional file 2].

To dissect the impact of age and hormone therapy use, we performed SAM analyses to identify differentially expressed genes according to MD, whereas stratifying for age and postmenopausal hormone therapy use. For healthy women younger than 50 years of age, the *UGT* genes were not significant at a FDR of 25%. For healthy women aged 50 years or older, 49 probes were significantly down-regulated in breasts with MD of 30% or higher (FDR < 25%). Of these, 17 were overlapping with those significantly down-regulated among healthy women in the unstratified analysis. The *UGT* genes were not in this list. We then stratified the women aged 50 years or older on current hormone therapy use. When only those currently using hormone therapy were included in the analysis, *UGT2B7* and *UGT2B11* were among the six genes differentially expressed with an FDR less than 10E-5 and *UGT2B28* with FDR less than 25%. For healthy women above 50 years of age and not currently using hormone therapy, several of the 24 genes were differentially expressed according to MD with an FDR of less than 25%, but again the *UGT* genes were not in this list [see Additional file 3].

These analyses were confirmed fitting a linear regression model. Although the other variables were excluded from the model with insignificant *P* values, age was kept in the model to control for the age-effect. After stepwise variable selection, the only significant variables remaining in the model were UGT2B10 (A_23_P7342)($P = 0.005$) and BMI ($P = 0.015$). Sensitivity analysis excluding extreme ages (30 years and younger) did not alter the results (UGT2B10 $P = 0.003$, BMI $P = 0.016$) and indicates the robustness of the results. ESR1 was borderline significant in both these analyses. These results were not significantly altered when MD was log2-transformed. For further stratification see Table 2.

Unsupervised hierarchical clustering of the 79 samples from healthy women showed two main clusters. MD was not significantly different between these two clusters [see Figure S3 in Additional file 2].

Table 2 Linear regression analysis of factors predicting mammographic density in all women and stratified for age and hormone therapy use

Women in model	N	Variables	Beta value	*P* value
All women	76	UGT2B10[1)]	-0.6	0.902
		UGT2B7	1.8	0.631
		UGT2B11	4.8	0.275
		ESR1	-3.8	0.055
		UGT2B10[2)]	**-5.6**	**0.005**
		BMI	**-1.5**	**0.015**
		age	-0.4	0.074
50 years or older	46	UGT2B11	0.2	0.987
		UGT2B10[1)]	1.0	0.946
		UGT2B7	3.5	0.486
		UGT2B10[2)]	-3.7	0.073
		BMI	-1.4	0.052
		ESR1	**-6.0**	**0.016**
		age	-0.9	0.061
50 years or older, currently on hormone therapy	11	UGT2B10[1)]	7.2	0.771
		UGT2B11	-5.8	0.695
		BMI	-2.9	0.103
		UGT2B7	6.8	0.418
		UGT2B10[2)]	**-27.0**	**0.000**
		ESR1	**-8.1**	**0.011**
		age	-0.9	0.103
50 years or older, never used hormone therapy	28	UGT2B11	-0.7	0.948
		UGT2B10[1)]	3.3	0.809
		UGT2B7	3.1	0.555
		UGT2B10[2)]	-1.4	0.607
		BMI	-0.9	0.348
		ESR1	**-6.0**	**0.033**
		Age	**-1.5**	**0.004**
Younger than 50 years	30	UGT2B7	0.4	0.950
		UGT2B10[1)]	-1.2	0.866
		ESR1	-0.9	0.835
		UGT2B11	8.4	0.225
		BMI	-1.4	0.216
		UGT2B10[2)]	**-6.2**	**0.040**
		Age	-0.3	0.610

1) A_24_P521559, 2) A_23_P7342

Factors predicting mammographic density (MD) after stepwise exclusion of non-significant factors are shown. Variables listed in the order of exclusion from the model. *P* value from the last equation including the variable is shown. Age is forced to stay in the model. UGT2B10 (A_23_P7342) is a significant, independent predictor of MD in all analyses with a majority of women under influence of female hormones; women younger than 50 years of age and women currently on hormone therapy. BMI, body mass index.

In the breast cancer group, MD was significantly associated with age and BMI, with higher MD in the younger women and in those with BMI less than 25. Both MD and UGT expression tended to be higher in women with receptor positive tumours, but this was not significant for any type of receptor. UGT-expression in tumours was unrelated to age, BMI, age at first birth and current hormone therapy (data not shown). There was a higher proportion of oestrogen receptor positive tumours among the breast cancer patients with high MD (≥ 30%) compared with low (< 30%) MD (10 of 10 vs 36 of 40, Fisher's = 0.001). There was no significant association between tumour subtype and level of MD as assessed by ANOVA. There was no indication that degree of MD was associated with the risk of relapse as assessed by the method of Parker et al [23] [see Figure S4 of Additional file 2].

Nine probes were differentially expressed according to MD in cancer samples (FDR < 25%; Table 3). None of these were overlapping with the 24 genes differentially

Table 3 Genes differentially expressed according to mammographic density in cancer samples

Agilent ID	Gene name	FDR (%)
A_32_P171923	730402	0.00
A_32_P480177	TNN	0.00
A_23_P200298	AGL	0.00
A_24_P87036	TMEM16A	0.00
A_23_P312150	EDN2	14.87
A_23_P83388	EPPK1	14.87
A_32_P60065	F2RL2	19.82
A_32_P158272	MRNA	19.82
A_23_P105012	HRASLS2	19.82

FDR, false discovery rate.

expressed in the samples from the breasts of healthy women.

Genetic polymorphisms

In order to identify genetic determinants of the expression of the *UGT* genes found to be associated with MD, we performed eQTL analyses of SNPs in these genes as available from an array based GWAS study and a candidate gene study. Twenty one SNPs in *UGT* genes were present on the 109 K array from Illumina, and 9 SNPs from the candidate gene analysis. Of these, 5 SNPs were associated with the expression of *UGT* genes or other genes in their vicinity at $P = 0.05$ [see Additional file 4]. Two of these SNPs, both located in *UGT2B10* (rs1828705, rs1828705), were significantly associated with gene expression of another *UGT* gene (*UGT2B7* and *UGT2B28*).

Discussion

Previously, whole genome expression profiling of normal breast tissue (all cell types included) has been performed to a limited extent [27,28]. Yang et al recently performed a study of cancer-free breast tissue obtained from mastectomies in breast cancer patients with high and low MD [29]. They identified a list of 73 genes differentially expressed between high and low MD samples. Specifically, this included the down-regulation of several transforming growth factor (TGF) β-related genes in samples with high MD. In the present study we analysed breast biopsies from 79 healthy women and tumours of 64 women with breast cancer. Twenty-four genes were differentially expressed according to MD in the healthy samples. In breast tumours, none of these 24 genes were found differentially expressed according to MD. Tumour-specific deregulation of a large number of mRNA transcripts may be expected to overshadow the MD signature. In addition, the sample size is limited and the two sample sets (cases and controls) are not directly comparable with respect to MD [see Figure S5 in Additional file 2].

In our study, three *UGT* genes (*UGT2B11, UGT2B10* and *UGT2B7*) were differentially expressed according to MD in the breasts of healthy women. All these three enzymes had decreased expression in dense breasts. Previous knowledge links the UGT enzymes to the metabolism of female hormones known to influence the mammary glands (Figure 1). The over-representation of *UGT* genes on the list of significant genes along with a biological link makes these genes particularly interesting. In a linear regression model with age as a confounding factor, BMI and one of two probes for UGT2B10 were the only significant variables independently predicting MD, with ESR1 as a borderline significant covariate. The expression of these three *UGT2B* genes is highly correlated to each other and as expected only one probe remained in the regression model as an independent predictor of MD. BMI is known to be the strongest and most consistent epidemiological predictor of MD, and is expected to remain in the model. It is noteworthy that one of the *UGT* genes has an independent predictive value of a greater significance and magnitude than BMI. MD is determined by multiple factors. In a study of limited sample size, we can only expect to identify the strongest predictors.

UGT2B7 is postulated to protect the breast tissue from oestrogen metabolites locally [30], and this is consistent with our findings that breasts with higher MD have reduced expression of this gene. The main metabolites of oestradiol and oestrone (hydroxyl- and methoxy-oestrogen compounds) bind to the oestrogen receptor, but with a reduced affinity compared with oestradiol. UGT2B10 and 11 are not yet reported to be associated with MD or breast cancer, but UGT2B10 is involved in the metabolism of tobacco-related nitrosamines [31]. Less is known about UGT2B11. The different *UGT2B* genes are located close to each other on chromosome 4 and there is great homology between the genes [see Figure S6 in Additional file 2]. UGT1A1, previously linked to MD and breast cancer [32], is not represented on the microarray used in this study.

We have identified a set of genes differentially expressed according to MD. Interestingly, the *UGT* genes seem, to a greater extent than the other genes, to be more similarly expressed between tumour samples and normal samples from breasts with high MD as compared with normal samples from breasts with low MD [see Table S4 and Figure S7 in Additional file 2]. The other differentially expressed genes generally express the same levels in the tumours and in the biopsies from the healthy women with low MD. We cannot exclude that the *UGT* genes confer risk for breast cancer development through increasing MD, but further studies would be needed to investigate this.

We found the *UGT* genes to be differentially expressed in young women and women over 50 years of age currently on hormone therapy. SAM analysis of MD in women younger than 50 years did not give any differentially expressed genes with an FDR of less than 25%. However, several UGT-probes are on the top of the list of genes down-regulated in samples from breasts with high MD. The lack of significance could be due to low sample size (n = 30). As UGT enzymes conjugate oestradiol metabolites, its effect will be greater when there is an increased level of oestradiol present, whether the oestradiol is endogenous or exogenous. The linear regression analysis showed that UGT2B10 was predicting MD independent of age in all women, younger women and women older than 50 years currently using hormones. This leads to the hypothesis that decreased UGT expression in the breast of a woman with increased levels of female hormones confers an increased MD and possibly an increased risk of breast cancer.

The biology in breasts with high and low MD may differ, partly due to differences in proportion of fatty tissue. Therefore, we looked for differentially expressed genes in a subset of samples including only samples from breasts with MD of more than 20%. The fact that the *UGT2B* gene family is so strongly represented among the down-regulated genes (six probes representing five different *UGT2B* genes are the only genes differentially expressed with an FDR < 10E-5) indicate that reduced UGT expression is of greater significance in breasts with higher MD and lower content of fatty tissue.

We find that *ESR1* is down-regulated in biopsies from healthy women with high MD compared with those with low MD. This is not consistent with previous findings [33] and contrary to what one would expect because *ESR1* induces transcription and epithelial growth and high MD may contain increased amounts of epithelial cells [34,35]. However, increased levels of oestradiol have been shown to decrease levels of ESR1 in breast cancer [36], and in normal breast tissue in monkeys [37] and in mice [38]. Increased levels of oestradiol may increase MD. Elevated expression of ESR1 is common postmenopausally [37] and represents non-proliferating cells. The association between reduced levels of ESR1 and high MD may reflect high levels of oestradiol. We found that ESR1 was only a borderline significant predictor of MD in models with stepwise exclusion of covariates. In a model including ESR1 with only age or age and UGT2B10, ESR1 was significantly predicting MD. The independent contribution of ESR1 in predicting MD was significant in older women, where the effect of UGT2B10 was not present. There could be a link between UGT-expression and ESR1-expression in that reduced metabolism of oestradiol-metabolites increases the levels of ESR1-ligands (oestradiol metabolites) and hence reduces ESR1-levels. The UGT-enzyme activity may be the cause of the alterations leading to increased MD by this mechanism. Reduced ESR1 is only borderline significant in predicting MD and could also be an intermediate factor.

MD is the result of complex biological processes without any single determining factor. BMI is the single most important factor found to date, and is also significant in this study. Age seems to have its effect mainly through hormonal influence, except for in postmenopausal women not taking hormones, where age has a significant, independent effect on MD. MD is not significantly different between the two main clusters from unsupervised hierarchical clustering of the samples from healthy women. MD is hence not related to the main variation in the normal samples.

The genes whose expression we have found to be associated with MD do have a fairly high FDR in a SAM analysis and are not significant in all stratified analyses, suggesting that they may play a role in only subsets of individuals and other factors also have a significant contribution. Despite this, in linear regression models UGT2B10 is an independent predictor of MD along with BMI.

There is a substantial heritable proportion of MD. SNPs in *UGT* genes with influence on the UGT expression have been described [8,39]. We identified two UGT-SNPs associated with the expression of other *UGT* genes. Due to their homology and co-localisation on the chromosome, they may share common control loci that affect the expression of multiple *UGT* genes. It remains to be investigated in larger and better powered epidemiological studies whether any of these SNPs are associated to MD *per se*.

We do not know enough about the variability of gene expression within normal breasts to know if the genes relevant for MD are adequately represented by one biopsy taken from an area with some MD. It is previously shown that two biopsies from the same breast tumour, before and after chemotherapy, cluster together [40]. The tumours may, however, be more homogenous than normal breast tissue. Variability in gene expression within each breast will make it difficult to detect genes with only a minor influence on MD so that only the strongest factors are identified. In an unpublished dataset we found no significant difference between UGT-expression in tumours and normal adjacent tissue tested by paired t-test [see Table S5 in Additional file 2]. This is merely an indication that the expression in one breast might be similar for different locations in the breast and hence be used to look for associations with MD.

In this study, healthy individuals had higher MD than the breast cancer patients. The women recruited in the study had been referred to a breast diagnostic centre for a second look. As high MD confers an increased risk for breast cancer and mammograms with high MD are more difficult to interpret, they most likely had a higher MD. In addition, the inclusion criterion of some visible MD for biopsy may have influenced the mean MD of the study population. The two populations are not directly comparable with respect to MD and related parameters. This lack of comparability on MD does not affect the analyses of gene expression among the healthy women only.

We obtained good quality microarrays from only 79 of 120 healthy women and from 64 of 66 breast cancer patients. This was due to low mRNA-yield or low mRNA-quality. The biopsies from healthy women consistently yielded less mRNA than the tumour samples. There is significantly higher MD in the breasts of healthy women with successful microarrays than in those with unsuccessful microarrays (37% vs 29%, P = 0.03). As samples from breasts with low MD are underrepresented in the microarray study, it is more difficult to identify genes that are differentially expressed between breast tissue with high and low MD. Despite these limitations, we have identified differentially expressed genes. These genes might have a greater significance than shown in this study.

Normal breast tissue yields less RNA than tumour tissue. The biopsies in this study were small and in agreement with the pathologist, all tissue from normal breasts was prioritised for RNA-extraction rather than histological evaluation. Imprint was not in routine use in the hospitals where we started this study. In order to make it possible for the staff to include women in this study in a busy schedule we had to use procedures already established. We do therefore not have any information about the cell types of the normal biopsies. Knowledge about the cell types present in the biopsies would have facilitated the analysis.

The two UGT2B10-probes behave differently in our dataset. Both probes map to the 3'end of the UGT2B10-gene by BLAT (98.4% homology for A_23_P7342 and 100% homology for A_24_P521559). The discrepancy in UGT2B10-expression detected by the two probes may be due to the fact that they both also share substantial sequence homology with other, but different UGT2B-genes.

Conclusions

We have identified a set of genes that are differentially expressed according to MD in breast samples from healthy women. Some of these genes are known to influence MD and breast cancer, such as *ESR1* and *UGT2B7*. Two less described *UGT* genes, *UGT2B10* and *UGT2B11*, are also differentially expressed. The expression of the three *UGT* genes is reduced in samples with high MD and also in tumour samples, but does not vary between different tumour subtypes or risk groups. The UGT enzymes are known to conjugate active oestrogen-metabolites. We show that UGT2B10 expression and BMI are independent predictors of MD. The influence of reduced UGT expression was strongest in women under exposure of female hormones. Two candidate SNPs are associated with the *UGT* gene expression *in cis*. We hypothesise that reduced expression of *UGT* genes in women exposed to female sex hormones, increase MD and that this may be associated with an increased risk of breast cancer. Further studies of these genes are needed to test the hypothesis that the gene products from these genes protect the breast from the oestrogen-induced MD and thereby reducing the risk of breast cancer.

Additional material

Additional file 1: Healthy SAM MD. Significance analysis of microarrays (SAM) for genes differentially expressed according to mammographic density (MD).

Additional file 2: Figures and tables. A collection of figures and tables describing the data set and the uridine 5'-diphospho-glucuronosyltransferase (UGT) genes. The main text refers to individual figures and tables in this file.

Additional file 3: Healthy SAM MD stratified. Significance analysis of microarrays (SAM) for genes differentially expressed according to mammographic density (MD) stratified on age and use of hormone therapy.

Additional file 4: eQTL. Expression quantitative trait loci (eQTL) analysis of single nucleotide polymorphism (SNPs) affecting the expression of uridine 5'-diphospho-glucuronosyltransferase (UGT) genes *in cis*.

Abbreviations

ANOVA: analysis of variance; BMI: body mass index; eQTL: expression quantitative trait loci; ESR1: oestrogen receptor; FDR: false discovery rate; GWAS: genome wide association studies; HER: human epidermal growth receptor; MD: mammographic density; SAM: significance analysis of microarrays; SNP: single nucleotide polymorphism; UGT: uridine 5'-diphospho-glucuronosyltransferase.

Acknowledgements

This study was funded primarily by The Research Council of Norway and South-Eastern Norway Regional Health Authority. We thank all the women who participated in the study and all the personnel in the hospitals who made the inclusion of these women possible, in particular the responsible radiologists: Einar Vigeland, Rolf O Næss and Else Berit Velken. We would also like to thank Lars Ottestad for help in the initiation of the project, Hilde Johnsen, Caroline Jevanord Frøyland and Marit Hilsen for lab assistance and to Eunjung Lee for statistical help. No medical writers were involved in this paper.

Author details

[1]Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Montebello, NO-0310, Norway. [2]Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, NO-0315,

Norway. [3]Biomedical Research Group, Department of Informatics, University of Oslo, Oslo, NO-0315, Norway. [4]Centre for Cancer Biomedicine, University of Oslo, Oslo, NO-0315, Norway. [5]Department of Radiology, Oslo University Hospital Radiumhospitalet, Oslo, NO-0310, Norway. [6]Department of Radiology, University Hospital of North Norway, Tromsø, NO- 9038, Norway. [7]Department. of Pathology, Vestfold Hospital, Halfdan Wilhelmsens Alle' 17, Tønsberg, NO-3103, Norway. [8]Department of Radiology, Innlandet Hospital, Brummundal, NO-2381, Norway. [9]Department of Radiology, Buskerud Hospital, Drammen, NO-3004, Norway. [10]Department for Clinical Molecular Biology (EpiGen), Institute for Clinical Medicine, Akershus University Hospital, University of Oslo, Oslo, NO-0315, Norway. [11]Department of Surgery, Akerhus University Hospital, Lørenskog, 1478, Norway. [12]Department of Nutrition, School of Medicine, University of Oslo, Oslo, NO-0315, Norway. [13]Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles, CA 90033, USA. [14]Department of Oncology, Oslo University Hospital Radiumhospitalet, Oslo, NO-0310, Norway.

Authors' contributions

The trial was designed by ALBD, ÅH, GU, MMH and VNK. ALBD and ÅH ensured funding. MMH, JOF, DN, LR, IKB and VDH assisted in data collection. MB and VNK are responsible for SNP analyses. VDH and TL contributed to the laboratory work. GU estimated the amount of mammographic density. OCL, VDH, HKS and MB performed statistical analyses of the data. ÅH, ALBD and VDH interpreted the results and wrote the paper. All authors were involved in reviewing the report.

Competing interests

The authors declare that they have no competing interests.

Received: 5 August 2010 Revised: 5 August 2010
Accepted: 27 August 2010 Published: 27 August 2010

References

1. Oza AM, Boyd NF: **Mammographic parenchymal patterns: a marker of breast cancer risk.** *Epidemiol Rev* 1993, **15**:196-208.
2. Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, Lockwood GA, Tritchler DL, Yaffe MJ: **Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study.** *J Natl Cancer Inst* 1995, **87**:670-675.
3. Ursin G, Ma H, Wu AH, Bernstein L, Salane M, Parisky YR, Astrahan M, Siozon CC, Pike MC: **Mammographic density and breast cancer in three ethnic groups.** *Cancer Epidemiol Biomarkers Prev* 2003, **12**:332-338.
4. Boyd NF, Lockwood GA, Martin LJ, Knight JA, Byng JW, Yaffe MJ, Tritchler DL: **Mammographic densities and breast cancer risk.** *Breast Dis* 1998, **10**:113-126.
5. Guo YP, Martin LJ, Hanna W, Banerjee D, Miller N, Fishell E, Khokha R, Boyd NF: **Growth factors and stromal matrix proteins associated with mammographic densities.** *Cancer Epidemiol Biomarkers Prev* 2001, **10**:243-248.
6. Ursin G, Lillie EO, Lee E, Cockburn M, Schork NJ, Cozen W, Parisky YR, Hamilton AS, Astrahan MA, Mack T: **The relative importance of genetics and environment on mammographic density.** *Cancer Epidemiol Biomarkers Prev* 2009, **18**:102-112.
7. Guillemette C, Levesque E, Harvey M, Bellemare J, Menard V: **UGT genomic diversity: beyond gene duplication.** *Drug Metab Rev* 2010, **42(1)**:22-42.
8. Mackenzie PI, Gregory PA, Lewinsky RH, Yasmin SN, Height T, McKinnon RA, Gardner-Stephen DA: **Polymorphic variations in the expression of the chemical detoxifying UDP glucuronosyltransferases.** *Toxicol Appl Pharmacol* 2005, **207**:77-83.
9. Cheng Z, Rios GR, King CD, Coffman BL, Green MD, Mojarrabi B, Mackenzie PI, Tephly TR: **Glucuronidation of catechol estrogens by expressed human UDP-glucuronosyltransferases (UGTs) 1A1, 1A3, and 2B7.** *Toxicol Sci* 1998, **45**:52-57.
10. Yong M, Schwartz SM, Atkinson C, Makar KW, Thomas SS, Newton KM, iello Bowles EJ, Holt VL, Leisenring WM, Lampe JW: **Associations between polymorphisms in glucuronidation and sulfation enzymes and mammographic breast density in premenopausal women in the United States.** *Cancer Epidemiol Biomarkers Prev* 2010, **19**:537-546.
11. Starlard-Davenport A, Lyn-Cook B, Radominska-Pandya A: **Identification of UDP-glucuronosyltransferase 1A10 in non-malignant and malignant human breast tissues.** *Steroids* 2008, **73**:611-620.
12. Wang H, Karesen R, Hervik A, Thoresen SO: **Mammography screening in Norway: results from the first screening round in four counties and cost-effectiveness of a modeled nationwide screening.** *Cancer Causes Control* 2001, **12**:39-45.
13. **Stanford Microarray Database.** [http://genome-www5.stanford.edu//].
14. **R library impute.knn.** [http://rss.acs.unt.edu/Rdoc/library/impute/html/impute.knn.html].
15. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS: **A genome-wide scalable SNP genotyping assay using microarray technology.** *Nat Genet* 2005, **37**:549-554.
16. Nordgard SH, Johansen FE, Alnaes GI, Bucher E, Syvanen AC, Naume B, Borresen-Dale AL, Kristensen VN: **Genome-wide analysis identifies 16q deletion associated with survival, molecular subtypes, mRNA expression, and germline haplotypes in breast cancer patients.** *Genes Chromosomes Cancer* 2008, **47**:680-696.
17. **Sequenom.** [http://www.sequenom.com/].
18. Ursin G, Astrahan MA, Salane M, Parisky YR, Pearce JG, Daniels JR, Pike MC, Spicer DV: **The detection of changes in mammographic densities.** *Cancer Epidemiol Biomarkers Prev* 1998, **7**:43-47.
19. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
20. **Significance Analysis of Microarrays.** [http://www-stat.stanford.edu/~tibs/SAM/].
21. **R project.** [http://r-project.org].
22. Altman DG: *Practical Statistics for Medical Research* Boca Raton: Chapman & Hall/CRC 1999.
23. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**:1160-1167.
24. Carlborg O, de Koning DJ, Manly KF, Chesler E, Williams RW, Haley CS: **Methodological aspects of the genetic dissection of gene expression.** *Bioinformatics* 2005, **21**:2383-2393.
25. **UNC Biostatistics Software Links.** [http://www.bios.unc.edu/~wsun/software.htm].
26. **DAVID Bioinformatics Resources 6.7.** [http://david.abcc.ncifcrf.gov/].
27. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de RM, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98**:10869-10874.
28. Nicolau M, Tibshirani R, Borresen-Dale AL, Jeffrey SS: **Disease-specific genomic analysis: identifying the signature of pathologic biology.** *Bioinformatics* 2007, **23**:957-965.
29. Yang WT, Lewis MT, Hess K, Wong H, Tsimelzon A, Karadag N, Cairo M, Wei C, Meric-Bernstam F, Brown P, Arun B, Hortobagyi GN, Sahin A, Chang JC: **Decreased TGFbeta signaling and increased COX2 expression in high risk women with increased mammographic breast density.** *Breast Cancer Res Treat* 2010, **119(2)**:305-314.
30. Guillemette C, Belanger A, Lepine J: **Metabolic inactivation of estrogens in breast tissue by UDP-glucuronosyltransferase enzymes: an overview.** *Breast Cancer Res* 2004, **6**:246-254.
31. Kaivosaari S, Toivonen P, Hesse LM, Koskinen M, Court MH, Finel M: **Nicotine glucuronidation and the human UDP-glucuronosyltransferase UGT2B10.** *Mol Pharmacol* 2007, **72**:761-768.
32. Dalhoff K, Buus JK, Enghusen PH: **Cancer and molecular biomarkers of phase 2.** *Methods Enzymol* 2005, **400**:618-627.
33. Verheus M, Maskarinec G, Erber E, Steude JS, Killeen J, Hernandez BY, Cline JM: **Mammographic density and epithelial histopathologic markers.** *BMC Cancer* 2009, **9**:182.
34. Li T, Sun L, Miller N, Nicklee T, Woo J, Hulse-Smith L, Tsao MS, Khokha R, Martin L, Boyd N: **The association of measured breast tissue characteristics with mammographic density and other risk factors for breast cancer.** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:343-349.
35. Rayter Z: **Steroid receptors in breast cancer.** *Br J Surg* 1991, **78**:528-535.
36. Dunbier AK, Anderson H, Ghazoui Z, Folkerd EJ, A'hern R, Crowder RJ, Hoog J, Smith IE, Osin P, Nerurkar A, Parker JS, Perou CM, Ellis MJ, Dowsett M: **Relationship between plasma estradiol levels and estrogen-**

responsive gene expression in estrogen receptor-positive breast cancer in postmenopausal women. *J Clin Oncol* 2010, **28**:1161-1167.

37. Cheng G, Li Y, Omoto Y, Wang Y, Berg T, Nord M, Vihko P, Warner M, Piao YS, Gustafsson JA: **Differential regulation of estrogen receptor (ER) alpha and ERbeta in primate mammary gland.** *J Clin Endocrinol Metab* 2005, **90**:435-444.
38. Saji S, Jensen EV, Nilsson S, Rylander T, Warner M, Gustafsson JA: **Estrogen receptors alpha and beta in the rodent mammary gland.** *Proc Natl Acad Sci USA* 2000, **97**:337-342.
39. Chen G, Dellinger RW, Gallagher CJ, Sun D, Lazarus P: **Identification of a prevalent functional missense polymorphism in the UGT2B10 gene and its association with UGT2B10 inactivation against tobacco-specific nitrosamines.** *Pharmacogenet Genomics* 2008, **18**:181-191.
40. Perou CM, Sorlie T, Eisen MB, van de RM, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
41. Chen DT, Nasir A, Culhane A, Venkataramu C, Fulp W, Rubio R, Wang T, Agrawal D, McCarthy SM, Gruidl M, Bloom G, Anderson T, White J, Quackenbush J, Yeatman T: **Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue.** *Breast Cancer Res Treat* 2010, **119(2)**:335-346.
42. Showe MK, Vachani A, Kossenkov AV, Yousef M, Nichols C, Nikonova EV, Chang C, Kucharczuk J, Tran B, Wakeam E, Yie TA, Speicher D, Rom WN, Albelda S, Showe LC: **Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease.** *Cancer Res* 2009, **69(24)**:9202-9210.
43. Jones C, Mackay A, Grigoriadis A, Cossu A, Reis-Filho JS, Fulford L, Dexter T, Davies S, Bulmer K, Ford E, Parry S, Budroni M, Palmieri G, Neville AM, O'Hare MJ, Lakhani SR: **Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer.** *Cancer Res* 2004, **64**:3037-3045.

doi:10.1186/bcr2632
Cite this article as: Haakensen *et al.*: Expression levels of uridine 5'-diphospho-glucuronosyltransferase genes in breast tissue from healthy women are associated with mammographic density. *Breast Cancer Research* 2010 **12**:R65.

Additional file 1: SAM MD, healthy women.

Current settings

Input parameters

Data type?	Quantitative
Arrays centered?	FALSE
Delta	0.1128214
Minimum fold change	0
Test statistic	standard
Regression method	standard
Are data are log scale?	FALSE
Number of permutations	500
Input percentile for exchangeability factor s0	Automatic choice
Number of neighbors for KNN	10
Seed for Random number generator	1234567

Computed values

Estimate of pi0 (proportion of null genes)	0.9931402
Exchangibility factor s0	0.0010565
s0 percentile	0
False Discovery Rate (%)	50.528184

List of Significant Genes for Delta = 0.113

Positive genes (3)

Row	Gene ID	Gene Name	Score(d)	Numerator(r)	Denominator(s+s0)	q-value(%)
4650	A_24_P756494	LOC730057	2.86058017	0.015237575	0.005326743	34.05051997
8807	A_24_P329065	BTN3A1	2.828849448	0.011028004	0.003898406	34.05051997
9112	A_24_P311917	BTN3A3	2.760141709	0.009885918	0.00358167	34.05051997

Negative genes (54)

Row	Gene ID	Gene Name	Score(d)	Numerator(r)	Denominator(s+s0)	q-value(%)	FDR<25
6760	A_24_P932736	729641	-2.939794306	-0.024561301	0.008354769	0	
6129	A_23_P427472	FLJ10404	-2.898093839	-0.017947299	0.006192794	0	
9647	A_24_P18802	VPS18	-2.874371883	-0.020756984	0.007221398	0	
4530	A_23_P7342	UGT2B11	-2.870323551	-0.020651449	0.007194816	0	
6993	A_24_P177236	CABP7	-2.86828496	-0.025243675	0.008800965	0	
9508	A_24_P131589	CD86	-2.834444103	-0.025353336	0.00894473	0	
2950	A_23_P212968	UGT2B11	-2.791516519	-0.029177501	0.010452204	0	
3485	A_23_P152570	580687	-2.743200198	-0.020483623	0.007467054	0	
7883	A_23_P254212	DIAPH2::RPA4	-2.687939803	-0.020478069	0.0076185	8.276168049	
8366	A_32_P199824	LMOD1	-2.675114954	-0.023358045	0.008731604	8.276168049	
7637	A_24_P521559	UGT2B10	-2.656354673	-0.023386251	0.008803889	8.276168049	
2978	A_23_P66543	PIK3R5	-2.655411884	-0.017592155	0.006625019	8.276168049	
6371	A_32_P107994	ATG7	-2.590985913	-0.021891741	0.008449193	8.276168049	
5506	A_23_P155463	LRRC2	-2.573213695	-0.0191423	0.007439063	15.68116051	
4399	A_23_P28733	RBL1	-2.537519034	-0.009506857	0.003746516	15.68116051	
1537	A_23_P69699	NPY1R	-2.536481879	-0.025025147	0.009866086	15.68116051	
8728	A_23_P144244	810781	-2.511297337	-0.018143577	0.007224783	15.68116051	
4605	A_32_P80016	593535	-2.5104764	-0.019578524	0.007798728	15.68116051	
4562	A_23_P204277	H2AFJ	-2.505132635	-0.011727729	0.00468148	15.68116051	
5316	A_32_P35668	666399	-2.481846445	-0.016741092	0.006745418	15.68116051	
6564	A_24_P640617	Transcribed	-2.434158993	-0.016788046	0.006896857	18.91695554	
6272	A_32_P20997	Transcribed	-2.418337293	-0.011416998	0.004721011	21.59000361	
2288	A_23_P136671	UGT2B7	-2.403406049	-0.029123074	0.012117417	21.59000361	
3534	A_23_P309739	ESR1	-2.388785868	-0.017363723	0.007268849	21.59000361	
2471	A_23_P119448	SAPS1	-2.365047597	-0.016885621	0.007139654	23.83536398	

9547	A_24_P125894	PPM1F	-2.329501946	-0.013607624	0.005841	26.73838908
8792	A_24_P180243	UGT2B28	-2.294362745	-0.023268398	0.010142	28.37543331
7921	A_24_P693448	ZNF552	-2.292709794	-0.014196218	0.006192	28.37543331
9066	A_24_P36890	RAP1GAP	-2.277257711	-0.014689339	0.00645	30.82159135
9523	A_24_P287664	PLCB2	-2.264592578	-0.013955797	0.006163	33.1046722
7706	A_24_P178834	LOC132205::LOC2854	-2.251631636	-0.010829326	0.00481	33.1046722
3043	A_23_P256033	EEF1A2	-2.245056305	-0.00961755	0.004284	33.1046722
4986	A_24_P17691	UGT2B17	-2.191409165	-0.017356122	0.00792	40.26243916
2186	A_23_P90273	CHST8	-2.184547528	-0.011603115	0.005311	40.26243916
3978	A_23_P436284	OSTbeta	-2.173318271	-0.015594705	0.007176	40.26243916
6354	A_24_P844100	710943	-2.154978079	-0.010025909	0.004652	40.74421193
8645	A_24_P734406	CDNA	-2.139040689	-0.011402345	0.005331	40.74421193
8396	A_24_P913847	797019	-2.126742515	-0.013371348	0.006287	40.74421193
4352	A_32_P163469	NFE2L1	-2.122614145	-0.008279792	0.003901	40.74421193
27	A_32_P149404	537146	-2.121890923	-0.011735979	0.005531	40.74421193
8825	A_24_P8721	HIST2H2AC	-2.116853263	-0.007737682	0.003655	40.74421193
9195	A_24_P315014	825337	-2.103838237	-0.011041271	0.005248	44.13956293
2652	A_32_P59549	GFRA1	-2.089580714	-0.010791871	0.005165	44.13956293
9608	A_24_P585430	837185	-2.089248101	-0.008927945	0.004273	44.13956293
9565	A_24_P575267	835938	-2.085134242	-0.01329088	0.006374	44.13956293
4339	A_23_P55616	SLC14A1	-2.080633257	-0.010976489	0.005276	44.13956293
3525	A_23_P14072	KRT8	-2.073510618	-0.008600155	0.004148	44.13956293
5841	A_24_P171043	DKFZP547L112	-2.072368793	-0.014787908	0.007136	44.13956293
6833	A_23_P102071	A_23_P102071	-2.066516081	-0.012184098	0.005896	44.13956293
1659	A_23_P428184	HIST1H2AD	-2.062725933	-0.009282874	0.0045	44.13956293
2325	A_23_P40761	OSTalpha	-2.060190729	-0.009576087	0.004648	44.13956293
6618	A_23_P73526	CITED1	-2.048192347	-0.014631918	0.007144	47.8512989
7156	A_24_P46484	RBM22	-2.039381041	-0.015274032	0.00749	48.770276
7002	A_23_P384816	SLC45A4	-2.021143562	-0.012840954	0.006353	50.52818388

Estimated Miss rates for Delta=0.112821421147983

Quantiles	Cutpoints	Miss Rate(%)
0 -> 0.05	-2.012 -> -1.228	0
0.05 -> 0.1	-1.228 -> -0.969	3.49
0.1 -> 0.15	-0.969 -> -0.783	0
0.15 -> 0.2	-0.783 -> -0.633	0
0.2 -> 0.25	-0.633 -> -0.5	0
0.25 -> 0.75	-0.5 -> 0.533	0.44
0.75 -> 0.8	0.533 -> 0.654	8.29
0.8 -> 0.85	0.654 -> 0.807	0.71
0.85 -> 0.9	0.807 -> 1.002	0.76
0.9 -> 0.95	1.002 -> 1.254	11.81
0.95 -> 1	1.254 -> 2.505	0

Additional file 2 - Figures and tables

Figure S1: The percentage of samples with low expression of *UGT* genes (<-0.5) within tumour samples and healthy women with high (≥30%) or low (<30%) MD.

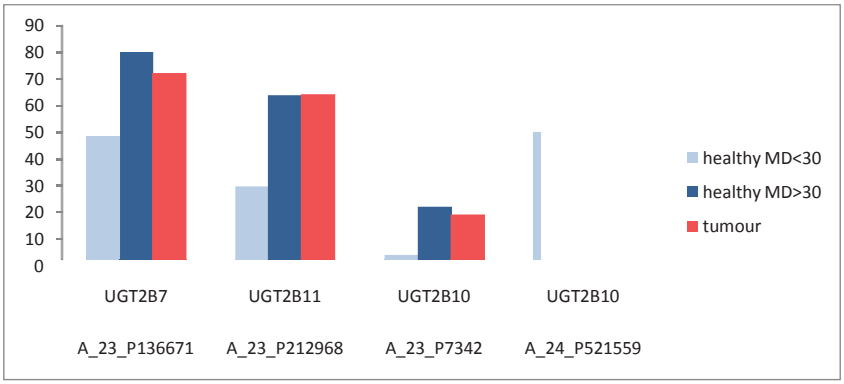


Figure S2: Clustering of the genes significantly down regulated in high MD samples. The three UGT genes cluster separately and tightly together. Samples are sorted according to MD.

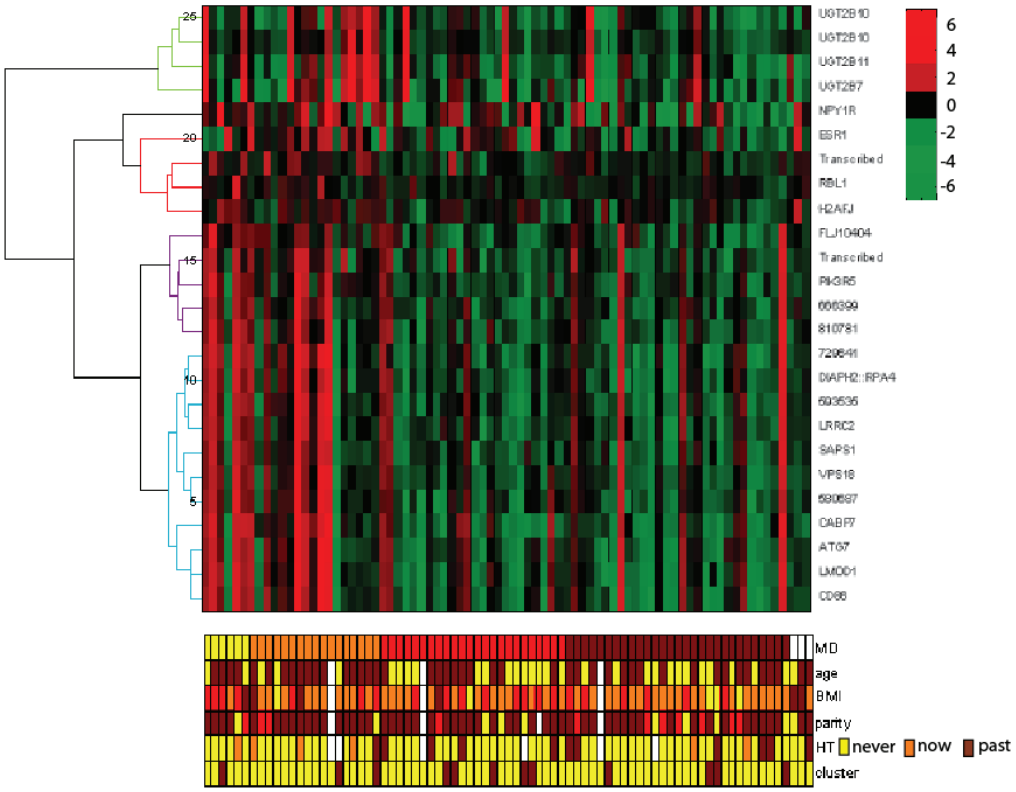


Figure S3: The expression of the different *UGT* genes is highly correlated. Expression of the four probes representing UGT-transcripts (y-axis) for each sample (x-axis) for A) healthy women and B) breast cancer patients respectively

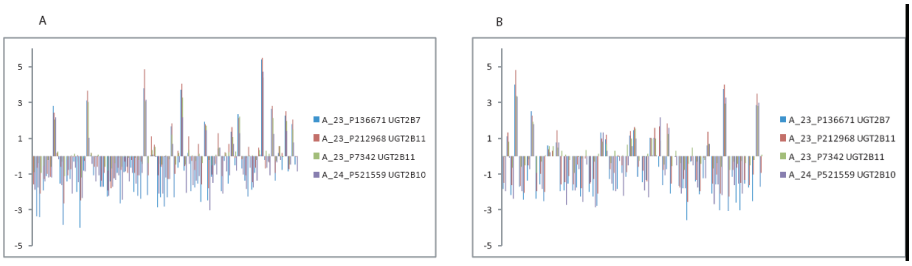


Table S1: Correlation between the expression of different UGT genes

		A_23_P136671 UGT2B7	A_23_P212968 UGT2B11	A_23_P7342 UGT2B10	A_24_P521559 UGT2B10
A_23_P136671	UGT2B7	-	0.92	0.91	0.90
A_23_P212968	UGT2B11	0.92	-	0.93	0.94
A_23_P7342	UGT2B10	0.91	0.93	-	0.90
A_24_P521559	UGT2B10	0.90	0.94	0.90	-

Table S2: MD and expression of UGT genes in samples from healthy women

		MD	A_23_P136671 UGT2B7 expression	A_23_P7342 UGT2B10 expression	A_24_P521559 UGT2B10 expression	A_23_P212968 UGT2B11 expression
age	mean <50 (n=31)	40.4	-0.83	0.15	-0.74	-0.07
	mean ≥50 (n=45)	35.2	-0.68	0.22	-0.70	-0.07
	p-value	0.25	0.72	0.78	0.92	0.99
BMI	mean <25 (n=49)	41.2	-0.84	0.12	-0.77	-0.11
	mean ≥25 (n=27)	30.0	-0.57	0.31	-0.63	0.02
	p-value	0.01	0.53	0.45	0.66	0.73
Age at first birth	mean <25 (n=38)	31.5	-0.66	0.25	-0.60	-0.03
	mean ≥25 (n=18)	39.0	-0.64	0.22	-0.48	0.00
	p-value	0.14	0.96	0.93	0.77	0.95
Current hormone therapy	mean yes (n=64)	46.5	-0.97	-0.12	-1.04	-0.50
	mean no (n=12)	35.6	-0.71	0.24	-0.64	0.01
	p-value	0.07	0.65	0.25	0.35	0.30

MD and expression of UGT genes in samples from healthy women in relation to epidemiological factors. All p-values are from two-sided t-tests not corrected for multiple testing.

Figure S3: Unsupervised hierarchical clustering showed two main clusters. MD was not significantly different between these two clusters.

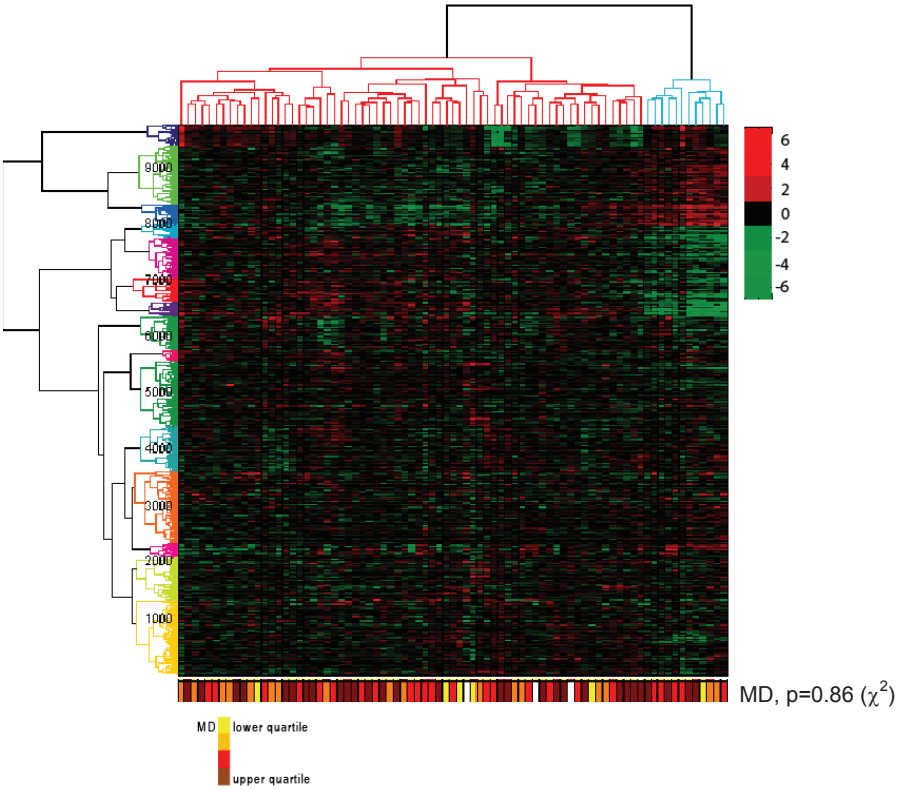
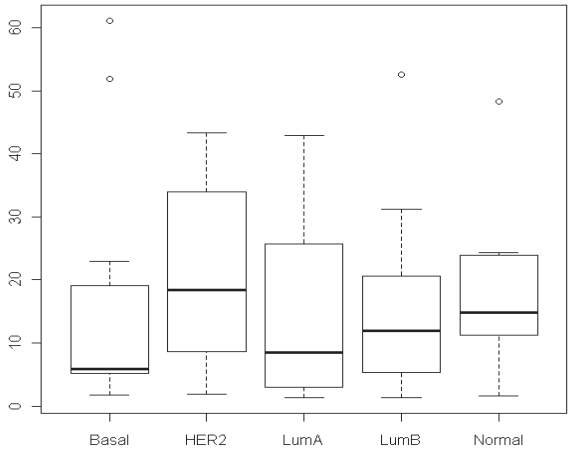


Figure S4: MD in relation to tumour subgroups

a) Boxplot of MD vs subtypes



a) Boxplot of MD for each pam50 risk group (1=high, 3=medium, 5=low risk).

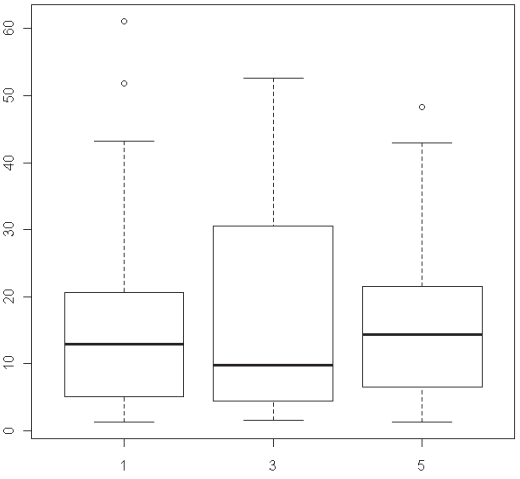
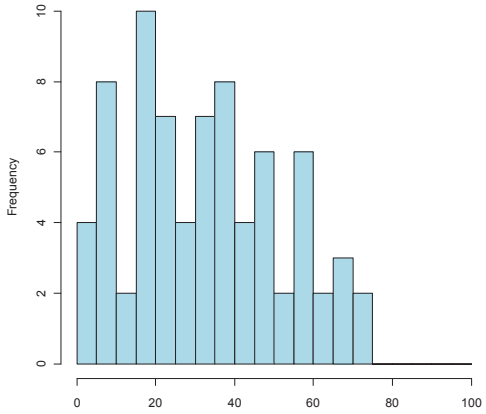
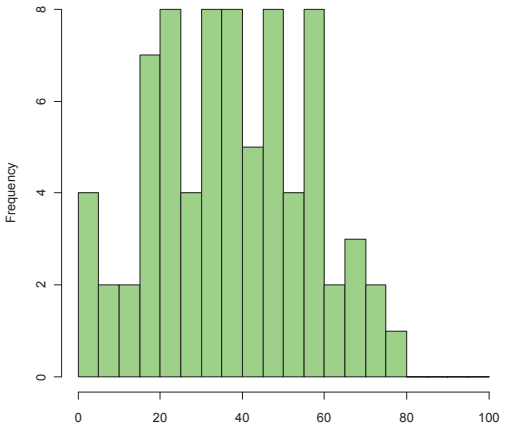


Figure S5: Distribution of MD in a) all samples (mean=), b) healthy women (mean=37%) and c) breast cancer patients (mean=16%)

a) MD frequency all women



b) MD frequency healthy women



c) MD frequency breast cancer patients

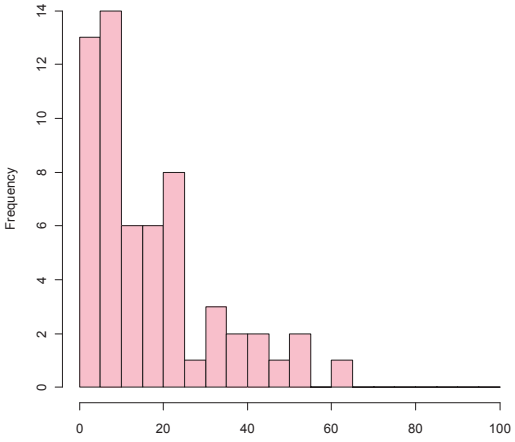


Figure S6: Localisation of UGT genes on chromosome 4:

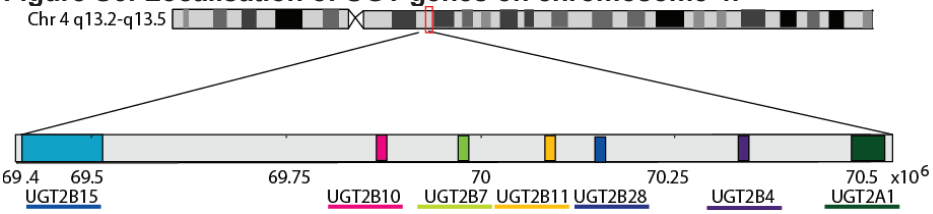


Table S4: Gene expression in samples from healthy women with differing MD compared with tumour samples.

A) MD < 30% vs MD>30%: The mean expression of all four probes representing *UGT* genes is not significantly different between tumour samples and normal samples from breasts with high MD. For three UGT probes, mean expression is significantly different between tumour samples and normal samples from breasts with low MD. This is not the case for most other genes.

Agilent ID	SYMBOL	mean t	mean n MD <30%	mean n MD>30%	ttest t vs MD<30%	ttest t vs n MD>30%
A_23_P119448	SAPS1	0.87	0.98	0.13	0.693	0.000
A_23_P136671	**UGT2B7**	**-1.00**	**0.21**	**-1.24**	**0.005**	**0.402**
A_23_P144244	810781	1.16	1.12	0.28	0.889	0.000
A_23_P152570	580687	1.53	1.78	0.87	0.430	0.003
A_23_P155463	LRRC2	1.39	1.60	0.65	0.478	0.001
A_23_P204277	H2AFJ	1.62	1.20	0.79	0.047	0.000
A_23_P212968	**UGT2B11**	**-0.43**	**0.77**	**-0.50**	**0.003**	**0.768**
A_23_P254212	DIAPH2::RPA4	1.50	1.62	0.68	0.705	0.000
A_23_P28733	RBL1	-0.90	-0.77	-1.12	0.306	0.010
A_23_P309739	ESR1	2.32	1.96	1.32	0.282	0.000
A_23_P427472	FLJ10404	1.21	1.44	0.51	0.466	0.002
A_23_P66543	PIK3R5	0.89	0.95	0.07	0.833	0.000
A_23_P69699	NPY1R	1.42	2.01	1.35	0.267	0.902
A_23_P7342	**UGT2B10**	**0.16**	**0.79**	**-0.13**	**0.010**	**0.054**
A_24_P131589	CD86	1.57	1.93	0.79	0.305	0.002
A_24_P177236	CABP7	1.63	1.95	0.68	0.378	0.000
A_24_P18802	VPS18	1.44	1.64	0.69	0.510	0.000
A_24_P521559	**UGT2B10**	**-0.70**	**0.07**	**-1.08**	**0.072**	**0.148**
A_24_P640617	Transcribed	0.97	1.04	0.18	0.801	0.000
A_24_P932736	729641	1.55	1.68	0.59	0.714	0.000
A_32_P107994	ATG7	1.51	1.78	0.74	0.413	0.001
A_32_P199824	LMOD1	1.55	1.91	0.77	0.320	0.002
A_32_P20997	Transcribed	0.25	0.56	0.10	0.008	0.092
A_32_P35668	666399	1.31	1.51	0.62	0.499	0.001
A_32_P80016	593535	1.59	1.72	0.75	0.676	0.000

Table S4 cont B) MD<20% vs MD>40%: The mean expression of all four probes representing UGT genes is not significantly different between tumour samples and normal samples from breasts with high MD, as opposed to most other probes. There is no significant difference in mean expression between tumour samples and normal samples from breasts with low MD. Contrary to most other probes, the mean expression of the UGT genes in tumours is more similar to the mean expression in normal samples from breasts with high than low MD.

			MD <20% vs >40%							
Agilent ID	SYMBOL	mean t	mean n MD <20%	mean n MD >40%	ttest t vs n MD <20%	ttest t vs n MD >40%	t-n low MD	|t-n low MD|	t-n high MD	t closer to
A_23_P119448	SAPS1	0.87	1.31	0.18	0.23	0.01	-0.44	0.44	0.69	low
A_23_P136671	**UGT2B7**	**-1.00**	**-0.39**	**-1.22**	**0.22**	**0.52**	**-0.61**	**0.61**	**0.21**	**high**
A_23_P144244	810781	1.16	1.46	0.37	0.42	0.00	-0.31	0.31	0.79	low
A_23_P152570	580687	1.53	2.12	0.98	0.13	0.04	-0.59	0.59	0.55	high
A_23_P155463	LRRC2	1.39	1.78	0.72	0.29	0.01	-0.39	0.39	0.67	low
A_23_P204277	H2AFJ	1.62	1.18	0.73	0.11	0.00	0.44	0.44	0.89	low
A_23_P212968	**UGT2B11**	**-0.43**	**0.24**	**-0.52**	**0.15**	**0.75**	**-0.66**	**0.66**	**0.09**	**high**
A_23_P254212	DIAPH2::RPA4	1.50	1.98	0.75	0.23	0.01	-0.48	0.48	0.76	low
A_23_P28733	RBL1	-0.90	-0.70	-1.15	0.18	0.01	-0.20	0.20	0.26	low
A_23_P309739	ESR1	2.32	1.89	1.21	0.32	0.00	0.43	0.43	1.12	low
A_23_P427472	FLJ10404	1.21	1.70	0.53	0.24	0.01	-0.49	0.49	0.68	low
A_23_P66543	PIK3R5	0.89	1.14	0.17	0.51	0.01	-0.25	0.25	0.72	low
A_23_P69699	NPY1R	1.42	2.21	1.26	0.25	0.73	-0.78	0.78	0.16	high
A_23_P7342	**UGT2B10**	**0.16**	**0.50**	**-0.15**	**0.20**	**0.07**	**-0.34**	**0.34**	**0.30**	**high**
A_24_P131589	CD86	1.57	2.26	0.87	0.10	0.02	-0.68	0.68	0.70	low
A_24_P177236	CABP7	1.63	2.25	0.77	0.17	0.01	-0.62	0.62	0.86	low
A_24_P18802	VPS18	1.44	1.88	0.77	0.23	0.01	-0.44	0.44	0.67	low
A_24_P521559	**UGT2B10**	**-0.70**	**-0.38**	**-1.14**	**0.52**	**0.17**	**-0.32**	**0.32**	**0.44**	**low**
A_24_P640617	Transcribed	0.97	1.25	0.32	0.42	0.01	-0.28	0.28	0.65	low
A_24_P932736	729641	1.55	2.10	0.64	0.19	0.00	-0.55	0.55	0.91	low
A_32_P107994	ATG7	1.51	2.07	0.83	0.17	0.02	-0.57	0.57	0.68	low
A_32_P199824	LMOD1	1.55	2.22	0.84	0.12	0.02	-0.67	0.67	0.71	low
A_32_P20997	Transcribed	0.25	0.69	0.15	0.00	0.28	-0.44	0.44	0.10	high
A_32_P35668	666399	1.31	1.71	0.71	0.28	0.02	-0.41	0.41	0.60	low
A_32_P80016	593535	1.59	2.02	0.82	0.26	0.00	-0.42	0.42	0.78	low

Figure S7: The Kullback-Leibler divergence between UGT expression in tumours and healthy samples with high and low MD respectively. Small divergence means more similar distribution in the two populations tested. For three of four UGT probes, the distribution in tumour samples is more similar to the distribution in healthy individuals with high-MD than with low-MD. (The first UGT2B10-probe is A_23_P7342, the probe that is significant in the GLM-analysis, the second is A_24_P521559).

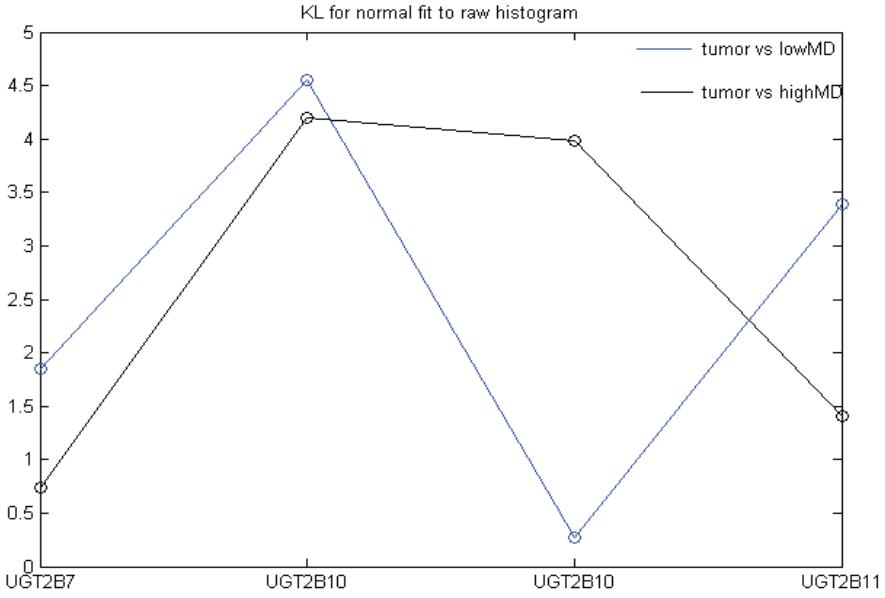


Table S5: Gene expression of UGT2B10 in tumour samples (T) and normal adjacent samples (N) from the same breast in an unpublished dataset. There is no significant difference in mean expression by pair wise t-test.

	A_23_P7342 UGT2B10		A_23_P7342 UGT2B10
CM 1N	9.75	CM 1T	7.52
CM 9N	8.62	CM 9T	8.49
CM 10N	8.77	CM 10T	7.27
CM 11N	9.26	CM 11T	6.65
CM 13N	10.32	CM 13T	7.43
CM 18N	10.69	CM 18T	8.79
CM 19N	9.12	CM 19T	9.11
CM 26N	8.78	CM 26T	7.31
CM 31N	8.00	CM 31T	8.07
CM 32N	7.73	CM 32T	7.23
CM 38N	10.96	CM 38T	7.32
CM 41N	9.07	CM 41T	11.71
CM 46N	8.74	CM 46T	11.21
CM 47N	11.08	CM 47T	8.00
CM 54N	7.80	CM 54T	13.14
CM 56N	7.92	CM 56T	17.48
CMG24N	6.66	CMG24T	13.89
CMG43N	8.24	CMG43T	10.59
CM 44N	8.89	CM 44T	7.75
average	8.97	average	9.42
p-value	0.60	(pair wise t-test)	

Table S6: Range of MD in the breasts of healthy women

max	**77.3133466**
min	**1.28417454**
mean	**28.1668549**
median	**23.5504612**

Supplemental discussion

Gene expression microarray analyses using tissue adjacent to a breast tumour have previously been done [1,2]. The expression profile in these normal samples will be influenced by the neighbouring breast tumour [3]. Breast reduction mammoplasties have also been used in analysis of healthy breast tissue [4]. These samples are generally collected from large breasts with a higher than average proportion of fatty tissue which may also skew the analyses to some extent. Our study analysed a population more representative of the population of women at risk for developing breast cancer, since we have studied normal breast tissue from women with no malignant disease and not undergoing breast reduction mammoplasties.

Reference List

1. Chen DT, Nasir A, Culhane A, Venkataramu C, Fulp W, Rubio R, Wang T, Agrawal D, McCarthy SM, Gruidl M et al.: **Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue.** *Breast Cancer Res Treat* 2009.

2. Yang WT, Lewis MT, Hess K, Wong H, Tsimelzon A, Karadag N, Cairo M, Wei C, Meric-Bernstam F, Brown P et al.: **Decreased TGFbeta signaling and increased COX2 expression in high risk women with increased mammographic breast density.** *Breast Cancer Res Treat* 2009.

3. Showe MK, Vachani A, Kossenkov AV, Yousef M, Nichols C, Nikonova EV, Chang C, Kucharczuk J, Tran B, Wakeam E et al.: **Gene Expression Profiles in Peripheral Blood Mononuclear Cells Can Distinguish Patients with Non-Small Cell Lung Cancer from Patients with Nonmalignant Lung Disease.** *Cancer Res* 2009.

4. Jones C, Mackay A, Grigoriadis A, Cossu A, Reis-Filho JS, Fulford L, Dexter T, Davies S, Bulmer K, Ford E et al.: **Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer.** *Cancer Res* 2004, **64:**3037-3045.

Additional file 3: SAM MD, Healthy women, stratified

Healthy women MD 20+

n= 61

Down-regulated in high MD

Gene ID	Gene Name	Score(d)	q-value(%)
A_23_P212968	UGT2B11	-2.77969	0
A_24_P521559	UGT2B10	-2.70433	0
A_24_P180243	UGT2B28	-2.49195	0
A_24_P17691	UGT2B17	-2.49119	0
A_23_P7342	UGT2B11	-2.43767	0
A_23_P136671	UGT2B7	-2.39515	0
A_24_P575267	835938	-2.22042	16.58061
A_23_P309739	ESR1	-2.11901	29.01607

Healthy women age 50+

n= 43

Down-regulated in high MD

Gene ID	Gene Name	Score(d)	q-value(%)
A_24_P932736	729641	-2.84599	0
A_23_P102071	A_23_P102071	-2.79848	0
A_24_P131589	CD86	-2.73066	0
A_24_P18802	VPS18	-2.68478	0
A_23_P144244	810781	-2.63993	0
A_23_P427472	FLJ10404	-2.58166	0
A_23_P254212	DIAPH2::RPA4	-2.54627	0
A_32_P199824	LMOD1	-2.54021	0
A_24_P107994	ATG7	-2.51581	0
A_24_P111096	PFKFB3	-2.49518	0
A_32_P35668	666399	-2.47058	0
A_23_P152570	580687	-2.42633	0
A_32_P133840	TMCC2	-2.40001	5.212564
A_24_P375205	MKL2	-2.39587	5.212564
A_23_P69699	NPY1R	-2.36047	5.212564
A_24_P177236	CABP7	-2.35609	5.212564
A_32_P80016	593535	-2.34829	5.212564
A_24_P693448	ZNF552	-2.34813	5.212564
A_23_P155463	LRRC2	-2.31267	5.212564
A_23_P366376	TDGF3	-2.28853	5.212564
A_23_P66543	PIK3R5	-2.24972	5.212564
A_24_P640617	Transcribed	-2.24753	5.212564
A_24_P913847	797019	-2.21721	5.212564
A_24_P46484	RBM22	-2.21441	5.212564
A_23_P204277	H2AFJ	-2.15817	7.217396
A_32_P111996	MGC39584	-2.155	7.217396
A_23_P39095	CGB::CGB1	-2.12366	8.796201
A_23_P119448	SAPS1	-2.11473	8.796201
A_32_P20997	Transcribed	-2.11014	8.796201
A_24_P287664	PLCB2	-2.10975	8.796201
A_32_P77416	554489	-2.106	8.796201
A_24_P919640	CD44	-2.10298	8.796201
A_24_P125894	PPM1F	-2.08066	11.03837
A_24_P349633	FLJ32679::GOLGA8F	-2.07426	11.03837
A_23_P116694	RPS26	-2.0591	11.03837
A_23_P436284	OSTbeta	-2.01381	13.03141
A_24_P315014	825337	-2.00604	13.03141
A_24_P36890	RAP1GAP	-1.98816	14.81466
A_23_P44663	SERPINA1	-1.93374	16.41958
A_32_P149404	537146	-1.92918	16.41958
A_32_P79313	FLJ45244	-1.91531	18.30754
A_24_P110601	834483	-1.90173	18.30754
A_23_P414793	CP	-1.90022	18.30754
A_23_P38732	CDH2	-1.89132	20.39699
A_24_P82880	TPM4	-1.88579	20.39699
A_32_P147241	PKM2	-1.88028	20.39699
A_32_P168431	RPS26	-1.85413	23.45654
A_23_P428184	HIST1H2AD	-1.84698	23.45654
A_24_P719081	786677	-1.83241	24.89265

Healthy women age <50

n= 30

Down-regulated in high MD

Gene ID	Gene Name	Score(d)	q-value(%)
A_23_P31816	DEFA1	-2.42047	108.98
A_24_P945408	A_24_P945408	-2.10504	108.98
A_23_P28485	GCA	-2.09072	108.98
A_23_P133606	SLC12A2	-1.96186	108.98
A_23_P155666	ASAHL	-1.85119	108.98
A_24_P521559	UGT2B10	-1.85097	108.98
A_23_P57961	PLXNB1	-1.85011	108.98
A_23_P251002	A_23_P251002	-1.75684	108.98
A_24_P180243	UGT2B28	-1.73066	108.98
A_24_P682550	805257	-1.73046	108.98
A_24_P926053	EEF1D	-1.73024	108.98
A_24_P234732	MXD4	-1.71981	108.98
A_23_P7342	UGT2B10	-1.71718	108.98
A_23_P218144	LTBP2	-1.67917	108.98
A_23_P136671	UGT2B7	-1.67406	108.98
A_23_P55616	SLC14A1	-1.66802	108.98
A_24_P649357	LOC153561::SMA3	-1.65662	108.98
A_24_P105913	660721	-1.63813	108.98
A_23_P1833	B3GAT1	-1.6233	108.98
A_23_P66481	RTN4RL1	-1.61872	108.98
A_24_P942694	C10orf118	-1.6163	108.98
A_23_P212968	UGT2B11	-1.61244	108.98

Healthy women currently using HT

n=11

Down-regulated in high MD

Gene ID	Gene Name	Score(d)	q-value(%)
A_23_P150979	SBEM	-2.02043	0
A_23_P8702	PIP	-1.99379	0
A_23_P136671	UGT2B7	-1.76847	0
A_23_P393099	TFF3	-1.73037	0
A_24_P701582	755742	-1.71496	0
A_23_P212968	UGT2B11	-1.70231	0
A_24_P180243	UGT2B28	-1.61254	14.85761091

Healthy women ≥ 50 not currently using HT

n= 32

Down-regulated in high MD

Gene ID	Gene Name	Score(d)	q-value(%)
A_23_P39095	CGB::CGB1	-2.30668	13.41
A_24_P131589	CD86	-2.24301	13.41
A_23_P102071	A_23_P102071	-2.24149	13.41
A_32_P199824	LMOD1	-2.16352	13.41
A_24_P932736	729641	-2.13709	13.41
A_23_P144244	810781	-2.11942	13.41
A_24_P640617	Transcribed	-2.10214	13.41
A_24_P18802	VPS18	-2.094	13.41
A_23_P366376	TDGF3	-2.07299	13.41
A_23_P436284	OSTbeta	-2.06873	13.41
A_24_P111096	PFKFB3	-2.06416	13.41
A_23_P254212	DIAPH2::RPA4	-2.0602	13.41
A_32_P80016	593535	-2.05224	13.41
A_23_P152570	580687	-2.04327	13.41
A_23_P155463	LRRC2	-2.03448	13.41
A_24_P913847	797019	-2.02459	13.41
A_23_P66543	PIK3R5	-1.99902	13.41
A_24_P919640	CD44	-1.98346	13.41
A_32_P133840	TMCC2	-1.97499	13.41
A_32_P107994	ATG7	-1.97384	13.41
A_32_P149404	537146	-1.88689	20.43

Additional file 4: eQTL

UGT transcripts the expression of which is associated to SNPs in their own or other UGT genes in cis

Probe_ID	Gene exp	SNP_rs	SNP_gene	b1_p
A_23_P41553	Ncaml	rs1828705	UGT2B10	0.01
A_24_P575267	835938	rs1828705	UGT2B10	0.02
A_23_P136671	UGT2B7	rs1828705	UGT2B10	0.04
A_23_P41553	Ncaml	rs941389	UGT2B4	0.05
A_24_P180243	UGT2B28	rs1828705	UGT2B10	0.05
A_24_P180243	UGT2B28	rs2288741	UGT2A1	0.07
A_23_P7342	UGT2B11	rs2288741	UGT2A1	0.08
A_23_P58407	UGT2B15	rs1828705	UGT2B10	0.08
A_24_P180243	UGT2B28	rs4554145	UGT2B4	0.08
A_24_P575267	835938	rs4554145	UGT2B4	0.08
A_24_P575267	835938	rs13139888	UGT2B4	0.08
A_23_P41553	Ncaml	rs4557343	UGT2B4	0.08
A_23_P7342	UGT2B11	rs1560605	UGT2A1	0.09
A_24_P180243	UGT2B28	rs1560605	UGT2A1	0.09
A_23_P212968	UGT2B11	rs1828705	UGT2B10	0.10
A_24_P575267	835938	rs3775782	UGT2A1	0.10
A_24_P180243	UGT2B28	rs10026603	UGT2A1	0.11
A_24_P575267	835938	rs2288741	UGT2A1	0.13
A_24_P180243	UGT2B28	rs1432329	UGT2A1	0.13
A_24_P575267	835938	rs1560605	UGT2A1	0.15
A_24_P180243	UGT2B28	rs3775782	UGT2A1	0.15
A_23_P58407	UGT2B15	rs1454254	UGT2B15	0.16
A_24_P575267	835938	rs1131878	UGT2B4	0.16
A_24_P521559	UGT2B10	rs10026603	UGT2A1	0.16
A_23_P7342	UGT2B11	rs1432329	UGT2A1	0.19
A_24_P575267	835938	rs1432329	UGT2A1	0.20
A_23_P7342	UGT2B11	rs10026603	UGT2A1	0.21
A_24_P180243	UGT2B28	rs13139888	UGT2B4	0.21
A_24_P180243	UGT2B28	rs1131878	UGT2B4	0.21
A_23_P212968	UGT2B11	rs2288741	UGT2A1	0.23
A_23_P212968	UGT2B11	rs1560605	UGT2A1	0.23
A_23_P7342	UGT2B11	rs1828705	UGT2B10	0.23
A_24_P521559	UGT2B10	rs2288741	UGT2A1	0.24
A_24_P180243	UGT2B28	rs4557343	UGT2B4	0.25
A_23_P7342	UGT2B11	rs4557343	UGT2B4	0.25
A_24_P521559	UGT2B10	rs1560605	UGT2A1	0.28
A_23_P7342	UGT2B11	rs3775782	UGT2A1	0.29
A_24_P521559	UGT2B10	rs1828705	UGT2B10	0.29
A_24_P521559	UGT2B10	rs4554145	UGT2B4	0.29
A_23_P7342	UGT2B11	rs4554145	UGT2B4	0.30
A_24_P17691	UGT2B17	rs1828705	UGT2B10	0.31
A_23_P41553	Ncaml	rs3775782	UGT2A1	0.31
A_24_P521559	UGT2B10	rs4235126	UGT2B28	0.33
A_23_P212968	UGT2B11	rs7668258	UGT2B7	0.34
A_23_P212968	UGT2B11	rs4521414	UGT2B7	0.34
A_23_P41553	Ncaml	rs1513559	UGT2B10	0.35
A_23_P212968	UGT2B11	rs10026603	UGT2A1	0.35
A_24_P575267	835938	rs4557343	UGT2B4	0.36
A_23_P212968	UGT2B11	rs7439366	UGT2B7	0.36
A_23_P41553	Ncaml	rs844342	UGT2B10	0.37
A_23_P136671	UGT2B7	rs4554145	UGT2B4	0.38
A_23_P212968	UGT2B11	rs3775782	UGT2A1	0.38
A_23_P136671	UGT2B7	rs10026603	UGT2A1	0.38
A_24_P180243	UGT2B28	rs4694211	UGT2B4	0.40
A_24_P521559	UGT2B10	rs3775782	UGT2A1	0.41
A_24_P521559	UGT2B10	rs1432329	UGT2A1	0.41
A_23_P7342	UGT2B11	rs13139888	UGT2B4	0.41
A_23_P136671	UGT2B7	rs3775782	UGT2A1	0.42
A_23_P136671	UGT2B7	rs2288741	UGT2A1	0.44
A_23_P41553	Ncaml	rs1454254	UGT2B15	0.44
A_23_P41553	Ncaml	rs4554145	UGT2B4	0.44
A_23_P7342	UGT2B11	rs1131878	UGT2B4	0.45
A_24_P575267	835938	rs2045100	UGT2B15	0.45
A_24_P180243	UGT2B28	rs2045100	UGT2B15	0.46
A_23_P212968	UGT2B11	rs1432329	UGT2A1	0.46
A_24_P521559	UGT2B10	rs13139888	UGT2B4	0.47
A_23_P58407	UGT2B15	rs2045100	UGT2B15	0.47

Probe_ID	Gene exp	SNP_rs	SNP_gene	b1_p
A_24_P521559	UGT2B10	rs1131878	UGT2B4	0.48
A_23_P136671	UGT2B7	rs1560605	UGT2A1	0.48
A_24_P521559	UGT2B10	rs903446	UGT2B4	0.49
A_24_P575267	835938	rs10026603	UGT2A1	0.49
A_23_P41553	Ncaml	rs13139888	UGT2B4	0.50
A_23_P212968	UGT2B11	rs4554145	UGT2B4	0.51
A_23_P212968	UGT2B11	rs2045100	UGT2B15	0.51
A_24_P575267	835938	rs4694211	UGT2B4	0.51
A_23_P58407	UGT2B15	rs1513559	UGT2B10	0.55
A_23_P212968	UGT2B11	rs4557343	UGT2B4	0.57
A_24_P521559	UGT2B10	rs7439366	UGT2B7	0.59
A_23_P58407	UGT2B15	rs7668258	UGT2B7	0.59
A_23_P58407	UGT2B15	rs4521414	UGT2B7	0.59
A_23_P136671	UGT2B7	rs13139888	UGT2B4	0.59
A_23_P136671	UGT2B7	rs4557343	UGT2B4	0.60
A_23_P7342	UGT2B11	rs7668258	UGT2B7	0.60
A_23_P7342	UGT2B11	rs4521414	UGT2B7	0.60
A_23_P136671	UGT2B7	rs2045100	UGT2B15	0.61
A_24_P180243	UGT2B28	rs7439366	UGT2B7	0.62
A_23_P136671	UGT2B7	rs941389	UGT2B4	0.62
A_24_P575267	835938	rs4148279	UGT2A1	0.63
A_24_P17691	UGT2B17	rs1513559	UGT2B10	0.63
A_24_P575267	835938	rs1513559	UGT2B10	0.64
A_23_P136671	UGT2B7	rs1432329	UGT2A1	0.64
A_24_P575267	835938	rs7668258	UGT2B7	0.65
A_24_P575267	835938	rs4521414	UGT2B7	0.65
A_24_P575267	835938	rs4235126	UGT2B28	0.65
A_23_P136671	UGT2B7	rs4694211	UGT2B4	0.66
A_23_P212968	UGT2B11	rs13139888	UGT2B4	0.67
A_23_P136671	UGT2B7	rs1131878	UGT2B4	0.67
A_24_P17691	UGT2B17	rs1454254	UGT2B15	0.68
A_23_P7342	UGT2B11	rs7439366	UGT2B7	0.68
A_24_P17691	UGT2B17	rs844342	UGT2B10	0.69
A_23_P7342	UGT2B11	rs4148279	UGT2A1	0.69
A_24_P521559	UGT2B10	rs4694211	UGT2B4	0.70
A_23_P7342	UGT2B11	rs844342	UGT2B10	0.70
A_24_P521559	UGT2B10	rs7668258	UGT2B7	0.71
A_24_P521559	UGT2B10	rs4521414	UGT2B7	0.71
A_24_P521559	UGT2B10	rs844342	UGT2B10	0.72
A_23_P212968	UGT2B11	rs1131878	UGT2B4	0.73
A_23_P41553	Ncaml	rs4235126	UGT2B28	0.73
A_24_P180243	UGT2B28	rs4235126	UGT2B28	0.73
A_24_P180243	UGT2B28	rs844342	UGT2B10	0.74
A_23_P7342	UGT2B11	rs903446	UGT2B4	0.74
A_24_P521559	UGT2B10	rs941389	UGT2B4	0.75
A_23_P136671	UGT2B7	rs4235126	UGT2B28	0.75
A_24_P521559	UGT2B10	rs2045100	UGT2B15	0.76
A_24_P575267	835938	rs7439366	UGT2B7	0.77
A_23_P7342	UGT2B11	rs941389	UGT2B4	0.77
A_24_P575267	835938	rs844342	UGT2B10	0.78
A_24_P180243	UGT2B28	rs7668258	UGT2B7	0.78
A_24_P180243	UGT2B28	rs4521414	UGT2B7	0.78
A_23_P136671	UGT2B7	rs4148279	UGT2A1	0.78
A_23_P7342	UGT2B11	rs4235126	UGT2B28	0.79
A_23_P212968	UGT2B11	rs4694211	UGT2B4	0.79
A_23_P41553	Ncaml	rs1131878	UGT2B4	0.79
A_24_P575267	835938	rs1454254	UGT2B15	0.80
A_23_P7342	UGT2B11	rs1513559	UGT2B10	0.80
A_23_P58407	UGT2B15	rs7439366	UGT2B7	0.80
A_23_P41553	Ncaml	rs7439366	UGT2B7	0.82
A_23_P41553	Ncaml	rs7668258	UGT2B7	0.83
A_23_P41553	Ncaml	rs4521414	UGT2B7	0.83
A_24_P521559	UGT2B10	rs4557343	UGT2B4	0.83
A_23_P58407	UGT2B15	rs844342	UGT2B10	0.83
A_23_P41553	Ncaml	rs4694211	UGT2B4	0.83
A_23_P7342	UGT2B11	rs2045100	UGT2B15	0.84
A_23_P212968	UGT2B11	rs903446	UGT2B4	0.84

Probe_ID	Gene exp	SNP_rs	SNP_gene	b1_p
A_23_P212968	UGT2B11	rs4235126	UGT2B28	0.84
A_23_P212968	UGT2B11	rs844342	UGT2B10	0.84
A_24_P180243	UGT2B28	rs1454254	UGT2B15	0.85
A_24_P180243	UGT2B28	rs903446	UGT2B4	0.87
A_23_P212968	UGT2B11	rs1513559	UGT2B10	0.87
A_23_P136671	UGT2B7	rs844342	UGT2B10	0.88
A_24_P180243	UGT2B28	rs4148279	UGT2A1	0.90
A_24_P521559	UGT2B10	rs4148279	UGT2A1	0.90
A_24_P521559	UGT2B10	rs1513559	UGT2B10	0.90
A_23_P41553	Ncaml	rs903446	UGT2B4	0.90
A_23_P136671	UGT2B7	rs7668258	UGT2B7	0.90
A_23_P136671	UGT2B7	rs4521414	UGT2B7	0.90
A_23_P212968	UGT2B11	rs941389	UGT2B4	0.91
A_24_P521559	UGT2B10	rs1454254	UGT2B15	0.93
A_23_P7342	UGT2B11	rs1454254	UGT2B15	0.93
A_23_P212968	UGT2B11	rs1454254	UGT2B15	0.94
A_23_P41553	Ncaml	rs2045100	UGT2B15	0.95
A_24_P17691	UGT2B17	rs2045100	UGT2B15	0.95
A_24_P180243	UGT2B28	rs1513559	UGT2B10	0.96
A_23_P136671	UGT2B7	rs903446	UGT2B4	0.96
A_24_P575267	835938	rs941389	UGT2B4	0.97
A_23_P136671	UGT2B7	rs1454254	UGT2B15	0.98
A_23_P7342	UGT2B11	rs4694211	UGT2B4	0.98
A_23_P212968	UGT2B11	rs4148279	UGT2A1	0.98
A_23_P41553	Ncaml	rs2288741	UGT2A1	0.98
A_23_P136671	UGT2B7	rs1513559	UGT2B10	0.99
A_23_P136671	UGT2B7	rs7439366	UGT2B7	0.99
A_24_P180243	UGT2B28	rs941389	UGT2B4	1.00
A_23_P362694	C4orf7	rs903446	UGT2B4	0.01
A_23_P41365	SMR3A	rs941389	UGT2B4	0.06
A_23_P362694	C4orf7	rs7439366	UGT2B7	0.10
A_23_P362694	C4orf7	rs4148279	UGT2A1	0.10
A_23_P362694	C4orf7	rs7668258	UGT2B7	0.14
A_23_P362694	C4orf7	rs4521414	UGT2B7	0.14
A_23_P110234	CSN1S1	rs3775782	UGT2A1	0.18
A_23_P110234	CSN1S1	rs13139888	UGT2B4	0.18
A_23_P110234	CSN1S1	rs1131878	UGT2B4	0.19
A_23_P362694	C4orf7	rs1131878	UGT2B4	0.22
A_23_P362694	C4orf7	rs4554145	UGT2B4	0.23
A_23_P110234	CSN1S1	rs4554145	UGT2B4	0.24
A_23_P41365	SMR3A	rs3775782	UGT2A1	0.27
A_23_P362694	C4orf7	rs4235126	UGT2B28	0.33
A_23_P110234	CSN1S1	rs4694211	UGT2B4	0.34
A_23_P41365	SMR3A	rs4557343	UGT2B4	0.40
A_23_P362694	C4orf7	rs4694211	UGT2B4	0.43
A_23_P110234	CSN1S1	rs1432329	UGT2A1	0.49
A_23_P362694	C4orf7	rs3775782	UGT2A1	0.52
A_23_P41365	SMR3A	rs13139888	UGT2B4	0.52
A_23_P362694	C4orf7	rs2288741	UGT2A1	0.58
A_23_P110234	CSN1S1	rs4557343	UGT2B4	0.58
A_23_P41365	SMR3A	rs903446	UGT2B4	0.61
A_23_P41365	SMR3A	rs4554145	UGT2B4	0.64
A_23_P110234	CSN1S1	rs4235126	UGT2B28	0.68
A_23_P41365	SMR3A	rs4148279	UGT2A1	0.70
A_23_P110234	CSN1S1	rs941389	UGT2B4	0.75
A_23_P362694	C4orf7	rs1432329	UGT2A1	0.75
A_23_P362694	C4orf7	rs941389	UGT2B4	0.78
A_23_P110234	CSN1S1	rs1560605	UGT2A1	0.78
A_23_P110234	CSN1S1	rs2288741	UGT2A1	0.79
A_23_P41365	SMR3A	rs1131878	UGT2B4	0.81
A_23_P110234	CSN1S1	rs844342	UGT2B10	0.81
A_23_P362694	C4orf7	rs10026603	UGT2A1	0.81
A_23_P41365	SMR3A	rs4694211	UGT2B4	0.84
A_23_P110234	CSN1S1	rs7439366	UGT2B7	0.85
A_23_P41365	SMR3A	rs1432329	UGT2A1	0.85
A_23_P362694	C4orf7	rs1560605	UGT2A1	0.85
A_23_P110234	CSN1S1	rs7668258	UGT2B7	0.86

Probe_ID	Gene exp	SNP_rs	SNP_gene	b1_p
A_23_P11023	CSN1S1	rs4521414	UGT2B7	0.86
A_23_P11023	CSN1S1	rs10026603	UGT2A1	0.88
A_23_P36269	C4orf7	rs4557343	UGT2B4	0.89
A_23_P41365	SMR3A	rs2288741	UGT2A1	0.90
A_23_P41365	SMR3A	rs10026603	UGT2A1	0.91
A_23_P36269	C4orf7	rs13139888	UGT2B4	0.92
A_23_P11023	CSN1S1	rs903446	UGT2B4	0.92
A_23_P41365	SMR3A	rs1560605	UGT2A1	0.94
A_23_P41365	SMR3A	rs4235126	UGT2B28	0.96
A_23_P11023	CSN1S1	rs4148279	UGT2A1	0.98

Paper III

Margarethe Biong, Matthew Suderman, Vilde D. Haakensen, Bettina Kulle, Paul R. Berg, Inger Torhild Gram, Vanessa Dumeaux, Giske Ursin, Åslaug Helland, Michael Hallett, Anne-Lise Børresen-Dale, Vessela N. Kristensen.

Candidate SNP analysis integrated with mRNA expression and hormone levels reveal influence on mammographic density and breast cancer risk.

Submitted to Cancer Epidemiology Biomarkers and Prevention.

III