UiT The Arctic University of Norway

Faculty of Health Sciences

A Systems Epidemiology Approach to Breast Cancer and Parity

The Norwegian Women and Cancer (NOWAC) Study

Sanda Krum-Hansen, MD

A dissertation for the degree of Philosophiae Doctor (Ph.D.) December 2023

Table of Contents

1 Background 1

1.1 Breast cancer, a global public health issue 1

1.2 Development of the mammary gland 2

1.3 Breast Cancer 4

1.4 Risk factors 7

Parity and breast cancer 7

Potential mechanisms for parity-induced protection against breast cancer 8

Breastfeeding 10

Body mass index and obesity 11

Menopausal hormone therapy (MHT) and hormonal contraceptives 11

Alcohol consumption 12

Smoking 13

1.5 Integrated systems epidemiology approach 13

1.6 Gene expression, microarray, PAM 50 14

The central dogma of molecular biology 14

Gene expression 16

Microarray technology 17

PAM 50 19

1.7 Aims 19

2 Material and methods 19

2.1 Study population 20

The Norwegian Women and Cancer Study (NOWAC) (Paper I) 20

Normal breast tissue biopsies (Paper II and III) 22

Breast cancer biopsies (Paper IV) 23

.... 25

Follow-up with Cancer Registry of Norway and Central Population Registry 26

2.2 Ethical approval 26

2.3 Variables 26

2.4 Statistical Methods 29

3 Results-summary of papers 30

3.1 Paper I 30

3.2 Paper II 31

3.3 Paper III 31

3.4 Paper IV ... 32
4 Discussion ... 32
4.1 General discussion ... 32
Study design and population ... 32
Validity of the variables ... 36
Technical consideration in gene expression analysis ... 37
4.2 Discussion of the main results ... 37
5 Main conclusion ... 40
6 Future perspectives ... 40
Works cited ... 42

Error! Bookmark not defined.

List of Figures

Figure 1: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for Breast Cancer in 185 Countries ... 1
Figure 2: Breast Cancer ... 5
Figure 3: Overview of the Central Dogma of Molecular Biology and Processes Involved in Relaying the Flow of Genetic Information ... 16
Figure 4: Schematic overview of cDNA microarray experiment ... 18
Figure 5: Enrollment to Norwegian Woman and Cancer Study ... 21
Figure 6. Flowchart depicts normal breast biopsy collected. ... 24
Figure 7. Flowchart depicts breast cancer biopsy collected. ... 25

Foreword and acknowledgements

A large portion of the work in this thesis was the year I spent collecting normal breast tissue from healthy women attending the mammography screening program at the Breast Diagnostic Center at the University Hospital of Northern Norway in Tromsø. It was my pleasure and privilege to meet these women who were so interested in contributing to our research. Rarely does a researcher have the opportunity to meet so many of the participants in her research. These encounters gave me priceless insight into the practical process of participating in a research project, as well as the women's motivations for doing so. Thank you to each and every woman who participated in our research projects! Without you, this work would have been impossible.

Thank you to the University of Tromsø for the opportunity to conduct this research, and to all my colleagues at the university who I have met along the way.

In particular, thank you to my three supervisors (Professor Ruth H. Paulsen, Professor emeritus Eiliv Lund and Associated professor Karina Standahl Olsen) for your guidance, understanding and especially your patience.

Thank you to my patients and colleagues at the hospitals in Harstad, Tromsø and Stavanger, who help me remember how every research project is a steppingstone to improving the lives of the patients we meet every day.

Thank you to all my friends who have supported me for all these years.

And finally, the biggest thank you to my husband, Niels, and to my mother, Emina. You are always there for me when I need you.

Summary

The incidence and prevalence of breast cancer (BC) are increasing worldwide. The reasons for this trend are partly unknown. BC has several well documented modifiable and non-modifiable risk factors. The only known, natural protective factors are breastfeeding and parity. The current paradigm is that the protective effect of parity is mediated by permanent genomic changes in a woman's breast tissue during her first pregnancy. Many gene expression studies examining breast carcinogenesis, have used a variety of non-cancerous, but not truly normal samples of breast tissue as controls. We performed a series of studies with a systems epidemiology approach to explore association between breast cancer and parity in a postmenopausal population.

The NOWAC study is a prospective national cohort study of 172 000 Norwegian women included between 1991 and 2007. Data is collected from questionnaires and linked national registries.

We found that the women in the Norwegian Women and Cancer Study had a decreasing cumulative incidence rate (CIR) of BC of 0.8% per child, rather than only for the first child. This finding was consistent irrespective of other risk factors and is in line with past studies.

We performed two gene expression studies using microarray technology in order to explore the biological processes underlying this protective effect. To ensure the best possible source of normal tissue, we collected breast tissue samples of 400 healthy postmenopausal women from the same NOWAC cohort. Gene expression profiles of 311 of these samples showed no genomic changes associated with parity, although we did find genomic changes associated with obesity, smoking and alcohol.

Next, we performed a nested case control study of 311 pairs of the same healthy cohort and women with breast cancer in the NOWAC study. We found no impact of parity on global gene expression levels in either cohort.

In conclusion we find that parity is a protective factor for BC for each additional child, not only the first full term pregnancy. However, we find no evidence that parity is associated with permanent genomic changes in normal breast tissue or breast cancer tissue.

Sammendrag (Summary in Norwegian)

Insidensen og prevelansen av brystkreft (BK) øker globalt. Årsakene til dette er ikke kjent. BK har en rekke veldokumenterte modifiserbare og ikke-modifiserbare risikofaktorer. De eneste kjente beskyttende faktorene er amming og paritet. Den rådende teorien for å forklare dette er at den beskyttende effekten av paritet er mediert av permanente genomiske endringer i en kvinnes brystvev i løpet av hennes første svangerskap. Mange genekspresjonsstudier som undersøker utvikling av brystkreft, har brukt ulike ikke-maligne, men ikke egentlige normale, prøver fra brystvev som kontroller. Vi utførte en serie studier med en systemepidemiologisk tilnærming for å utforske dette nærmere i en postmenopausal populasjon. NOWAC-studien er en prospektiv nasjonal kohortstudie av 172 000 norske kvinner inkludert mellom 1991 og 2007. Data samles fra spørreskjema og koblede nasjonale registre.

Vi fant at kvinner i Kviner og Kreft studien hadde reduksjon av kumulative insidens-rate for bryst kreft på 0.8% per barn, ikke kun for det første barnet. Dette funnet var konsistent uavhengig av andre risikofaktorer, og i tråd med tidligere studier.

Vi utførte to genekspresjonsstudier med mikromatrise-teknologi for å undersøke de biologiske prosessene bak denne beskyttende effekten. For å benytte den beste kilden til normalt vev, samlet vi brystvevsbiopsier fra 400 friske postmenopausale kvinner fra den samme NOWAC-kohorten. Genekspresjonsprofiler av 311 av disse biopsiene viste ingen genomiske endringer assosiert med paritet, men vi fant genomiske endringer assosiert med overvekt, røyking og alkohol.

Deretter utførte vi en nøstet kasus-kontroll-studie av 311 par fra den samme friske kohorten og kvinner med BK i NOWAC-studien. Vi fant ingen innvirkning av paritet på globale genekspresjonsnivå i noen av kohortene.

For å konkludere, finner vi at paritet er en beskyttende faktor for BK for hvert barn, ikke bare det første fullgåtte svangerskapet. Derimot finner vi ingen holdepunkter for at paritet er assosiert med permanente genomiske forandringer i normalt brystvev eller i brystkreftvev.

List of papers

This thesis is based on the four manuscripts, which are referred to in the text by their Roman numerals I-IV:

Paper I

Sanda Krum-Hansen, Arne Bastian Wiik, Karina Standahl Olsen, Marko Lukic, Ruth H. Paulssen, Eiliv Lund. Parity and cumulative incidence rates of breast cancer in the Norwegian Woman and Cancer Study (NOWAC).

Submission in progress

Paper II

Sanda Krum-Hansen, Karina Standahl Olsen. Practical and Ethical Issues in Establishing a Collection of Normal Breast Tissue Biopsies— Part of the NOWAC Post-Genome Cohort.

Published 2020 in: Lund, E. (Ed.) (2020). Advancing Systems Epidemiology in Cancer. Exploring Trajectories of Gene Expression. Chapter 4. Oslo: Scandinavian University Press. DOI: https://doi.org/10.18261/9788215041193-2020-04

Paper III

Sanda Krum-Hansen, Karina Standahl Olsen, Endre Anderssen, Jan Ole Frantzen, Eiliv Lund and Ruth H. Paulssen. Associations of breast cancer related exposures and gene expression profiles in normal breast tissue – The Norwegian Women and Cancer (NOWAC) normal breast tissue study.

Published January 2023 in Cancer Reports DOI: 10.1002/cnr2.1777 ***PMID: 36617746***

Paper IV

Eiliv Lund, Sanda Krum-Hansen, Karina Standahl Olsen, Nikita Shvetsov· Igor Snapkow, Oxana Gavriluk, Lill-Tove Rasmussen Busund, Jan Ole Frantzen, Marit Holden, Lars Holden

No impact of parity on global gene expression levels in breast cancer tissue and normal breast tissue – a nested case-control study in the NOWAC Postgenomic biobank.

Submitted to BMC Cancer

List of abbreviations

BC	Breast cancer
BK	Brystkreft
BMI	Body mass index
CIR	Cumulative incidence rate
EFTP	Early full-term pregnancy
ER	Estrogen receptors
FFTP	First full-term pregnancy
GO	Gene ontology
IQR	Inter quantile range
LIMMA	Linear models for microarrays
Lob	Lobule
MHT	Menopausal hormone therapy
NOWAC	The Norwegian Women and Cancer Study
PAM50	Prediction Analysis of Microarray 50
PR	Progesterone receptor
ROS	Reactive oxygen species
SHBG	Sex- hormone binding globulin
TDLUs	Terminal-duct lobular units
WHO	World Health Organization

1 Background

1.1 Breast cancer, a global public health issue

Breast cancer (BC) is the predominant cancer in women worldwide. In 2020, the BC incidence became the leading cause of global cancer. Yearly, there are an estimated 2.3 million new BC cases, accounting for 11.7 % of new cancer cases worldwide. BC is the fifth leading cause of cancer mortality worldwide, with 685,000 yearly deaths, after lung, colorectal, liver, and stomach cancer (1). It is also the leading cause of cancer death in women worldwide, responsible for 1 in 6 cancer deaths (1).

Incidence rates of BC are higher in developed countries than in developing countries, 55.9 and 29.7 per 100,000, respectively. In contrast to this, mortality rates are higher in developing countries than in developed countries (15.0 and 12.8 per 100,000) (Figure 1).

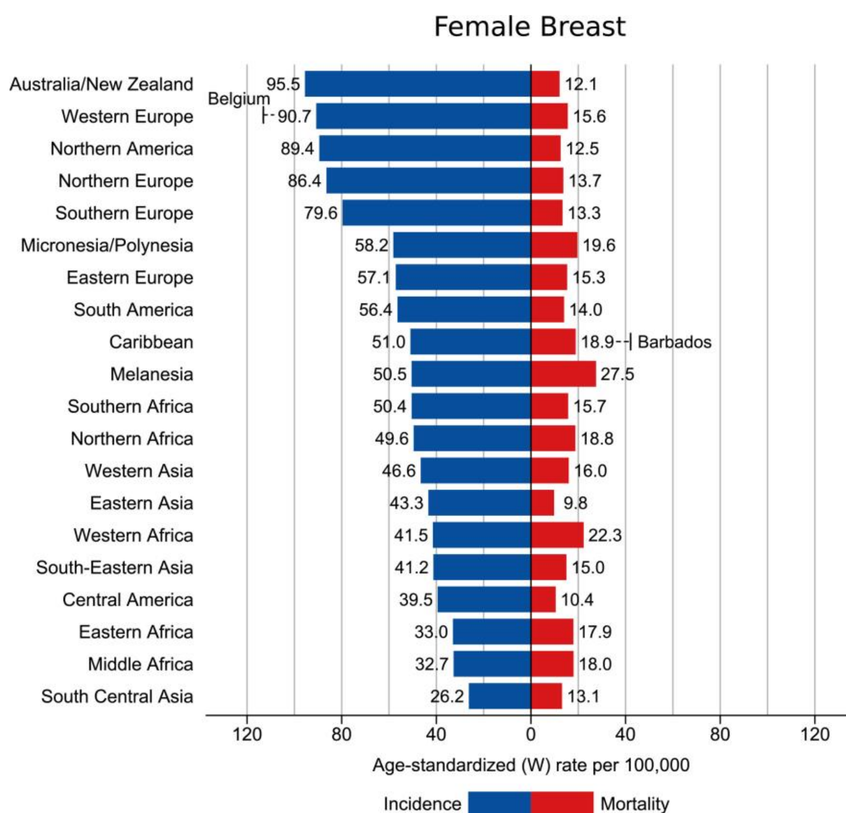


Figure 1: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for Breast Cancer in 185 Countries

CA A Cancer J Clinicians, Volume: 71, Issue: 3, Pages: 209-249, First published: 04 February 2021, DOI: (10.3322/caac.21660)

The high incidence of BC in developed countries reflects a risk profile with increased occurrence of reproductive and hormonal risk factors such as high age at first birth, lower parity, less breastfeeding, as well as lifestyle risk factors (alcohol intake, obesity, physical inactivity). Earlier diagnoses through mammography screening also contributed to a higher incidence. Higher mortality rates in developing countries reflect undeveloped health infrastructure without organized screening programs. Consequently, many women present with higher stage disease at diagnosis and do not have access to modern treatment.

The incidence rates of BC are rising globally, including in developing countries. An increasing female workforce in growing economies changes the women's risk profile towards that of women in developed countries, with late first pregnancy and fewer births, as well as increasing obesity and sedentary lifestyle (1).

With the increasing incidence of both BC and the prevalence of risk factors, BC will continue to be a global public health care issue. Modern treatment options available in developed countries reduce mortality but increase the number of women living with BC and the late complications of treatment. This will contribute to even more BC related morbidity in the future.

Given these factors, there is a need to develop better strategies for preventing and treating BC. Research on parity-induced protection against BC can potentially contribute to this development. It is well known that early first full-term pregnancy before age 20 or 25 significantly reduces BC risk. Each additional pregnancy reduces a woman's risk of developing BC by another 7%. On the other side, nulliparity or a first full term pregnancy later in life (after age 35) increases BC risk (2-5). A complete understanding of the physiological processes involved in decreased BC risk in early parous women may lead to identifying biomarkers to predict the risk for BC development, especially in women who are at high risk. This knowledge can also help develop more effective preventive and treatment strategies.

Unfortunately, the precise molecular mechanisms involved in these changes and the significance of these observations in protecting against mammary carcinogens is still poorly defined. To understand these factors in depth, it is essential first to understand physiology of the normal breast tissue, the type and stages of BC, and the risk factors of the disease.

1.2 Development of the mammary gland

The mammary gland is a unique organ not completely developed at birth. The mammary gland evolved to produce and store milk to provide foodstuff for the newborn. The structure

of the human breast reflects this development and function. Milk production and storage occur in the apocrine glandular tissue, surrounded, and supported by connective tissue (stroma). The areola and nipple on the skin mark the outflow from the internal ducts, allowing milk to exit and providing the site at which the child may suckle.

The stroma of the breast, surrounding the glandular tissue, contains white fat and fibrovascular tissue. The quantities of fat to fibrovascular tissues vary from woman to woman and over time within individuals. The fibroblasts in the stroma produce growth factors necessary to support and nourish apocrine cells and – as regulators of the stromal microenvironment - are essential contributors to tumour development in breast cancer (6).

The functional glandular tissue of the breast consists of milk-producing apocrine parenchymal tissue. This parenchyma is divided into 15-25 lobes, each comprising 20-40 lobules. Milk is produced in terminal-duct lobular units (TDLUs), drained by terminal collecting ducts into 4-18 lactiferous ducts. These finally drain to the nipple itself. Lobular units in normal breasts are composed of four main lobular structures, numbered type 1 through 4, and referred to as Lob 1-4, based on their degree of development. These subtypes represent a gradual differentiation of alveolar buds, transitioning from Lob 1 to 2 and Lob 2 to 3. Lob 1 is the dominant subtype in breasts of nulliparous women, while Lob 2 and 3 can occasionally be apparent during early reproductive age (7). Lob 3 numbers increase significantly in parous women with early full-term pregnancy and remain the prominent structure during the reproductive years. Lob 3 decreases in number due to regression of the breast after the fourth decade of life, and again after menopause, when they involute to Lob 2 and Lob 1. Lob 4 represents the maximal expression of lobular development and cell differentiation in the adult human mammary gland, occurring at the time of lactation (7, 8). After menopause, both nulliparous and parous women have Lob 1 as the dominant subtype in the breast. Despite morphological similarity, Lob 1 in nulliparous and Lob 1 in parous women differ in their susceptibility to carcinogenesis. Several studies have explored these differences (9-11).

As demonstrated by the varying presence of these lobular subtypes, the composition of the breast tissues changes during a woman’s lifetime. Six stages are defined by these changes: fetal or embryonic, pre-pubertal, pubertal, pregnancy, lactation, and involution. This lobular development is driven by estrogens and progesterone, which modulate the proliferation of the breast epithelium and, thus, the differentiation of lobules through the subtypes mentioned above. Some data suggest that the degree of breast tissue differentiation is an important

determinant in the expression of estrogen receptors (ER) and progesterone receptors (PR). It is theorized that the percentage of cells expressing these receptors varies as a function of lobular development and, therefore, of the type of lobular structure analyzed. Lob 1 has a higher percentage of ER-alpha and PR-positive cells than lob 2,3 and 4. The differentiation of lob 1 to later subtypes (either in response to hormones or other physiological triggers during early full- term pregnancy) is associated with lower cell proliferation, lower quantities of ER and PR, and more efficient DNA repair. This appears to generate a stem cell in the differentiated breast that is more refractory to carcinogenesis. This process depends upon pregnancy, so the final stages of development will never occur if a woman remains nulliparous.

The glandular tissue atrophies progressively during menopause, reducing lobule size and number – primarily through the involution of the milk-producing acini. The structure of the stromal tissue also undergoes predictable changes, with fibrovascular tissue being replaced by adipocytes. These changes, like menopause itself, are hormonally driven (12). Thus, the extensive modern use of hormone-replacement therapy has altered considerably the appearance of postmenopausal breast tissue (7).

1.3 Breast Cancer

BC is not a single disease but a group of heterogeneous disorders in the breast. BC has been classified into various subtypes using histological and molecular characteristics. This classification continues to improve with advances in cancer research, leading to better treatment and prognosis. BC is both etiologically and genetically a heterogeneous disease. It consists of several sub-types with different molecular profiles and biological and clinical behaviour. Different subtypes are associated with varying risk profiles and present challenges in both understanding breast carcinogenesis and clinical management (13). An overview of the different BC subtypes is presented in Figure 1.

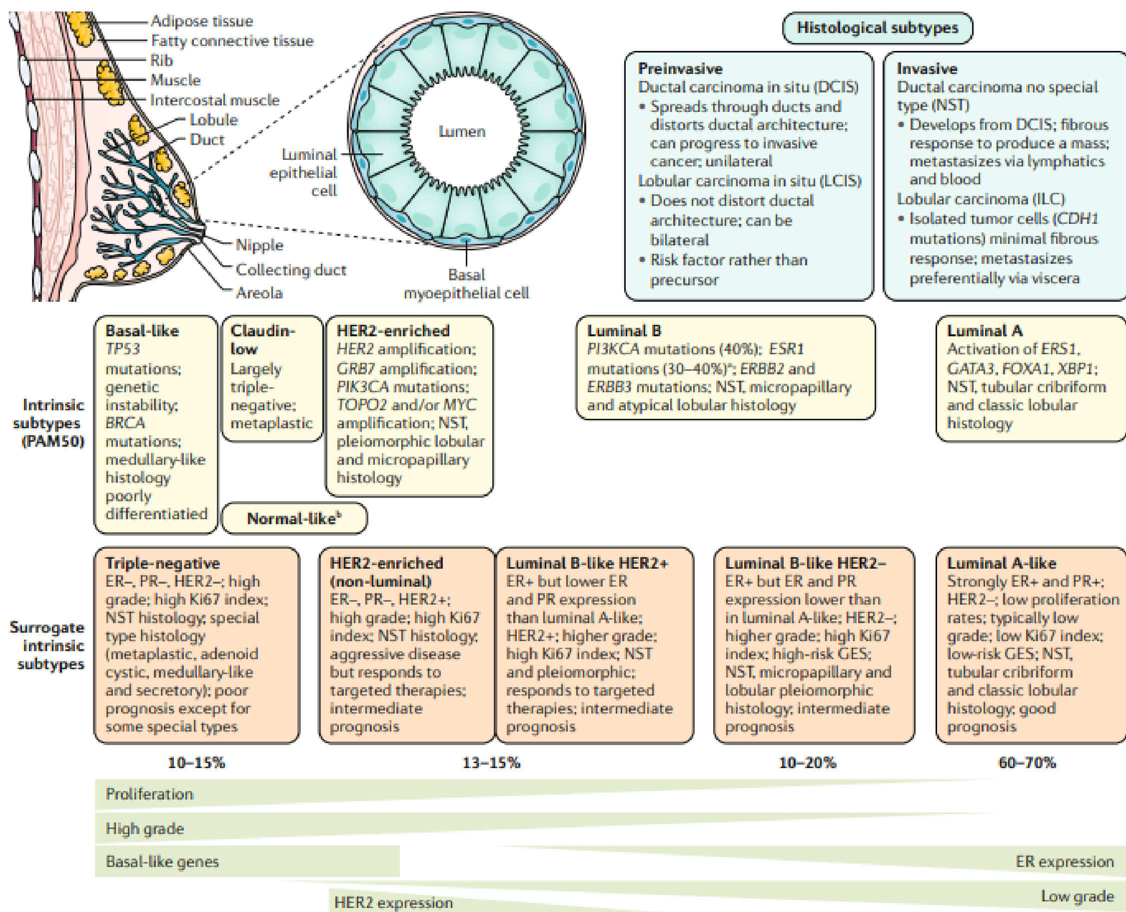


Fig. 1 | **Breast cancer.** All breast cancers arise in the terminal duct lobular units (the functional unit of the breast) of the collecting duct. The histological and molecular characteristics have important implications for therapy, and several classifications on the basis of molecular and histological characteristics have been developed. The histological subtypes described here (top right) are the most frequent subtypes of breast cancer; ductal carcinoma (now referred to as 'no special type' (NST)) and lobular carcinoma are the invasive lesions; their preinvasive counterparts are ductal carcinoma in situ and lobular carcinoma in situ (or lobular neoplasia), respectively. The intrinsic subtypes of Perou and Sorlie[1] are based on a 50-gene expression signature (PAM50)[121]. The surrogate intrinsic subtypes are typically used clinically and are based on histology and immunohistochemistry expression of key proteins: oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and the proliferation marker Ki67. Tumours expressing ER and/or PR are termed 'hormone receptor-positive'; tumours not expressing ER, PR and HER2 are called 'triple-negative'. The relative placement of the boxes align with the characteristics (for example, proliferation and grade) in green. –, negative; +, positive. GES, gene expression signature. [a]*ESR1* mutations induced by aromatase inhibitor targeted therapy. [b]Artefact; expression of normal breast components due to low tumour cellularity.

NATURE REVIEWS | DISEASE PRIMERS | Article citation ID: (2019) 5:66 3

Figure 2: Breast Cancer

Illustration used with permission, Springer Nature and Copyright Clearance Center.
License Number 5443120112809

There are different types of BC depending on the specific structures of the breast that are affected, including ducts, lobules, or stroma. BC can be broadly divided into sarcomas and carcinomas based on the cell of origin. Sarcomas arise from stroma consisting mainly of myofibroblasts and blood vessel cells. Sarcomas are rare BC types, representing less than 1% of cases, and are not the topic of this work.

Most breast cancers are carcinomas, particularly adenocarcinomas, originating in glandular tissue. Carcinomas arise from the epithelial cells lining the lobules and terminal duct. There are several different types of breast carcinomas and treatment options, and prognosis depends on pathological features and invasiveness (13).

Based on these criteria, BC can be divided into three major groups: non-invasive (in situ), invasive, and metastatic BCs. Non-invasive, premalignant breast lesions are relatively common. Ductal carcinoma in situ (DCIS) is a relatively common, non-invasive BC that develops inside the normal duct without breaching through basement membrane. This condition is not included in the cancer definition of our studies and are therefore not discussed further.

In invasive BC, cells have spread outside the normal lobules and ducts, infiltrating stromal tissue. Invasive BC can be classified as invasive carcinomas of no special type (formerly ductal carcinoma), the most commonly diagnosed type of BC, invasive lobular carcinoma and the other carcinomas of special type (tubular, medullary, mucinous, papillary, and cribriform carcinoma of the breast) (13-15). The majority of BC are carcinomas, particularly adenocarcinomas that originate in the duct or lobules.

BC may be divided into molecular subtypes based on gene expression profiles (16, 17). Luminal A and Luminal B are the two intrinsic subtypes expressing estrogen receptors. Basal-like and human epidermal growth factor receptor 2 (HER-2) enriched subtypes, in most cases, do not express steroid hormone receptors and are more aggressive subtypes. This classification predicts treatment choices as well as response and prognosis.

The treatment strategies for breast cancer are rapidly developing. Considering biological aspects and heterogeneity of disease, modern treatment algorithms give women an improved chance of survival and reduce the adverse effects of treatment. Prosigna test is an example of use of research results in treatment decisions. Prosigna is commercially available test based on PAM50 gene signature (18). Prosigna test identify patients in low-risk group of developing distant recurrence that can avoid overtreatment (19, 20).

Surgery is the cornerstone of curative breast cancer treatment. Surgical treatment, breast-conserving, or mastectomy is often combined with radiation therapy and some systemic treatment before and/or after surgery. Systemic treatment includes chemotherapy, endocrine treatment, and targeted therapy. The Norwegian Breast Cancer Group (21) develops treatment guidelines published by the Norwegian Directorate for Health.

1.4 Risk factors

Many different factors are recognized/established as risk factors for BC. One way to categorize them is into non-modifiable (we cannot do anything about them), semi-modifiable (intervention is not feasible or is unethical), and modifiable risk factors (we can do something about them). Research on risk factors for BC can be challenging because of the complexity of their relations to BC and the heterogeneity of the disease. Different exposures have different effects on the risk of BC development depending on, for example, menopausal status (before or after menopause), when in her lifetime a woman is exposed (for instance, before or after first full-term pregnancy), and what type of BC the woman develops (importance of different risk factors varies by molecular subtypes of BC) (22).

Timing and length of exposure to risk factors are important. BC carcinogenesis takes time, and breast tissue is more susceptible to developing BC at different periods of life.

The primary exposure/risk factor in this work was parity. We naturally included variables directly connected to parity, such as age of the mother at first birth, breastfeeding, and number of children. Further, we included the main modifiable risk factors conditioned by lifestyle as alcohol consumption, smoking, obesity, and menopausal hormone therapy (MHT).

Parity and breast cancer

Parity, early age at first full-term pregnancy (FFTP), and prolonged breastfeeding are some of the factors associated with reduced risk of breast cancer.
Bernardino Ramazzini observed the association between parity and BC in a treatise on occupational diseases published in the 1700s, where childless nuns were reported to have a higher risk of BC (23). A similar observation was reported in 1842 (23, 24). Since then, several extensive epidemiological studies have investigated the association between parity and risk of BC. They have concluded that women who gave birth before the age of 20 or 25 have a significantly reduced risk of developing BC compared with women who never gave birth and compared to those who delivered their first child aged 35 or older (2-5). Women who had their FFTP between 30 and 34 years of age have the same risk of developing BC as nulliparous women. The risk increases even further after 35 years. Husby *et al.* showed that a pregnancy must last 34 weeks or longer to observe this protective effect (25). It must be noted, however, that this protective effect of early first-term pregnancy is not immediate: the

risk of breast cancer undergoes a transient increase after delivery, with the risk decreasing over 5-10 years (26, 27).

Women with a first full-term pregnancy after the age of 30 have the highest transient risk of breast cancer post-delivery (28). This short-term adverse effect of pregnancy, its amplitude, and timing is modulated by the age of the mother at the first and the last birth, the number of pregnancies, the time between adjacent pregnancies, and the time since the last pregnancy. FFTP is the most crucial determinant of the dual effect of pregnancy. Subsequential full-term pregnancies (FTP) executes a similar effect, yet a long-term decrease follows a quantitative, less substantial, short-term increase in BC risk. Lower age at FFTP and subsequential pregnancies and shorter intervals between adjacent pregnancies showed greater long-term protection. Comparing uniparous to nulliparous women, the short-term increase in BC risk peaks around 5 years after FFTP, while long-term protection is fully attained within 15 years. Epidemiological studies also demonstrate a further reduction in lifetime BC risk with multiple full-term pregnancies, with each additional pregnancy reducing the risk by approximately 7%. However, a short-term increase in risk precedes this protective effect. The magnitude and timing of this adverse effect differs by maternal age at delivery, time between pregnancys and pregnancy number.

Potential mechanisms for parity-induced protection against breast cancer

Though the long-term effect of early full-term pregnancy (EFTP) on the risk of BC development has been known for centuries, the mechanisms by which this protection occurs are poorly understood. Several theories have been proposed (29-32).

Hormonal Changes: Pregnancy triggers significant hormonal shifts, including elevated estrogen, progesterone, and placental hormones. These changes are believed to influence breast tissue, reducing the risk of cancer development.

Cellular Differentiation: Pregnancy induces differentiation in breast cells, making them more specialized for milk production. This process may leave cells less susceptible to cancer, as less differentiated or more stem-like cells are thought to be more prone to malignancy.

Cellular Senescence: Pregnancy may induce cellular senescence, where cells cease to divide. Senescent cells are less likely to become cancerous, potentially contributing to breast cell protection during pregnancy and lactation.

Immune System Modulation: Pregnancy modulates the immune system to tolerate genetically different fetal cells. This immune modulation may extend to cancer surveillance, offering protection against breast cancer development.
Changes in Breast Microenvironment: Pregnancy induces alterations in the breast microenvironment, including changes in blood flow and the extracellular matrix. These changes may create an environment less conducive to cancer development.
Genetic and Epigenetic Factors: Genetic and epigenetic factors associated with pregnancy could contribute to the protective effect. Upregulation or downregulation of certain genes and epigenetic modifications as methylation, may alter breast cancer risk.

The terminal differentiating of the breast epithelia.

Full-term pregnancy (FFTP) induces terminal differentiation of the breast epithelium, considered a protective factor against breast cancer (BC). This differentiated state persists post-lactation, leaving the mammary tissue less proliferative and more resistant to carcinogenesis. Involution following lactation induces apoptosis in differentiated lobules, potentially eliminating carcinogenic cells and reducing BC risk. Genome-wide studies have identified a specific genomic signature associated with the differentiated state of the parous breast (9). However, in animal study (33) differentiation induced by medication did not showed the same protective effects, suggesting another mediator beyond differentiation itself.

Prolactin is hormone crucial for breast tissue differentiation and is controlled by dopamine from the hypothalamus. Studies using the dopamine receptor antagonist Perphenazine revealed mammary gland differentiation but failed to confer protection against carcinogenesis (34). This indicates that differentiation alone is insufficient for protection. Conversely, short-term estrogen treatment, without inducing complete differentiation, effectively inhibits mammary carcinogenesis (35). Other studies (25, 36) suggest that the protective effect of FFTP is unlikely to be solely explained by terminal differentiation.

It's important to note that while parity is linked to a reduced overall risk of breast cancer, this relationship can be influenced by factors such as age at first pregnancy, the number of children, and breastfeeding. Additionally, the protective effect may vary across different subtypes of breast cancer.

Further exploration of the molecular and cellular mechanisms underlying the protective effects of parity against breast cancer is essential. This understanding is necessary for developing targeted preventive and therapeutic strategies.

Breastfeeding

Breastfeeding is a potentially modifiable risk factor for breast cancer. An extensive review was conducted by Lipworth *et al*., including studies from 1966 through 1998. It showed a relatively weak protective effect on BC, limited to premenopausal women (37, 38). Most of the included studies had a case-control design. Yang *et al.* continued with a systematic review of 31 epidemiological studies published between 1999 and 2007. Their findings were inconsistent regarding the risk reduction in BC associated with breastfeeding or cumulative breastfeeding duration (39). A meta-analysis conducted by the Collaborative Group on Hormonal Factors in Breast Cancer, including 47 epidemiological studies from 30 countries, showed a risk reduction of 4.3% for every 12 months of breastfeeding (40).
Several recent studies reviewed by Islami et.al. (41) explored combined effect of parity and breastfeeding. In 2019, Fortner *et al*. conducted a large prospective study investigating the combined effects of parity and breastfeeding on breast cancer risk, considering receptor status and molecular phenotype (42). Their findings revealed that parous women exhibited a lower risk of estrogen receptor-positive (ER+) breast cancer compared to nulliparous women. However, no significant association was observed for estrogen receptor-negative (ER-) breast cancer. Among parous women, those who breastfed had a reduced risk of ER- breast cancer, but no such association was found for ER+ breast cancer.
Examining the molecular phenotype, the study found that increased parity in multiparous women was inversely associated with Luminal B breast cancer, regardless of breastfeeding. Conversely, comparing nulliparous with multiparous women, increased parity had an inverse association with Luminal A breast cancer in breastfeeding women. Basal-like breast cancer showed an increase in women with higher parity who did not breastfeed, while no association was found in women who never breastfed. These findings highlight the nuanced interactions between parity, breastfeeding, receptor status, and molecular subtypes in influencing breast cancer risk.
The reason for lactation-induced risk reduction is not understood. Several mechanisms have been postulated. These include hormonal changes, such as reduced estrogen; removal of estrogens through breast fluid; excretion of carcinogens from breast tissue through breast-feeding; physical changes in the mammary epithelial cells reflecting maximal differentiation;

and delay of the re-establishment of ovulation (43). So, despite several plausible biological mechanisms behind the reduction of risk for BC by breastfeeding, findings in epidemiological studies are inconclusive.

Body mass index and obesity

Overweight and obesity are increasing problems for public health globally (44, 45). Increasing prevalence of obesity is observed also in Norway (46, 47). However, the association between high body-mass-index (BMI) and BC risk varies, primarily in relation to menopausal status. Overweight, obesity, and also adult weight gain increases the risk of hormone receptor positive BC in postmenopausal women (48, 49). In contrast, high BMI in childhood and adolescence shows a negative association with the risk of hormone receptor positive BC in premenopausal women (50, 51). However, truncal obesity may increase the risk ofa triple-negative BC, more common in premenopausal women (52).

There are several mechanisms, affecting both hormonal and inflammatory pathways, involved in obesity related breast carcinogenesis (53).

Proliferation of BC cells can be mediated through various hormonal changes. Insulin resistance, with elevated insulin and IGF-1 level, promotes proliferation, and inhibits cell apoptosis, but also stimulate estrogen dependent carcinogenesis in the breast. IGF-1 stimulated aromatase activity and increase levels of estrogen and insulin reduces synthesis of sex hormone-binding globulin (SHBG) and increases estradiol level (54). Adipose tissue is an active endocrine and metabolic organ that secrets several bioactive molecules in the surrounding microenvironment. They are called adipokines and they can induce low-grade systemic inflammation (55-57).

Menopausal hormone therapy (MHT) and hormonal contraceptives

Exposures to exogenous hormones, MHT, and oral contraceptive lead to a transient increased risk of BC during use, and 2 years, and 10 years after cessation, respectively (58, 59). Current users of MHT have a higher risk of developing BC than never users (60, 61) . Risk increases with extended duration of use (62) and is higher for users of combination MHT containing both estrogen and progesterone compared to estrogen alone. The risk associated with MHT is influenced by the age at the start of use, time from menopause to first use and BMI (59). The risk associated with MHT use reduces after MHT is stopped and disappears after 5-10 years, depending on the duration of use (63, 64). The recurrence rate of BC is higher for MHT users,

especially in those with lymph node negative tumors and patients who were not treated with aromatase inhibitors (65, 66).

We did not include exposure to hormonal contraceptives in our studies. Regarding parity, we can speculate that hormonal contraceptives prolong nulliparity and decrease the lifetime numbers of deliveries. Indirectly, contraceptives can, therefore, increase the risk of BC by postponing pregnancy and raising the age at first pregnancy.

Alcohol consumption

Alcohol intake is established as a risk factor of BC, with causal effect (67, 68). Extensive epidemiological data have consistently found an increased risk of BC with increasing alcohol intake, for both pre- and post-menopausal BC (69-72). A moderate alcohol intake, defined as 10 g of alcohol per day, increases risk of BC by 7-9%, and a higher intake of alcohol increases risk further in an approximately linear fashion (73-75). Alcohol consumption, even in small amounts, > 1 drink per day, increases the risk of BC (76). In a meta-analysis, using pooled data from 113 individual studies, Seitz et al. showed a modest but significant increase of 5% in the risk of BC in light drinkers (<1 drink per day) compared with non-drinkers (77). Alcohol consumption is closely associated with smoking, and this association can affect results by possible confounding between these two exposures. In the study conducted by Hamajima et al., they concluded that light to moderate drinking was associated with increased risk of BC in women who never smoked (67).

The possible underlying mechanisms for alcohol mediated BC carcinogenesis are complex and not fully understood. Several potential mechanisms explaining the association between alcohol use and initiation and progression of BC have been proposed (77, 78). Alcohol may induce carcinogenesis through several mechanisms, some of them involving the metabolism of alcohol in the body. Alcohol is metabolized by alcohol dehydrogenase, the enzyme which converts alcohol into acetaldehyde. Alcohol dehydrogenase is expressed in human epithelial cells in the breast, enabling alcohol metabolism (79). Acetaldehyde is a carcinogen and can accumulate in fat tissue of the breast. It can damage DNA in cells and cause gene mutation and induce carcinogenesis (80-83). Alcohol can be metabolized in breast tissue producing ROS (reactive oxygen species or free radicals) causing oxidative stress (84). ROS are known to cause chromosomal aberrations, DNA damage, chromosomal aberrations and mutations that inactivate tumor suppressor genes or increase expression of proto-oncogenes (85)

potentially inducing tumorigenesis. ROS and oxidative stress also have the potential to promote breast cancer aggressiveness (78). Additionally, alcohol can induce abnormal DNA methylation, which is important for tumor initiation and progression. Alcohol can increase the level of estrogen in the body, and high levels of estrogen have been linked to an increased risk of BC (72, 86, 87).

Smoking

Many epidemiological studies have produced a large body of evidence that places smoking as an emerging risk factor of BC (88-90). High accumulated exposure and high intensity of smoking, particularly in premenopausal women, has been linked to higher risk of BC (91). In addition, women who initiated smoking in early age and before first full-term pregnancy, are at higher risk of BC then never smokers (92, 93).

The exact mechanisms by which smoking increases risk of BC are not fully understood, but it is thought to involve the chemicals in tobacco smoke. Of several thousand chemicals in tobacco, 69 are known carcinogens. There is biological evidence that some of them can be taken up in breast tissue (94, 95). There are several biological mechanisms that can explain the association between smoking and increased risk of BC, but most of the evidence is from laboratory studies on cell lines (96).

Despite a large amount of epidemiological data, the association between smoking and risk of BC is still an ongoing debate (68, 97, 98). There are several reasons for this. A significant problem in epidemiological studies regarding smoking and BC is confounding, especially by alcohol consumption, which is positively associated with smoking. Smoking history is often collected by self-reporting, which may be unreliable. Current smokers are less likely to attend mammography screening (99), and consequently the number of undetected cancers in a population of unscreened active smokers may be high. Finally, the association between smoking and BC is most likely weaker than for other smoking related cancers.

1.5 Integrated systems epidemiology approach

Traditional epidemiological methods were developed to estimate the occurrence of disease in a population. This allows the discovery of associations between disease and exposures that might predispose to disease. All this assists in improving public health management through targeted primary and secondary messaging strategies to those at risk. Such approaches,

however, do not lend themselves well to studying the mechanisms of disease, and often incapable of taking advantage of growing knowledge around the biological underpinnings of revealed associations. Epidemiology becomes, in this fashion, a sort of “black box” (100) wherein the internal functioning of the process under study is essentially unknowable.

In contrast, molecular science focuses on how specific single genes or proteins can influence a biological phenotype. High-throughput molecular biological technologies – genomics, proteomics, epigenomics – have opened a revolution of “multilevel omics approaches” and led to the spread of integrated “systems” disciplines, such as systems biology and systems immunology. A system, in such a paradigm, is a set of interacting parts together with their relationships, which can be examined to identify the processes that maintain them (101). Systems approaches in biology, allow research to focus on the complex signaling networks that exist between cells, stroma, and organs, to determine how changes in such networks affect the organism as a whole. Such an approach allows us to study the manners in which the intracellular environment and the normal cell change during carcinogenesis, and so potentially improve early diagnostics and develop new treatment methods.

Taken in this context, the NOWAC study (102) is an example of a systems epidemiology approach. It represents a large cohort with a large sample size, allowing the examination of association between a chosen disease and exposures to relevant risk factors. It also includes biological samples collected during different periods of the adult life of participants. These open the possibility of analyzing biological material on the high dimensional gene expression level according to available exposure data. This can potentially aid us in examining how risk factors affect gene signatures in both case subjects with, for example breast cancer, and disease-free control subjects, in a nested case-control study. Such an integrated, multi-level systems epidemiological approach may represent an important tool for filling in the gaps in our knowledge of the multifactorial nature of mammary carcinogenesis (36, 103).

1.6 Gene expression, microarray, PAM 50

The central dogma of molecular biology

DNA and RNA are composed of nucleotides chains, long strings of sugars – either ribose in the case of RNA, or deoxyribose in the case of DNA - bound by phosphate groups, with an attached base. There are four possible bases in both DNA and RNA, though the bases

themselves may differ, with bases being either purines (adenine and guanine), or pyrimidines (cytosine and thymine in DNA, and replacing thymine in RNA). Purines and pyrimidines exhibit complementary bonding between strands of DNA, with adenine binding thymine and guanine b DNA and RNA are composed of nucleotides chains, long strings of sugars – either ribose in the case of RNA, or deoxyribose in the case of DNA - bound by phosphate groups, with an attached base. There are four possible bases in both DNA and RNA, though the bases themselves may differ, with bases being either purines (adenine and guanine), or pyrimidines (cytosine and thymine in DNA, and replacing thymine in RNA). Purines and pyrimidines exhibit complementary bonding between strands of DNA, with adenine binding thymine and guanine binding cytosine. It is this complementarity which allows the base sequence to act as an information molecule: after RNA polymerase bonds to an open DNA strand during the transcription process, the base sequence of the resultant messenger RNA (mRNA) molecule will be complementary to the original DNA. Where the DNA sequence had adenine, the mRNA will have guanine, and so on. When the mRNA is translated by the ribosome, each sequence of three bases will bind a transfer RNA (tRNA) with an attached amino acid. The ribosome thus produces a protein from these amino acids (Figure 3). This process of DNA transcription to RNA, and RNA translation to protein is referred to as the central dogma of molecular biology and is the basis of the science of genetics (104, 105). Understanding this relationship between the base sequence, base-pair complementarity, and RNA as an intermediate between DNA and protein, is essential to the understanding of modern DNA analytic techniques.inding cytosine. It is this complementarity which allows the base sequence to act as an information molecule: after RNA polymerase bonds to an open DNA strand during the transcription process, the base sequence of the resultant *messenger* RNA (mRNA) molecule will be complementary to the original DNA. Where the DNA sequence had adenine, the mRNA will have guanine, and so on. When the mRNA is translated by the ribosome, each sequence of three bases will bind a *transfer* RNA (tRNA) with an attached amino acid. The ribosome thus produces a protein from these amino acids (Figure 2). This process of *DNA transcription* to RNA, and RNA *translation* to protein is referred to as the *central dogma* of molecular biology and is the basis of the science of genetics. Understanding this relationship between the base sequence, base-pair complementarity, and RNA as an intermediate between DNA and protein, is essential to the understanding of modern DNA analytic techniques.

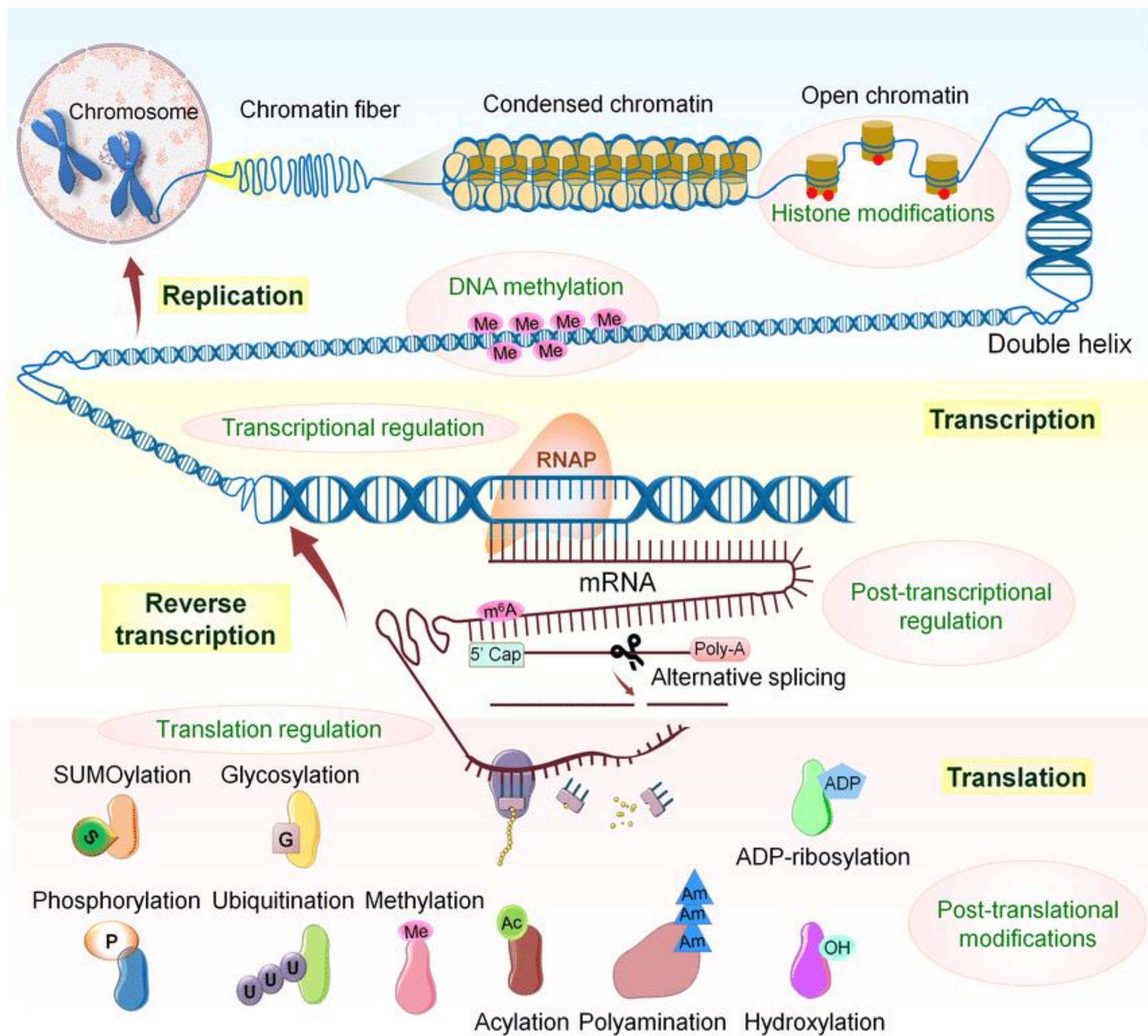


Figure 3: Overview of the Central Dogma of Molecular Biology and Processes Involved in Relaying the Flow of Genetic Information

Molecular Plant DOI:10.1016/j.molp.2020.22.002

Gene expression

Not all cells express all genes; cells differ in function, and this is entirely due to the differing repertoires of protein produced within the cells. To transcribe the proteins necessary for its function, a cell in one tissue will therefore have different regions of the genome available for transcription than a cell in a dissimilar tissue. An apocrine cell responsible for lactation will require different proteins than a fibroblast in the stroma. Changes in gene expression in response to cellular stimuli – for example, the introduction of growth factors or other signaling molecules – allows the cell to adapt to changing conditions or demands from the

world outside its cell membrane (predominantly, alterations in the extracellular matrix environment or stroma). Of interest to those studying disease, such stimuli can include risk factors that might lead to disease for the organism, including risk factors that might damage or alter expression of the genes.

Analysis of this gene expression can – based on the central dogma – be performed at several steps. One might examine the DNA itself, the messenger RNAs produced by the active DNA, or the proteins present in the cytosol of the cell. Messenger RNAs are a useful measure of DNA activity and gene expression, as they represent the genes actively being transcribed within the nucleus (106). Since they are actively transported from the nucleus to the cytosol, they are also more readily accessed than nuclear components. The base sequence any mRNA is a complementary reflection of the base sequence of the gene that produced it. This allows us to gain insight into genomic variation, using probes which bind either DNA or RNA. Such a probe is, simply, a complementary DNA or RNA (cDNA or cRNA) sequence bound to a labelling molecule. The label is typically a small molecule with a specific, high affinity for another molecule that is either fluorescent, allowing visualization (such as in enzyme-linked immunosorbent assay), or bound to a medium allowing separation of the information molecule from its environment. A typical system involves probes attached to biotin, which will form a complex with streptavidin or avidin. The latter molecules might be bound to fluorescent substances for visualization or quantification, or to magnetic beads, allowing later separation.

Microarray technology

Microarrays are a high-throughput variant of the above-mentioned probe-label analytic techniques, giving us the capability of measuring the expression of large numbers of genes simultaneously (Figure 2). Analysis typically involves several steps:

Microarray construction: Probes are prepared using cDNA fragments or oligonucleotides complementary to a set of coding and non-coding genes. A given platform may be capable of analyzing around 40,000 genes in parallel. These probes are then spotted onto a solid substrate such as membranes or glass slides.

Sample preparation: Blood or tissue to be analyzed must be collected, the mRNA isolated and purified, and pcDNA or cRNA synthesized from it. This cDNA or cRNA may be fluorescent labelled, radiolabeled or bound to a separation medium such as a resin bead medium.

Hybridization: At this step, selective complementary binding occurs between the labelled probes and the study samples, and unbound, unwanted material is washed away.

Analysis: RNA signal intensity/mRNA abundance is quantified via a chemiluminescence detector. Intensity of signal correlates with the level of gene expression. Signal data is then appropriately analyzed.

For the purposes of this thesis, mRNA was obtained from human breast tissue, probed with biotin-labelled cRNA hybridized to bead chips, and analysed using Illumina BeadArray Reader.

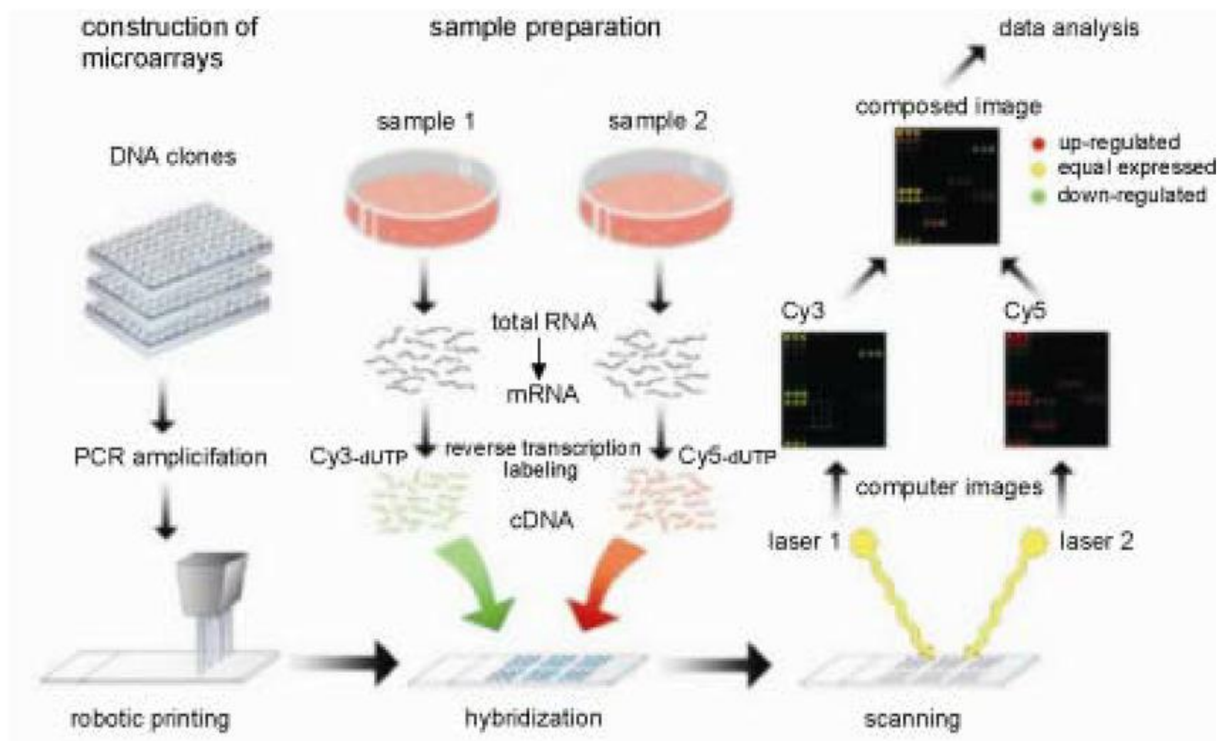


Figure 4: Schematic overview of cDNA microarray experiment

DNA microarrays - Techniques and applications in microbial systems
February 2004 Folia Microbiologica 49(6):635-64 DOI: 10.1007/BF02931546

PAM 50

PAM50 (Prediction Analysis of Microarray 50) is a 50-gene signature that classifies breast cancer in five molecular intrinsic subtypes: Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like (107). Each of the five molecular subtypes varies by their biological properties and prognosis (16). PAM50 gene signature is based on results of gene expression studies using microarray and hierarchical clustering in breast cancer tissue (16, 108). The intrinsic subtypes of BC, defined by PAM50, have been shown to be predictive of risk of recurrence and benefit of adjuvant therapy. Prosigna test is commercially available test that measures the expression of PAM50 gene set in tumor tissue, provide classification into intrinsic tumor subtypes, and calculate risk of recurrence score (ROR) for each patient (19, 20, 109).

1.7 Aims

The overall aim of the thesis was to compare gene expression in human normal breast tissue and breast cancer tissue to advance the understanding of the parity related effects on BC carcinogenesis.

Sub aims:

-Describe the cumulative incidence of breast cancer by parity in NOWAC, (Paper I).

- Establish representative sample of normal breast tissue biopsies among NOWAC women; (Paper II); describe gene expression in normal breast tissue associated with parity and some major risk factors for BC, (Paper III).

- Compare gene expression profiles in normal breast and breast cancer tissues (all taken within the NOWAC study) in relation to/depending on to parity, (Paper IV).

2 Material and methods

This thesis contains 3 different studies presenting a systems epidemiology approach to BC and parity. We use two-step integrated analyses in the same cohort; a novel design named an integrated systems epidemiology approach (36, 110).

In the first study we used epidemiological data from the NOWAC study. Data on different exposures was self-reported and questionnaire based. We conducted traditional epidemiological research, calculating cumulative incidence rates (CIR) for BC regarding parity (Paper I).

In the second study, we collected a large sample of normal breast tissue biopsies from healthy women through mammography screening program. Using microarray technology, we explored gene expression in normal breast tissue and described the association with relevant exposures (Paper II and III).

In the third study, using a nested case-control design and microarray technology we compared gene expression in normal breast tissue and breast cancer tissue, relating to parity (paper IV).

The analytic cohort was restricted to postmenopausal women.

2.1 Study population

The Norwegian Women and Cancer Study (NOWAC) (Paper I)

The NOWAC study (102, 111) was initially intended to investigate the relationship between BC and the use of oral contraceptives in the Norwegian population. It is a population-based, prospective cohort study, which began in 1991. Norway is an attractive site for such a study, as it has a complete population register, and a near-complete cancer registry. This allows for random sampling and follow-up of the whole population. Eligible women were those born between 1927-1965 with a Norwegian personal identification number. Selected women received a letter sent from the Central Bureau of Statistics, explaining the nature of the study, a questionnaire identified by a serial number, and a pre-paid envelope allowing them to return the completed questionnaire to the study center. Enrollment occurred in three waves and was nationwide. The final cohort consisted of approximately 172,000 women, 86% of which were born between 1943-1957 (Figure 5). Follow-up questionnaires were sent at 5–7-year intervals. Follow-up in 1998-2002 achieved a response rate of 81%. Information from the Cancer Registry of Norway regarding the incidence of BC and mortality was, in effect, complete.

Figure 5: Enrollment to Norwegian Woman and Cancer Study

The NOWAC questionnaire gathered demographic and lifestyle information on eligible women. This included information on oral contraceptive use, age at menarche, menopausal state, number of children, breastfeeding, smoking, medical history, physical activity, and diet. Follow up questionnaires contained different questions depending on research questions of interest at those times (lifestyle factors such as hormone replacement therapy, alcohol consumption, other medication use, sunbathing habits, more detailed dietary investigations). Length of follow-up questionnaires varied from 2-8 pages.

Normal breast tissue biopsies (Paper II and III)

During the period 2010-2012 we collected breast tissue biopsies, buffered blood samples and a one-page questionnaire from 368 women with a normal mammogram participating in the national screening program at the University hospital of North Norway in Tromsø. Women were asked if they had participated in the NOWAC study. The response rate was 64%.

A detailed description of the recruitment process and study population are presented in Paper II (112). Briefly, study participants were recruited through the national mammography screening program at the Breast Diagnostic Center at the University Hospital of North Norway (UNN), Tromsø, Norway, from October 2010 to May 2011.They were not referred due to pathological clinical findings or irregularities on previous mammograms but attended a scheduled routine mammogram. Eligible women were aged between 53-67 years, were post-menopausal, and were already participating in the nationally representative Norwegian Women and Cancer study (NOWAC) (102). Exclusion criteria for the present study included self-reported previous history of breast cancer, positive mammogram, other current malignant diseases, and use of anticoagulation therapy with warfarin, heparin, dipyridamole or clopidogrel. Eligible women who agreed to participate received written and oral information, signed an informed consent form, and answered a two-page questionnaire regarding menopausal status, weight and height, smoking and alcohol consumption, use of MHT and other medications. The number of included participants was 317. Three years after inclusion, data was linked to the Cancer Registry of Norway, using the unique personal identification number. This resulted in the exclusion of five participants who developed breast cancer within 3 years after the biopsy was taken, and one participant due to prior lymphoma diagnosis with unknown treatment. Thus, the final number of women included for statistical analysis was 311 (Figure 6).

Breast cancer biopsies (Paper IV)

From 2006 – 2010 ten Norwegian hospitals throughout the country participated in the sampling of tumor biopsies from participants in the NOWAC study (113, 114) Approximately 1 out of 3 women born 1943-1957 in Norway was at that time enrolled in the NOWAC study by answering one or more questionnaires. In collaboration with the Norwegian Breast Cancer Group, women who were admitted to one of the collaborating hospitals for diagnostic breast biopsy or BC surgery, was asked to donate a biopsy to be taken after the diagnostic one. Tissue samples were mailed overnight for biobanking at −70°C in Tromsø. A total of 372 biopsies were collected (Figure 7). Additionally, a blood sample and a two-page questionnaire were collected. Sampling and research on human biological material has been approved by Regional Committees for Medical and Health Research Ethics in Norway and is in accordance with the Norwegian law on biobanking

Study population Paper IV: In this nested case-control study we used 311 normal breast tissue biopsies (controls) matched by age with cancer biopsies(cases). The cases and controls within each pair were kept together through all laboratory procedures minimizing batch effect and technical noise. The dataset comprised 622 samples combined in 311 pairs, with 47323 probes.

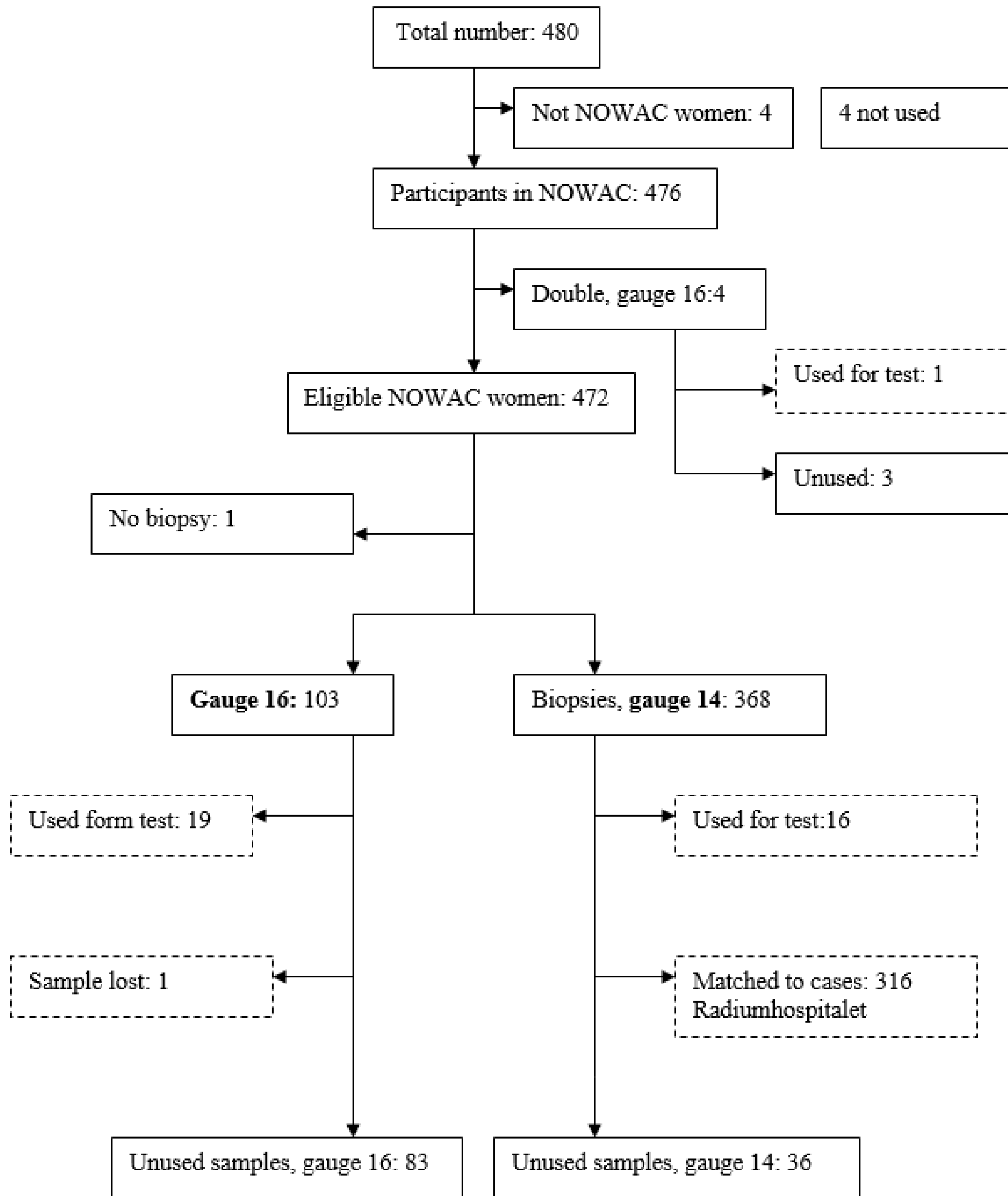


Figure 6. Flowchart depicts normal breast biopsy collected.

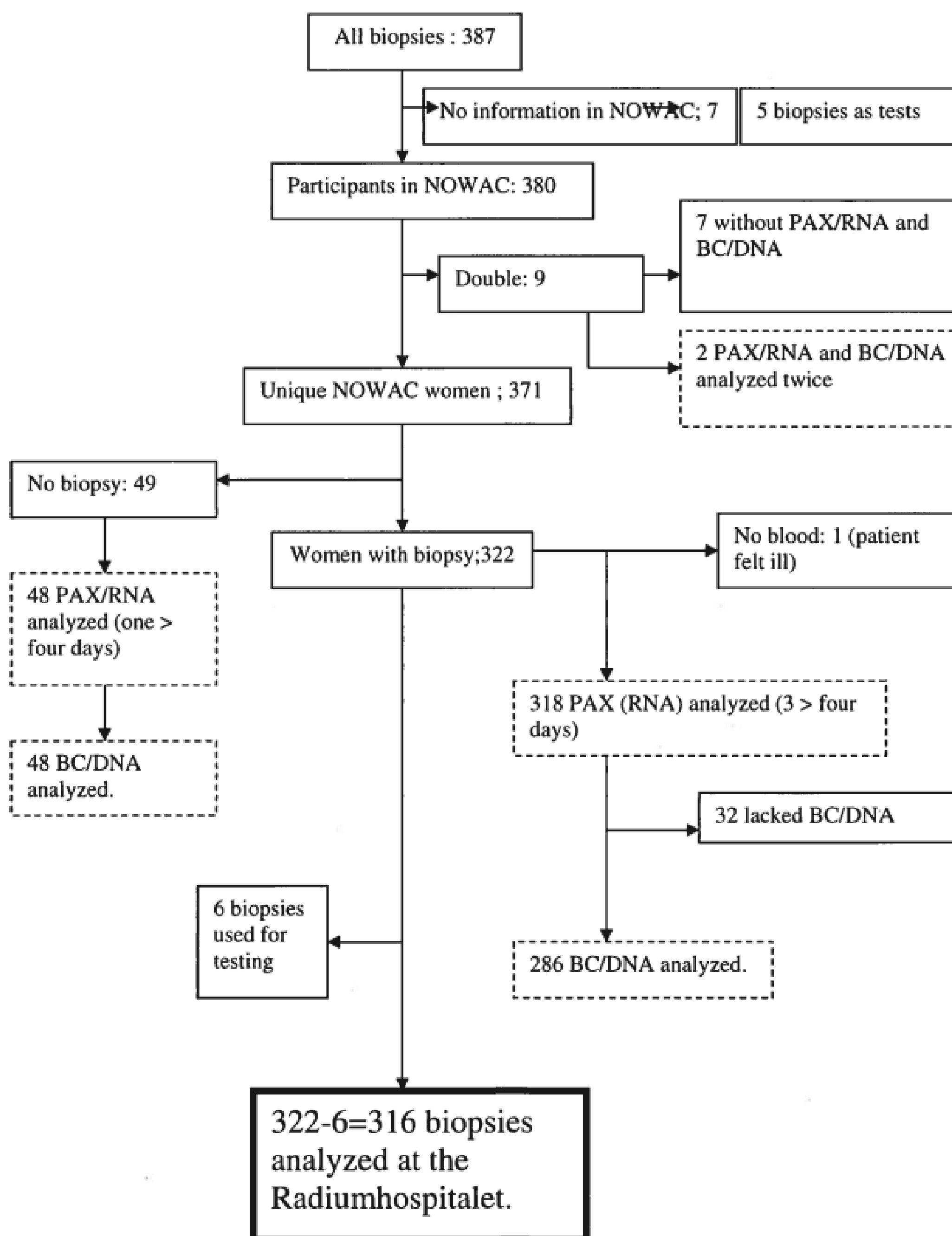


Figure 7. Flowchart depicts breast cancer biopsy collected.

Follow-up with Cancer Registry of Norway and Central Population Registry

Death and emigration status were updated through linkage to the Cause of Death Registry at Statistics Norway and Central Population Registry, respectively. The exit date from NOWAC was determined either by the date of cancer diagnosis, death, emigration, or end of follow-up on 31.12.2018.

The Cancer Registry of Norway was established in 1951 and is one of the oldest national cancer registries in the world. It is mandatory for all medical practitioners in Norway to report new cancer cases and update case information with clinical information (115). Nearly 100% of BC cases diagnosed in Norway are registered (116).

2.2 Ethical approval

The NOWAC study was approved by the Region Ethic Committee and by the Norwegian Data Inspectorate. Each participants gave informed written consent for enrolment and follow-up by linkage to the Norwegian Cancer Registry and the Norwegian Cause of Death Registry. They were informed that they, at any time, could withdraw from the study and have their data deleted.

Gene expression studies included in this thesis were performed in compliance with the Helsinki Declaration. At the inclusion in the studies, each participant gave informed written consent. Statistics of Norway provided information on cancer diagnosis via linkage to the Cancer Registry of Norway. Participants were informed about this linkage and information to the researchers were given using serial numbers instead of the original identity number. Study participants who donated tissue biopsies were informed that samples would be used for different gene expression analyses. All women included, were also informed that we are not providing any individual information regarding results of analyses.

The North Norway Regional Committee for Medical and Health Research Ethics (REK-Nord case number # 200603551) and the Norwegian Data Inspectorate approved the gene expressions studies in accordance with the Norwegian Biobank Act.

2.3 Variables

The main variable or exposure in this thesis is parity. However, in article I and III we also used other exposures relevant to BC (BMI, alcohol consumption, smoking, MHT, age at first birth, and breastfeeding).

Parity, age at first birth, and breastfeeding

Information on respective variables was collected from the NOWAC questionnaires.

Paper I: Women with parity > 6 were excluded due to a low number of participants. Parity was analyzed both as discrete numbers in the range from zero to six, and as four parity groups 0, 1-2, 3-4 and 5-6 children, respectively. In stratified analyses we used four parity groups except stratified for breastfeeding, where we had two parity groups, 1-3 and 4-6 children. Breastfeeding was self-reported in months of duration per child. In the analyses, women who breastfed each child for ≥6 months were compared with those who breastfed less. Maternal age at first birth was calculated by the year of birth of the first child.

Paper II and III: The parity variable was dichotomized into parous versus non-parous for analyses of gene expression.

Paper IV: Parity was analyzed as discrete numbers in the range from zero to eight for descriptive analyses. Parity data was categorized into three groups: 0, 1-3 and 4-8 when modelling gene expression as linear.

Body mass index (BMI):

BMI was calculated as weight divided by height squared (kg/m^2). Information about height and weight was extracted from questionnaires. For paper I we used information from the questionnaire filled in at baseline (first time the participants filled the questionnaire containing these questions). For paper II and III information was collected from the questionnaire filled in at the same time as the biopsy was taken.

Paper I: The variable BMI was dichotomized into BMI<25 (normal weight) and BMI≥25 (pre-obesity or overweight), according to BMI classification of the World Health Organization (WHO). We used the cutoff between normal weight and overweight/obesity, as many health effects occur in this transition.

Paper II and III: BMI was calculated, and obesity was defined according to the definition of the WHO (BMI > 30). When studying gene expression, small effect sizes are often expected. To increase our chances of identifying any effects of BMI, the cutoff was pushed to 30.

BMI variable was not included in Paper IV.

Alcohol consumption

Paper I: Alcohol consumption is based on information from the questionnaire concerning alcohol consumption the last two weeks; none versus 1+. Alcohol consumption was defined as ever-drinkers and teetotalers. We know that alcohol consumption tends to be underreported in NOWAC. Hence, a cutoff between consumers and non-consumers was employed.

Paper II and III: Women who had consumed alcohol during the week prior to the biopsy, regardless of the type or amount, were defined as alcohol consumers. Alcohol may influence gene expression; however, the dose-response and time dynamics of this influence has not been documented in the literature. Hence, a cutoff between consumers and non-consumers was employed here as well.

Smoking

The smoking status in NOWAC was coded into current, former, or never smokers in the first questionnaire.

Paper I: For this investigation current and former smokers were combined into one category, ever smokers. Changes in smoking status during follow-up was not used in the analyses.

Paper II and III: Women who had smoked during the week prior to biopsy were defined as smokers. For sensitivity analysis we combined two sources of exposure data: the detailed, 8-page questionnaire answered 0-20 years prior to the biopsy, and the 2-page questionnaire answered at the time of the biopsy.

Menopausal hormone therapy (MHT):

Paper II and III: Only women who were current users of systemic MHT (tablets or patch) were defined as HT users.

For analytical purposes, stratification variables were dichotomized. More refined cut of points generally gives finer stratifications, identify subgroups, and provide more detailed insight into data. On the other hand, more subgroups increase the complexity of the analyses and make

the interpretation of the results more challenging. Even if our study has a large sample size, we chose dichotomized variables partly due to the explorative nature of study analyses.

2.4 Statistical Methods

Statistical analyses for gene expression were not the main focus of this thesis. Detailed information on statistical methods used can be found in respective papers. Here is a brief description:

Paper I: The Statistical software used in in this investigation was R statistical environment and especially the Epi package for Epidemiological research. We calculated cumulative incidence rates (CIR) as an estimate of the risk of having a BC diagnosis within a given age-interval, here 35-84 years. The cumulative incidence rates were calculated for each age-group as the number of incident cases of BC divided by the number of person-years (per 100.000) until death, emigration, or a BC diagnosis. We used Poisson regression analyses for this calculation. We also compared cumulative incidence rates between two strata for respective variables (age at first birth, breastfeeding, BMI, alcohol use, and smoking). The log-rank test was applied as a test of significance (117).

Paper II and III: Gene expression profiles from normal breast tissue biopsies were analyzed using Illumina bead chip arrays. PCR (principal component analysis) including all genes and K-means clustering including genes with the most variability (inter quantile range (IQR) > 1 log2 unit) was used for initial analysis of the dataset. Gene clusters were analyzed for overrepresented gene ontology (GO) terms using the clusterProfiler package (118). The association of exposures and covariates with gene expression was determined using LIMMA (linear models for microarrays). Camera (119) was used to identify pathways and gene ontology terms that were related to the exposure variables.

Data analysis was done using R (r-project.org). Raw files were quantile normalized using the Bioconductor lumi package (120). P-values from the linear models were corrected for multiple testing using the method of Benjamin and Hochberg (121).

STATA (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.) was used for descriptive statistics. T-tests were used for BMI and age as continuous variables, and chi-square tests were used for the categorical variables.

Paper IV: Statistical analyses was performed by Norwegian Computing Center. Several steps were applied during preprocessing of initial dataset. PCA was used for clustering analysis. For gene annotation lumiHumanIDMapping (122) was utilized. PAM50 classifier was applied to all samples. For identification of differentially expressed genes, R with Bioconductor packages was used.

3 Results-summary of papers

3.1 Paper I

Parity and cumulative incidence rates of breast cancer in the Norwegian Woman and Cancer Study (NOWAC)

Increasing parity is a known risk reducing factor for the development of breast cancer, with a risk reduction of 7-8% per child. This study examined a prospective cohort of 165,238 women followed over 18 years to determine cumulative incidence rates and describe the lifelong risk of breast cancer in relation to party. We used Poisson regression analysis to calculate the cumulative incidence rate of breast cancer for parity, stratifying by other known risk factors such as maternal age at first birth, breastfeeding, body mass index, smoking, and alcohol consumption. 8210 of the cohort developed breast cancer after 17.3 years average follow-up. Incidence increased among women aged 60-64 and again at ages 80-84, with a reduction in incidence in the 75-79 age group. Cumulative incidence by age 84 was 11.7%, very close to the national cumulative incidence of 11.3%. Stratification by parity showed a cumulative incidence of 12.6% for parity 0, 13.3% for parity 1, 11.8% for parity 2, 10.5% for parity 3, 9.6 % for parity 4 and 8.7% for parity 5-6. Within the same parity group, women exposed to others BC risk factors had higher CIR. The parity-specific cumulative incidence rate of breast cancer showed the same pattern of decrease when stratified for other breast cancer risk factors. This confirmed a consistent decrease in cumulative incidence rate for each additional child across strata in established risk factors for breast cancer. To our knowledge this is the first study using CIR as risk estimate for BC in relation to parity (per child and per parity groups) and stratified for other risk factors.

3.2 Paper II

Practical and Ethical Issues in Establishing a Collection of Normal Breast Tissue Biopsies— Part of the NOWAC Post-Genome Cohort.

For tissue-based studies of breast cancer, getting access to truly normal, well-annotated tissue can be a challenge. To address that need, we collected 368 breast tissue biopsies and buffered blood samples from healthy postmenopausal women. Volunteers were part of the Norwegian Women and Cancer (NOWAC) Postgenome cohort, recruited through the national mammography screening program. We discussed logistics and feasibility of biopsy taking from a healthy population and ethical aspects of genomic research on healthy people. The NOWAC normal breast tissue collection for gene expression analysis will provide a valid basis for comparison in case-control studies.

3.3 Paper III

Associations of breast cancer related exposures and gene expression profiles in normal breast tissue – The Norwegian Women and Cancer (NOWAC) normal breast tissue study.

Breast tissue is utilized in tissue-based studies of breast carcinogenesis. While gene expression in breast tumor tissue is well explored, our knowledge of transcriptomic signatures in normal breast tissue is still incomplete. The aim of this study was to investigate variability of gene expression in a large sample of normal breast tissue biopsies, according to breast cancer related exposures (obesity, smoking, alcohol, hormone therapy, and parity). Methods: We analyzed gene expression profiles from 311 normal breast tissue biopsies from cancer-free, post-menopausal women, using Illumina bead chip arrays. Principal component analysis and K-means clustering was used for initial analysis of the dataset. The association of exposures and covariates with gene expression was determined using linear models for microarrays. Results: Heterogeneity of the breast tissue and cell composition had the strongest influence on gene expression profiles. After adjusting for cell composition, obesity, smoking, and alcohol showed the highest numbers of associated genes and pathways, whereas hormone therapy and parity were associated with negligible gene expression differences. Conclusion: Our results provide insight into associations between major exposures and gene expression profiles and provide an informative baseline for improved understanding of exposure-related molecular events in normal breast tissue of cancer-free, post-menopausal women

To our knowledge this is the first study describing the association of gene expression and BC related exposures in large sample of normal breast tissue biopsies of postmenopausal women

3.4 Paper IV

No impact of parity on global gene expression levels in breast cancer tissue and normal breast tissue – a nested case-control study in the Norwegian Women and Cancer Study

Increasing parity and age at first birth are associated with reduced risk of breast cancer. A prevailing hypothesis postulates that this risk reduction is due to persistent modifications of gene expression in the breast tissue, primarily during first pregnancy. We drew on the NOWAC cohort, asking that women enrolled between 2006-2010 who received breast cancer diagnoses supply a second biopsy for research purposes. Between October 2010 and May 2011normal breast tissue was also collected from NOWAC participants taking part in the national mammographic screening program. This provided 279 age-matched case control pairs. The matched-pair design was adhered to throughout all laboratory analyses, and samples were analyzed to determine differences in gene expression. Nearly all expressed genes were significant differential expression between cases and controls. However, when parity (stratified as 0, 1-3, 4-8 children) was incorporated into the model, only three genes displayed significant expression changes. Our findings do not support the hypothesis that reduced risk due to parity is mediated by permanent changes in gene expression in breast tissue following pregnancies.

4 Discussion

4.1 General discussion

Study design and population

The interpretation of results and generalizability of a study's findings depend on the quality of the data used in analyses and the choice of an appropriate study design.

In all studies, the data collected is subject, to some degree, to random or/and systematic errors (123, 124). Random errors are caused by accident, random circumstances, and are generally controlled by large sample size. Systematic errors, on the other side, introduce different types of bias that affect the results and conclusion of the study. Selection bias is systematic error that may occur due to systematic differences in the way study participants are included into

the study or during follow-up. Selection bias may also be related to the participants of the study, due to self-selection (responder bias). Information bias is another systematic error in the way information is collected from participants. Information bias may lead to misclassification into the exposure or outcome groups due to inadequate collection of information. Recall bias is a type of information bias where participants do not remember accurately past events. Internal and external validity of the studies are defined by the methods they used to handle these potential errors. Confounding is also a type of bias and may result in a misinterpretation of the association between an exposure and an outcome. Confounding occurs when the association between exposure and outcome is caused by a third factor, which is associated with both. We can adjust for the confounding effect of known confounders in the statistical analyses. However, residual confounding, unmeasured or unknown confounders, may still have a large impact on the results (123, 124).

In the NOWAC study, with 172,000 participants, the probability of random error is low. However, while a large sample size generally allows more robust statistical analyses and reliable generalization, random error can persist. Nondifferential misclassification is a potential source of random error that could lead to the obscureness of moderate associations because comparison groups are equally affected.

Participants in NOWAC are randomly selected nationwide, with relatively high response rates (102), so selection bias should not be a significant issue in our epidemiological study (Paper I). The population in the NOWAC study was found to be representative of the Norwegian female population (111), confirming the external validity of the study. In Paper IV, we analyzed tissue samples from two groups: cases and controls. Participants in both groups were already part of the NOWAC population, and biopsy collection was carried out in the same way. However, inclusion in the present study was slightly different, introducing the possibility of selection bias. Controls were included from October 2010 to May 2011 from the NOWAC population living in Northern Norway; they were healthy women participating in the mammography screening program. Cases were included between 2006 and 2010 across eleven different Norwegian hospitals and were women with breast cancer attending diagnostic breast biopsies at the breast diagnostic center. However, it is unlikely that some differences in sampling or potential temporal changes in the NOWAC population during sampling introduced selection bias that strongly affected our findings.

As part of the first paper, we compared the incidence of BC in our study cohort with incidence rates from the Cancer Registry of Norway and found that they were comparable. This provides further support for the generalizability of our findings.

The NOWAC study uses a prospective design by including healthy women and collecting information on exposures before the studied outcomes occur. The prospective design potentially avoids recall bias, ensures that the time-effect relationship is known, and makes it possible to study how various exposures affect the risk of developing a particular outcome. With this design, one can explore associations, although one cannot conclude with certainty regarding causality. The longitudinal follow-up in the study allows for an estimation of risk and interpretation of the primary exposures or variables as risk factors. We utilized the advantages of the NOWAC study design in Paper 1.

The traditional problems with case-control studies include among others, selection bias. To avoid this problem, we used a nested case-control study design in paper IV, where cases and controls are included from the same larger cohort. The recall bias, another potential problem in case-control studies, was handled with a prospective design.

There are several critical confounding factors to consider when studying parity and risk of BC. Previous studies have found that breastfeeding is associated with a reduced risk of BC, and the number of children and length of breastfeeding are closely linked. Similarly, alcohol and smoking, both associated with a higher risk of BC, are known to be associated with each other. We addressed this problem in Paper I by using stratified analyses.

Further, our gene expression studies used a rigorous methodology to ensure that biological samples from the cases and control pairs were kept together during laboratory work and statistical analyses. This minimized batch effect and technical noise in the data.

The normal breast tissue biopsies in our studies were also collected from women in the NOWAC study. The response rate for inclusion was 64%, which may be considered high for an invasive procedure with no diagnostic or therapeutic consequences for the study participants. Including cases and controls from the same large cohort ensured that information on the relevant variables was collected standardized from both BC cases and controls.

The women in the control group were recruited at mammography screening to make sure that they had neither malignant nor benign breast disease and to ensure successful age matching with our BC cases.

Including healthy women from one screening center may have decreased the representativeness of the control cohort. However, this reduced the possibility of pre-analytic variance, which is particularly important for genomic analyses.

Although the genomic changes within individual cancer cells are significant, there is an increasing awareness of the essential role of the tumor microenvironment. We, therefore, chose to analyze whole biopsies, as opposed to single-cell analyses as done in some other studies.

Population-based prospective cohorts with established biobanks collect a variety of biospecimens from cohort populations for different studies (DNA, genetics, and molecular biomarkers). The most collected biological materials are blood, saliva, buccal cells, urine, and umbilical cord. However, tissue samples from healthy individuals in population-based cohorts are rare (125). To our knowledge, the biopsies of normal breast tissue collected from postmenopausal women in the NOWAC study are a unique and valuable resource.

Regarding study design, in this thesis, we applied a novel design named an integrated systems epidemiology approach (110). We used two-step integrated analyses in the same cohort. In the first step, calculating CIR, we explored the numerical relation between parity as exposure and BC as outcome. For these explorative, hypothesis-generating analyses, we used the whole NOWAC cohort to ensure an assessment of the effect that was as correct as possible. The generated hypothesis was tested in two gene-expression studies in the nested case-control design as a second step of the analyses.

In Paper I, we used cumulative incidence rates to estimate the risk of developing BC. In oncology, cumulative risk measures the likelihood that a cancer-free individual will develop cancer by a specific age. This is reasonably approximated by the cumulative rate, which is the sum of the age-specific incidence rates of the disease in question from birth up to the specified age limit – in this case, age 80. Since the age-specific rates are computed in five-year groups, the cumulative rate will be five times the sum of the rates of disease in the

groups, assuming similar rates within the ages of these 5-year spans. The use of cumulative rate gives a more standard estimate of risk, eliminating the need for a reference population, and gives a more intuitive and interpretable measure of lifetime risk.

Validity of the variables

Information on parity and other covariates was collected from the NOWAC baseline questionnaire. Information was self-reported and may be subject to information bias. This methodological problem was addressed in several validation studies.

Parity, age at first birth, and breastfeeding: The number of children in our study was limited due to low fertility rates in Norway, resulting in a relatively small number of women who had given birth to more than four children (Paper IV). Previous analyses have confirmed that the NOWAC participants exhibit the same fertility level as the general population (126). In the questionnaire, women were asked about their age at each birth. Breastfeeding was self-reported as the duration of breastfeeding in months for each child. The variable is not validated.

Body-mass index (BMI): Skeie *et al*. performed a validation study (127), which showed a systematic error in self-reported weight towards a lower weight, hence a lower BMI. Misclassification of women into a lower BMI group is a type of bias that may lead to an underestimation of the effect of BMI on BC risk. However, a validation study of the NOWAC cohort showed that self-reported weight and height provided a valid ranking of BMI for middle-aged women (127).

Smoking and alcohol use: Information on smoking and alcohol were self-reported. Based on smoking trends in the general population, it is more likely that study participants quit smoking than started during the follow-up time. This may have underestimated the effect of this exposure. Changes in smoking status, duration of smoking, and number of cigarettes were not considered in the papers included in this thesis. The smoking variable has not been validated in NOWAC. However, smoking status at baseline and follow-up questionnaires were compared in 2018 (128). Only 1.8% of participants reported that they were ever smokers at baseline and never smokers at follow-up. A low degree of non-differential misclassification is therefore expected, and any bias is likely to be towards finding no association between BC and smoking. The alcohol variable was not validated in the NOWAC study.

Technical consideration in gene expression analysis

Our gene expression studies are exploratory and potentially hypothesis-generating, so we did not carry out validation of the methods used or the results.

There are many challenges regarding the interpretation and validation of all gene expression studies. Technical aspects that might introduce error include, but are not limited to, the pre-analytic steps of microarray processing and analysis: mRNA isolation, cDNA synthesis, labeling, hybridization, washing, and scanning. These can produce both random and systematic errors depending on the way in which the processes are carried out (129). In addition, using large-scale microarray technology in epidemiological studies has its own difficulties. Intra- and inter-sample variability, analytical issues, and a lack of standardization of methods may all influence the results. To reduce technical noise and systematic error, all cases and age-matched control pairs were processed pairwise through the entire analytic process.

4.2 Discussion of the main results

Our epidemiological study (Paper I) showed a clear association between parity and BC, where parity provided additional protection for each additional child. These results align with many other epidemiological studies, which find that full-term pregnancy, particularly at early maternal age, protects against BC and that each additional pregnancy gives an added risk reduction (ref). In the same study, we calculated the risk of BC in relation to parity stratified for other known risk factors: BMI (<25/≥25), smoking (yes/no), and alcohol(yes/no). Our study results showed an increased BC risk in participants exposed to the risk factors in all parity groups. While increasing risk for BC in exposed groups, selected risk factors did not affect the protective effect of parity. BC was consistently higher in the nulliparous women compared with parous groups. Increased BC risk in postmenopausal women with high BMI (130), who consume alcohol (68), or smoke before their FFTP (93) is well known from previous epidemiological studies. In our study we additionally showed parity as an independent risk factor for BC.

Our first gene expression study (Paper III) showed that the known risk factors for BC BMI, smoking, and alcohol had the strongest associations with gene expression profiles after correcting for heterogeneity of the tissue sample materials. They were associated with a

higher number of differentially expressed genes and pathways then other exposures. Apart from BMI (131), we have not found other studies that examine this association. Finding significant effects of other known risk factors on the breast tissue strengthens our finding that the lack of association between parity and gene expression profile in our cohort is valid and not a chance occurrence.

The carcinogenesis of BC is promoted by complex biological mechanisms of tumor cells in their microenvironment, affected by hormonal, chemical, and immunological factors (132). Both, our epidemiological data and gene expression studies support the existing theories that chemical factors like smoking and alcohol, increase BC risk.

The biological mechanism underlying the protective effect of FFTP against BC is unknown. The current paradigm is that a woman's first pregnancy induces permanent genetic changes in a woman's breast tissue. This is largely based on the study by Russo *et al*., where they found a large amount of significantly differentially expressed genes between nulliparous and parous BC samples and controls (9). Two other studies conducted after Russo found 208 differentially expressed genes between parous and nulliparous healthy postmenopausal women (133, 134). A study by Rotunno et al. explored gene expression signatures in tumours and non-tumours breast tissue (135). Comparing parous with nulliparous women, they found hundreds of significant differentially expressed genes. Surprisingly, we found no association between parity and gene expression profiles in either of our two studies, even though we used two different statistical methodologies (papers III and IV).

There are several possible causes of this discrepancy. Russo and colleagues examined a total of 74 samples, where the controls were women who had been biopsied on clinical indication. Histological findings included benign breast disease but without hyperplasia or atypia. In contrast, our study included 311 age-matched paired samples of BC tissue and normal breast tissue found at routine screening. Degnim *et al.* found that cancer-free breast tissue containing benign disease is significantly different from normal breast tissue (136), underscoring the importance of appropriate control samples.

Russo's study used laser capture microdissection from paraffin-embedded tissue samples. They focused on a crude comparison between nulliparous and parous women and did not address the effect of multiple children on gene expression. Our study used whole biopsies initially preserved in microarray later and studied the effect of each additional child on BC risk.

Rotunno developed a parity signature by analysing microRNA microarray data from 130 parous and nulliparous patients undergoing reduction mammoplasty. In our study, we used mRNA microarray in normal breast tissue. While both microRNA and mRNA microarrays use a similar technology, they focus on different parts of the transcriptome. mRNA microarrays measure alteration in the transcriptional activity of protein-coding genes, while microRNA microarrays measure expression levels of microRNA. microRNA plays an essential role in the regulation of gene expression post-transcriptionally (137). Using different sources of breast tissue and measuring different RNA molecules of interest makes comparing the studies difficult.

Many other studies have also studied the relationship between parity and BC. Unfortunately, significant methodological discrepancies, especially concerning the choice of control samples, make comparisons challenging.

Our studies were designed to elucidate the effect of parity and each consecutive pregnancy on the risk of BC, and further examine the effect of each pregnancy on tissue gene expression in normal and BC tissue. We found no significant findings associating parity and risk of BC in our gene expression studies, suggesting that other theories explaining the mechanism of the protective effect of parity should be pursued in future studies. As reviewed in the introduction, section 1.4, current theories include alterations in the pregnant woman's systemic hormonal environment, mammary gland differentiation, the role of the breast microenvironment and mammary gland stem cells.

Unfortunately, the studies supporting most of these theories are based on animal models which we know have limited transference value to human beings. Ideally, a prospective cohort study would involve collecting blood samples for hormonal analyses, as well as breast tissue samples from healthy women before, during and after each subsequent pregnancy, as well as collecting data on exposures to known risk factors for BC. Perhaps data from current pregnancy studies. Future studies should consider also including breast tissue samples of healthy women before, during and after pregnancy to further examine the mechanism of parity on BC risk. However, the cost in time and resources for such a study is considerable, from both the perspective of society and the individual woman.

In addition, BC is a heterogenous disease, and we know that various exposures, for instance breast feeding, affect the risk of different subtypes of BC. Obtaining a large enough cohort with various subtypes of BC is difficult outside a case-control design. This is further complicated by the knowledge that various risk factors affect the risk of BC differently at different periods of a woman's life, and large groups of women are necessary in order to assess the influence of various exposures.

5 Main conclusion

In conclusion, our research confirms that parity is a major protective factor of BC. Each childbirth gave equal risk reduction independently of other major risk factors. The results from our gene expression studies in both normal breast tissue and BC tissue did not show any differentially expressed genes regarding parity in either cases or controls. We find no evidence of systematic change of gene expression in the breast tissue of healthy women or cancer patients

6 Future perspectives

Genetic research is increasingly dependent on a multidisciplinary team. This type of study is both cost and labor intensive and should ideally be conducted in large international studies with highly specialized competency within all fields ranging from the technical to the bioinformatics and clinical. The systems epidemiology approach used in our studies is a good example of a framework for such work. However, future studies should use more modern technical methods, for instance whole genome sequencing. Ideally, future studies examining the relationship between parity and BC should be done in countries where more women have many children. Unfortunately, from this perspective, the rate of childbirth decreases as nations improve economically to the point where they can afford to finance large population-based prospective cohorts as are necessary for such research.

There is an increasing realization that we need international prospective large-scale cohorts that collect and store biological samples and epidemiological data (138). Such studies may contribute to uncovering what is the mechanism underlying the protective effect of parity on BC risk and would require both liquid and tissue samples from healthy women before, during and after subsequent pregnancies, as well as samples from women with BC from both healthy breast and tumor tissue with detailed histological, hormonal, and molecular characteristics.

The ethical dilemmas for each participant in such studies warrants further consideration, as detailed private and genetic data would be available to many researchers across the globe.

Also, future studies should use more modern technical methods, for instance whole genome sequencing. Ideally, future studies examining the relationship between parity and BC should be done in countries where more women have many children. Unfortunately, from this perspective, the rate of childbirth decreases as nations improve economically to the point where they can afford to finance large population-based prospective cohorts as are necessary for such research.

Perhaps the most important contribution of this current work is to demonstrate that healthy women are willing and able to participate in clinical studies involving invasive procedures, and future studies with well-formulated research questions and appropriate study designs can build on this.

Works cited

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.
2. MacMahon B, Cole P, Lin T, Lowe C, Mirra A, Ravnihar B, et al. Age at first birth and breast cancer risk. Bull World Health Organ. 1970;43:209 - 21.
3. Albrektsen G, Heuch I, Tretli S, Kvåle 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(6):604-11.
4. Hinkula M, Pukkala E, Kyyrönen P, Kauppila A. Grand multiparity and the risk of breast cancer: population-based study in Finland. Cancer Causes Control. 2001;12(6):491-500.
5. Ma H, Henderson KD, Sullivan-Halley J, Duan L, Marshall SF, Ursin G, et al. Pregnancy-related factors and the risk of breast carcinoma in situ and invasive breast cancer among postmenopausal women in the California Teachers Study cohort. Breast Cancer Res. 2010;12(3):R35.
6. Ellis H, Mahadevan V. Anatomy and physiology of the breast. Surgery (Oxford). 2013;31(1):11-4.
7. Russo J, Russo IH. Development of the human breast. Maturitas. 2004;49(1):2-15.
8. Alex A, Bhandary E, McGuire KP. Anatomy and Physiology of the Breast during Pregnancy and Lactation. Adv Exp Med Biol. 2020;1252:3-7.
9. Russo J, Balogh G, Russo I. Full-term pregnancy induces a specific genomic signature in the human breast. Cancer Epidemiol Biomarkers Prev. 2008;17:51 - 66.
10. Russo J, Moral R, Balogh G, Mailo D, Russo I. The protective role of pregnancy in breast cancer. Breast Cancer Res. 2005;7:131 - 42.
11. Peri S, de Cicco RL, Santucci-Pereira J, Slifker M, Ross EA, Russo IH, et al. Defining the genomic signature of the parous breast. BMC Med Genomics. 2012;5:46.
12. Watson CJ. Key stages in mammary gland development - Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. Breast Cancer Research. 2006;8(2):203.
13. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Nature Reviews Disease Primers. 2019;5(1):66.
14. Malhotra GK, Zhao X, Band H, Band V. Histological, molecular and functional subtypes of breast cancers. Cancer Biology & Therapy. 2010;10(10):955-60.
15. Tan PH, Ellis I, Allison K, Brogi E, Fox SB, Lakhani S, et al. The 2019 World Health Organization classification of tumours of the breast. Histopathology. 2020;77(2):181-5.
16. Sørlie 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 USA. 2001;98.
17. Prat A, Parker J, Karginova O, Fan C, Livasy C, Herschkowitz J, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12:R68.
18. Parker J, Mullins M, Cheang M, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. J Clin Oncol. 2009;27:1160 - 7.
19. Wallden B, Storhoff J, Nielsen T, Dowidar N, Schaper C, Ferree S, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. BMC Med Genomics. 2015;8:54.
20. Hequet D, Harrissart G, Krief D, Maumy L, Lerebours F, Menet E, et al. Prosigna test in breast cancer: real-life experience. Breast Cancer Res Treat. 2021;188(1):141-7.
21. NBCG. The Norwegian Breast Cancer Group 2023 [Available from: http://nbcg.no/.

22. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. Cancers (Basel). 2021;13(17):4287.
23. Franco G. Bernardino Ramazzini and women workers' health in the second half of the XVIIth century. J Public Health (Oxf). 2012;34(2):305-8.
24. MUSTACCHI P. Ramazzini and Rigoni-Stern on Parity and Breast Cancer: Clinical Impression and Statistical Corroboration. Archives of Internal Medicine. 1961;108(4):639-42.
25. Husby A, Wohlfahrt J, Øyen N, Melbye M. Pregnancy duration and breast cancer risk. Nature Communications. 2018;9(1):4255.
26. Rosner B, Colditz GA, Willett WC. Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. Am J Epidemiol. 1994;139(8):819-35.
27. Albrektsen G, Heuch I, Hansen S, Kvåle G. Breast cancer risk by age at birth, time since birth and time intervals between births: exploring interaction effects. Br J Cancer. 2005;92(1):167-75.
28. Schedin P. Pregnancy-associated breast cancer and metastasis. Nature Reviews Cancer. 2006;6(4):281-91.
29. Slepicka PF, Cyrill SL, dos Santos CO. Pregnancy and Breast Cancer: Pathways to Understand Risk and Prevention. Trends in Molecular Medicine. 2019;25(10):866-81.
30. Katz TA. Potential Mechanisms underlying the Protective Effect of Pregnancy against Breast Cancer: A Focus on the IGF Pathway. Frontiers in Oncology. 2016;6(228).
31. Sivaraman L, Medina D. Hormone-Induced Protection Against Breast Cancer. J Mammary Gland Biol Neoplasia. 2002;7(1):77-92.
32. Dall G, Risbridger G, Britt K. Mammary stem cells and parity-induced breast cancer protection- new insights. The Journal of Steroid Biochemistry and Molecular Biology. 2017;170:54-60.
33. Ginger M, Gonzalez-Rimbau M, Gay J, Rosen J. Persistent changes in gene expression induced by estrogen and progesterone in the rat mammary gland. Mol Endocrinol. 2001;15:1993 - 2009.
34. Guzman RC, Yang J, Rajkumar L, Thordarson G, Chen X, Nandi S. Hormonal prevention of breast cancer: mimicking the protective effect of pregnancy. Proc Natl Acad Sci U S A. 1999;96(5):2520-5.
35. Rajkumar L, Guzman RC, Yang J, Thordarson G, Talamantes F, Nandi S. Prevention of mammary carcinogenesis by short-term estrogen and progestin treatments. Breast Cancer Research. 2003;6(1):R31.
36. Lund E, Nakamura A, Snapkov I, Thalabard JC, Olsen KS, Holden L, et al. Each pregnancy linearly changes immune gene expression in the blood of healthy women compared with breast cancer patients. Clinical epidemiology. 2018;10:931-40.
37. Lipworth L, Bailey LR, Trichopoulos D. History of breast-feeding in relation to breast cancer risk: a review of the epidemiologic literature. J Natl Cancer Inst. 2000;92(4):302-12.
38. 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(2):81-7.
39. Yang L, Jacobsen KH. A systematic review of the association between breastfeeding and breast cancer. J Womens Health (Larchmt). 2008;17(10):1635-45.
40. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet. 2002;360(9328):187-95.
41. Islami F, Liu Y, Jemal A, Zhou J, Weiderpass E, Colditz G, et al. Breastfeeding and breast cancer risk by receptor status--a systematic review and meta-analysis. Ann Oncol. 2015;26(12):2398-407.
42. Fortner RT, Sisti J, Chai B, Collins LC, Rosner B, Hankinson SE, et al. Parity, breastfeeding, and breast cancer risk by hormone receptor status and molecular phenotype: results from the Nurses' Health Studies. Breast Cancer Res. 2019;21(1):40.
43. Ambrosone CB, Higgins MJ. Relationships between Breast Feeding and Breast Cancer Subtypes: Lessons Learned from Studies in Humans and in Mice. Cancer Res. 2020;80(22):4871-7.

44. Collaboration N. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. Lancet. 2016;387(10026):1377-96.
45. Dixon JB. The effect of obesity on health outcomes. Mol Cell Endocrinol. 2010;316(2):104-8.
46. Midthjell K, Lee CM, Langhammer A, Krokstad S, Holmen TL, Hveem K, et al. Trends in overweight and obesity over 22 years in a large adult population: the HUNT Study, Norway. Clin Obes. 2013;3(1-2):12-20.
47. Jacobsen BK, Aars NA. Changes in body mass index and the prevalence of obesity during 1994-2008: repeated cross-sectional surveys and longitudinal analyses. The Tromsø Study. BMJ Open. 2015;5(6):e007859.
48. Macinnis RJ, English DR, Gertig DM, Hopper JL, Giles GG. Body size and composition and risk of postmenopausal breast cancer. Cancer Epidemiol Biomarkers Prev. 2004;13(12):2117-25.
49. Suzuki S, Kojima M, Tokudome S, Mori M, Sakauchi F, Wakai K, et al. Obesity/weight gain and breast cancer risk: findings from the Japan collaborative cohort study for the evaluation of cancer risk. J Epidemiol. 2013;23(2):139-45.
50. van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol. 2000;152(6):514-27.
51. Agurs-Collins T, Ross SA, Dunn BK. The Many Faces of Obesity and Its Influence on Breast Cancer Risk. Front Oncol. 2019;9:765.
52. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008;109(1):123-39.
53. Zahid H, Simpson ER, Brown KA. Inflammation, dysregulated metabolism and aromatase in obesity and breast cancer. Curr Opin Pharmacol. 2016;31:90-6.
54. Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B, et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. Lancet. 1998;351(9113):1393-6.
55. Visser M, Bouter L, McQuillan G, Wener M, Harris T. Elevated C-reactive protein levels in overweight and obese adults. JAMA. 1999;282:2131 - 5.
56. Harvey AE, Lashinger LM, Hursting SD. The growing challenge of obesity and cancer: an inflammatory issue. Ann N Y Acad Sci. 2011;1229:45-52.
57. Kershaw EE, Flier JS. Adipose tissue as an endocrine organ. The Journal of clinical endocrinology and metabolism. 2004;89(6):2548-56.
58. 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. Lancet. 1996;347(9017):1713-27.
59. Narod SA. Hormone replacement therapy and the risk of breast cancer. Nature Reviews Clinical Oncology. 2011;8(11):669-76.
60. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet. 2003;362.
61. Bakken K, Alsaker E, Eggen AE, Lund E. Hormone replacement therapy and incidence of hormone-dependent cancers in the Norwegian Women and Cancer study. Int J Cancer. 2004;112.
62. Investigators WGftWsHI. Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal WomenPrincipal Results From the Women's Health Initiative Randomized Controlled Trial. JAMA. 2002;288(3):321-33.
63. Veronesi U, Boyle P, Goldhirsch A, Orecchia R, Viale G. Breast cancer. Lancet. 2005;365(9472):1727-41.
64. Cancer CGoHFiB. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. The Lancet. 2019;394(10204):1159-68.
65. Fahlén M, Fornander T, Johansson H, Johansson U, Rutqvist LE, Wilking N, et al. Hormone replacement therapy after breast cancer: 10 year follow up of the Stockholm randomised trial. Eur J Cancer. 2013;49(1):52-9.

66. Godina C, Ottander E, Tryggvadottir H, Borgquist S, Isaksson K, Jernström H. Prognostic Impact of Menopausal Hormone Therapy in Breast Cancer Differs According to Tumor Characteristics and Treatment. Front Oncol. 2020;10:80.
67. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer. 2002;87(11):1234-45.
68. IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 2012 [Available from: http://monographs.iarc.fr/ENG/Monographs/vol100E/.
69. Jung S, Wang M, Anderson K, Baglietto L, Bergkvist L, Bernstein L, et al. Alcohol consumption and breast cancer risk by estrogen receptor status: in a pooled analysis of 20 studies. Int J Epidemiol. 2016;45(3):916-28.
70. Chen W, Rosner B, Hankinson S, Colditz G, Willett W. Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. JAMA. 2011;306:1884 - 90.
71. Ellison RC, Zhang Y, McLennan CE, Rothman KJ. Exploring the relation of alcohol consumption to risk of breast cancer. Am J Epidemiol. 2001;154(8):740-7.
72. Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. Jama. 2001;286(17):2143-51.
73. 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(7):535-40.
74. Tjønneland A, Thomsen BL, Stripp C, Christensen J, Overvad K, Mellemkaer L, et al. Alcohol intake, drinking patterns and risk of postmenopausal breast cancer in Denmark: a prospective cohort study. Cancer Causes Control. 2003;14(3):277-84.
75. Tjønneland A, Christensen J, Olsen A, Stripp C, Thomsen BL, Overvad K, et al. Alcohol intake and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition (EPIC). Cancer Causes Control. 2007;18(4):361-73.
76. Cao Y, Willett WC, Rimm EB, Stampfer MJ, Giovannucci EL. Light to moderate intake of alcohol, drinking patterns, and risk of cancer: results from two prospective US cohort studies. Bmj. 2015;351:h4238.
77. Seitz HK, Pelucchi C, Bagnardi V, Vecchia CL. Epidemiology and Pathophysiology of Alcohol and Breast Cancer: Update 20122012 2012-05-01 00:00:00. 204-12 p.
78. Wang Y, Xu M, Ke ZJ, Luo J. Cellular and molecular mechanisms underlying alcohol-induced aggressiveness of breast cancer. Pharmacol Res. 2017;115:299-308.
79. Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: a look at the evidence. Women's health (London, England). 2015;11(1):65-77.
80. Secretan B, Straif K, Baan R, Grosse Y, El Ghissassi F, Bouvard V, et al. A review of human carcinogens—Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. The Lancet Oncology. 2009;10(11):1033-4.
81. Seitz HK, Stickel F. Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism. Genes Nutr. 2010;5(2):121-8.
82. Dumitrescu RG, Shields PG. The etiology of alcohol-induced breast cancer. Alcohol. 2005;35(3):213-25.
83. Castro GD, Castro JA. Alcohol drinking and mammary cancer: Pathogenesis and potential dietary preventive alternatives. World J Clin Oncol. 2014;5(4):713-29.
84. Seitz HK, Stickel F. Risk factors and mechanisms of hepatocarcinogenesis with special emphasis on alcohol and oxidative stress. Biol Chem. 2006;387(4):349-60.
85. Hecht F, Pessoa CF, Gentile LB, Rosenthal D, Carvalho DP, Fortunato RS. The role of oxidative stress on breast cancer development and therapy. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(4):4281-91.
86. Ginsburg ES, Mello NK, Mendelson JH, Barbieri RL, Teoh SK, Rothman M, et al. Effects of alcohol ingestion on estrogens in postmenopausal women. Jama. 1996;276(21):1747-51.
87. Purohit V. Moderate alcohol consumption and estrogen levels in postmenopausal women: a review. Alcoholism, clinical and experimental research. 1998;22(5):994-7.

88. Terry PD, Rohan TE. Cigarette Smoking and the Risk of Breast Cancer in Women: A Review of the Literature. Cancer Epidemiology Biomarkers & Prevention. 2002;11(10):953-71.
89. Miller MD, Marty MA, Broadwin R, Johnson KC, Salmon AG, Winder B, et al. The association between exposure to environmental tobacco smoke and breast cancer: a review by the California Environmental Protection Agency. Prev Med. 2007;44(2):93-106.
90. Johnson KC. Accumulating evidence on passive and active smoking and breast cancer risk. Int J Cancer. 2005;117(4):619-28.
91. Catsburg C, Miller AB, Rohan TE. Active cigarette smoking and risk of breast cancer. Int J Cancer. 2015;136(9):2204-9.
92. Bjerkaas E, Parajuli R, Weiderpass E, Engeland A, Maskarinec G, Selmer R, et al. Smoking duration before first childbirth: an emerging risk factor for breast cancer? Results from 302,865 Norwegian women. Cancer Causes Control. 2013;24(7):1347-56.
93. Gram IT, Braaten T, Terry PD, Sasco AJ, Adami H-O, Lund E, et al. Breast Cancer Risk Among Women Who Start Smoking as Teenagers. Cancer Epidemiology Biomarkers & Prevention. 2005;14(1):61-6.
94. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat Rev Cancer. 2003;3(10):733-44.
95. Thompson PA, DeMarini DM, Kadlubar FF, McClure GY, Brooks LR, Green BL, et al. Evidence for the presence of mutagenic arylamines in human breast milk and DNA adducts in exfoliated breast ductal epithelial cells. Environmental and molecular mutagenesis. 2002;39(2-3):134-42.
96. Kispert S, McHowat J. Recent insights into cigarette smoking as a lifestyle risk factor for breast cancer. Breast cancer (Dove Medical Press). 2017;9:127-32.
97. Johnson KC, Miller AB, Collishaw NE, Palmer JR, Hammond SK, Salmon AG, et al. Active smoking and secondhand smoke increase breast cancer risk: the report of the Canadian Expert Panel on Tobacco Smoke and Breast Cancer Risk (2009). Tob Control. 2011;20(1):e2.
98. National Center for Chronic Disease P, Health Promotion Office on S, Health. Reports of the Surgeon General. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US); 2014.
99. Lagerlund M, Drake I, Wirfält E, Sontrop JM, Zackrisson S. Health-related lifestyle factors and mammography screening attendance in a Swedish cohort study. Eur J Cancer Prev. 2015;24(1):44-50.
100. Weed DL. Beyond black box epidemiology. Am J Public Health. 1998;88(1):12-4.
101. Park J, Choi J, Choi JY. Network Analysis in Systems Epidemiology. J Prev Med Public Health. 2021;54(4):259-564.
102. Lund E, Dumeaux V, Braaten T, Hjartåker A, Engeset D, Skeie G, et al. Cohort Profile: The Norwegian Women and Cancer Study—NOWAC—Kvinner og kreft. International Journal of Epidemiology. 2008;37(1):36-41.
103. Lund E, Dumeaux V. Systems Epidemiology in Cancer. Cancer Epidemiology, Biomarkers & Prevention. 2008;17(11):2954-7.
104. Crick FH. On protein synthesis. Symp Soc Exp Biol. 1958;12:138-63.
105. Crick F. Central dogma of molecular biology. Nature. 1970;227(5258):561-3.
106. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270(5235):467-70.
107. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27(8):1160-7.
108. Perou C, Sorlie T, Eisen M, van de R, Jeffrey S, Rees C, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747 - 52.
109. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27.
110. Eiliv L. Advancing Systems Epidemiology in Cancer. Exploring Trajectories of Gene Expression. Oslo: Skandinavian Universtity Press; 2020.
111. Lund E, Kumle M, Braaten T, Hjartaker A, Bakken K, Eggen E, et al. External validity in a population-based national prospective study – the Norwegian Women and Cancer Study (NOWAC). Cancer Causes Control. 2003;14.

112. Olsen KS, Krum-Hansen S. 4. Practical and Ethical Issues in Establishinga Collection of Normal Breast Tissue Biopsies—Part of the NOWACPost-Genome Cohort. In: Lund E, editor. Advancing Systems Epidemiology in Cancer. Oslo, Norway: Scandinavian University Press; 2020. p. 63-76.
113. Dumeaux V, Børresen-Dale A-L, Frantzen J-O, Kumle M, Kristensen VN, Lund E. Gene expression analyses in breast cancer epidemiology: the Norwegian Women and Cancer postgenome cohort study. Breast Cancer Research. 2008;10(1):1-8.
114. Dumeaux V, Fjukstad B, Fjosne HE, Frantzen JO, Holmen MM, Rodegerdts E, et al. Interactions between the tumor and the blood systemic response of breast cancer patients. PLoS Comput Biol. 2017;13(9):e1005680.
115. Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer. 2009;45(7):1218-31.
116. Norway CRo. Cancer in Norway 2019. 2022.
117. Bland JM, Altman DG. The logrank test. Bmj. 2004;328(7447):1073.
118. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.
119. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic acids research. 2012;40(17):e133.
120. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24(13):1547-8.
121. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
122. Bioconductor. lumiHumanIDMapping. [Available from: http://bioconductor.org/packages/lumiHumanIDMapping/.
123. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet. 2002;359(9302):248-52.
124. Bhopal R. Concepts of Epidemiology. New York, USA: Oxford University Press; 2008.
125. Wilson H, Botfield B, Speirs V. A Global View of Breast Tissue Banking. In: Karimi-Busheri F, editor. Biobanking in the 21st Century. Cham: Springer International Publishing; 2015. p. 69-77.
126. Eiliv L, Merethe K, Tonje B, Anette H, Kjersti B, Elise E, et al. External validity in a population-based national prospective study – the Norwegian Women and Cancer Study (NOWAC). Cancer Causes Control. 2003;14(10):1001-8.
127. Skeie G, Mode N, Henningsen M, Borch KB. Validity of self-reported body mass index among middle-aged participants in the Norwegian Women and Cancer study. Clinical epidemiology. 2015;7:313-23.
128. Lukic M, Nilsson LM, Skeie G, Lindahl B, Braaten T. Coffee consumption and risk of rare cancers in Scandinavian countries. European Journal of Epidemiology. 2018;33(3):287-302.
129. Dumeaux V, Olsen KS, Nuel G, Paulssen RH, Børresen-Dale A-L, Lund E. Deciphering Normal Blood Gene Expression Variation?The NOWAC Postgenome Study. PLoS Genet. 2010;6(3):e1000873.
130. Bianchini F, Kaaks R, Vainio H. Overweight, obesity, and cancer risk. Lancet Oncol. 2002;3(9):565-74.
131. Coradini D, Gambazza S, Oriana S, Ambrogi F. Gene expression profile of normal breast tissue and body mass index. Breast Cancer. 2021;28(2):488-95.
132. Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, et al. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes & Diseases. 2018;5(2):77-106.
133. Belitskaya-Levy I, Zeleniuch-Jacquotte A, Russo J, Russo I, Bordas P, Ahman J, et al. Characterization of a genomic signature of pregnancy identified in the breast. Cancer Prev Res. 2011;4:1457 - 64.
134. Peri S, de Cicco R, Santucci-Pereira J, Slifker M, Ross E, Russo I, et al. Defining the genomic signature of the parous breast. BMC Medical Genomics. 2012;5(1):46.

135. Rotunno M, Sun X, Figueroa J, Sherman ME, Garcia-Closas M, Meltzer P, et al. Parity-related molecular signatures and breast cancer subtypes by estrogen receptor status. Breast Cancer Research : BCR. 2014;16(1):R74-R.
136. Degnim AC, Visscher DW, Hoskin TL, Frost MH, Vierkant RA, Vachon CM, et al. Histologic Findings in Normal Breast Tissues: Comparison to Reduction Mammaplasty and Benign Breast Disease Tissues. Breast Cancer Res Treat. 2012;133(1):169-77.
137. Macfarlane LA, Murphy PR. MicroRNA: Biogenesis, Function and Role in Cancer. Curr Genomics. 2010;11(7):537-61.
138. Collins R, Balaconis MK, Brunak S, Chen Z, De Silva M, Gaziano JM, et al. Global priorities for large-scale biomarker-based prospective cohorts. Cell Genom. 2022;2(6):100141.

Paper I

Krum-Hansen, S., Wiik, A.B., Olsen, K.S., Lukic, M., Paulssen, R.H. & Lund, E.

Parity and cumulative incidence rates of breast cancer in the Norwegian Woman and Cancer Study (NOWAC)

Submitted manuscript

Parity and cumulative incidence rates of breast cancer in the Norwegian Woman and Cancer Study (NOWAC)

Sanda Krum-Hansen[1,2,4], Arne Bastian Wiik[1,3], Karina Standahl Olsen[1], Marko Lukic[1], Ruth H. Paulssen [4], Eiliv Lund [1]

Affiliations

1) Department of Community Medicine, UiT Faculty of Health Sciences, UiT- The Arctic University of Norway, N-9037 Tromsø, Norway

2) Department of Hematology and Oncology, Stavanger University Hospital, N-4019 Stavanger, Norway

3) Centre for Care Research West. Western Norway University of Applied Sciences (HVL), N-5063 Bergen, Norway

4) Department of Clinical Medicine, UiT Faculty of Health Sciences, UiT- The Arctic University of Norway, N-9037 Tromsø, Norway

Corresponding author:

Sanda Krum-Hansen, e-mail: sanda.krum-hansen@uit.no, sanda.krum-hansen@sus.no

Keywords: parity, cumulative incidence, breast cancer, Norwegian Women and Cancer Study, risk factors

Abstract

Background

The reduced risk of breast cancer (BC) following increasing parity has been known for decades. Most prospective studies have presented the relative risk as the percentage decrease for each child during follow-up. Since the risk reduction is up to ten percent for each child, the overall lifelong BC risk reduction could be under communicated. In this study we use

cumulative incidence rates (CIR) to calculate and describe the lifelong risk of BC in relation to parity.

Methods

NOWAC is a prospective cohort study with 172,000 women recruited between 1991 and 2007 with follow-up through questionnaires and national registers of cancer and death. For the present analyses, we included 165 238 women with follow-up from 01.01.2000 until 31.12. 2018. We used Poisson regression analysis to calculate the CIR of BC for parity, stratified by other established BC risk factors (maternal age at first birth, breastfeeding, body mass index (BMI), smoking and alcohol consumption).

Results

After 17.3 years of average follow-up, 8120 women aged 35-84 years developed breast cancer. Age-specific incidence rates increased for each age group up to 60-64 years, decreased for the age group 75-79 years, and increased again among the oldest women aged 80-84. CIR for all participants up to 84 years was 11.7%, close to the national cumulative incidence rate of 11.3%.

In analyses stratified by parity, the CIR of BC for nullipara was 12.6%. For women who gave birth to one child, CIR was 13.3%, and for 2, 3, 4, and 5-6 children, CIR was 11.8, 10.5, 9.6 and 8.7%, respectively. The parity-specific CIR of BC had the same pattern of decrease in analyses stratified for other BC risk factors.

Conclusion:

Cumulative incidence showed a consistent decrease in BC risk for each additional child. The decrease was consistent in strata of other established BC risk factors.

Keywords: parity, cumulative incidence, breast cancer, Norwegian Women and Cancer Study, risk factors

1 Introduction

Breast cancer (BC) is the most common cancer among women, and the leading cause of cancer death in women worldwide (1). Globally, BC incidence has increased over the last

three decades (2), and the same trend has been observed in European countries, including Norway (3).

A large number of risk factors for developing BC have been studied: reproductive history (age at first birth, parity and duration of breastfeeding), body mass index (BMI), physical activity, diet including fat intake, alcohol consumptions, smoking, air pollution and radiation (4). Reproductive history and length of exposure to endogenous hormones (age at menarche, age at menopause) have a strong influence on the risk of developing BC (5). Exposure to exogenous hormones, use of oral contraceptive and hormonal therapy, lead to a transient increased risk during use, and 10 years, respective 2 years after cessation (6, 7).

Numerous epidemiological studies of parity have provided evidence of higher risk of BC in nulliparous compared with parous women, and declining incidence of BC with increasing number of children (8, 9). In many large prospective cohort studies and meta-analyses, the protective effect of parity for development of postmenopausal BC is around 7-8 % for every full-term pregnancy (9, 10, 11, 12). Prolonged breast feeding provides additional risk reduction (13, 14).

Most epidemiological studies have used relative risk (RR) as a measure for association between exposure (parity) and outcome (BC). RR shows the risk of outcome over the follow-up time but tells little about lifetime risk.

To our knowledge, only two previous study used CIR to investigate lifetime risk of BC (15, 16) neither of which calculated the change in CIR of BC for each pregnancy.

The aim of this study was to assess the parity-related lifetime risk of BC in a large prospective cohort by using CIR, and to investigate the parity-specific CIR of BC in strata of others BC risk factors.

2 Material and methods

2.1 Study design and participants/population

In this study we used data from The Norwegian Women and Cancer Study (NOWAC) and from the national registers of cancer and death. The NOWAC study (17) is a population-based prospective cohort study in which a sample of women living in Norway was recruited between 1991 and 2007 by random sampling from the Norwegian Central Population Registry. A total of 172 748 women participated with at least one questionnaire. Disease

87 status, death and emigration status were updated through linkage to the Cause of Death
88 Registry at Statistics Norway and to the Cancer Registry of Norway (www.ssb.no,
89 www.kreftregisteret.no), using the national personal identification number. The exit date was
90 defined as either the date of cancer diagnosis, death, emigration, or end of follow-up on
91 December 31st, 2018. The date of entry into the present study was January 1st, 2000, because
92 in 2000, most large counties in Norway had started mammography screening for BC. Since
93 the screening was anticipated to increase the incidence rates in women aged 50-65 years, use
94 of information from before 2000 could be difficult to interpret. After excluding prevalent
95 cases of breast cancer and women with 7 children or more (for data anonymization purposes),
96 165 238 Norwegian women were included in analyses.

97 BC cases were identified through the Norwegian Cancer Registry identified according to
98 organ site code C50 in the International Classification of Diseases, Tenth Revision (ICD-10)
99 (https://icd.who).

100

101 ## 2.2 Questionnaire information

102 Information on parity and other covariates was collected from the NOWAC baseline
103 questionnaire, and information on parity was not updated later. Parity was analyzed both as
104 discrete numbers in the range from zero to six, and as four parity groups 0, 1-2, 3-4 and 5-6
105 children, respectively.

106 Breastfeeding was self-reported in months of duration per child and calculated by summing
107 the total duration of breastfeeding and dividing by number of children. In the analyses,
108 women who breastfed ≥6 months per child were compared to those who breastfed less than 6
109 months per child. Maternal age at first birth was calculated by the year of birth of the first
110 child.

111 BMI (kg/m^2) was calculated from self-reported data on weight and height. The variable was
112 dichotomized into BMI<25 and BMI≥25. The smoking status in NOWAC was defined as
113 current, former, or never smokers. We combined current and former smokers into a single
114 category of ever-smokers. Alcohol consumption was defined as ever-drinkers and teetotalers.

115 The age range for the analyses of parity was 40-84 years, with certain exceptions described
116 below to take into account assumed menopausal status. However, in the analyses stratified

117 according to other BC risk factors, this was reduced to 40-79 years due to few BC cases in
118 women aged 80-84. In the analyses stratified by alcohol use, we also excluded age group 75-
119 79 years for parity group 5-6 children due to zero cases in alcohol users' strata.

120 An additional change was done in analyses stratified by maternal age at the first birth; by
121 taking into account the age of menopause, the youngest age group was changed from 40-45 to
122 45-50 to capture complete fertility history. For analyses stratified by breastfeeding, the
123 number of parity groups was changed from three groups (0, 1-2, 3-4, 5-6) to two groups (1-3
124 and 4-6), due to few cases in each parity group.

125

126
2.3 Statistical methods

127 The CIR was calculated for each age-group as the number of incident cases of BC divided by
128 the number of person-years (per 100.000) until death, emigration, or a BC diagnosis. The
129 entire cohort was used for calculation of CIR by each variable, starting at age 35 years. In
130 calculations combining parity and exposure variable, age groups started at 40 years as a
131 surrogate for perimenopausal/postmenopausal risk state. There was no adjustment to account
132 for that women aged and potentially moved into different age bands during the study period.
133 We used the log-rank test to assess differences in CIR (18, 19). Poisson regression analysis
134 was used for calculation of CIR. The statistical software used was R statistical environment
135 and the Epi package for Epidemiological research (www.rstudio.com).

136
3 Results

137
3.1 Baseline characteristics of study population

138 The mean age at enrollment for all the women in the population was 49.5 years. The mean
139 number of children was 2.2 (standard deviation (SD) 1.2) and the mean maternal age at first
140 birth was 23.8 years (SD 4.2). The average follow-up time was 17.3 years. During the follow-
141 up period, 8120 cases of breast cancer were diagnosed.

142 Table 1 shows the distribution of exposures and baseline characteristics of the total study
143 population and according to parity groups. One third of the total study population had BMI
144 ≥25, most pronounced among those with high parity. The prevalence of alcohol consumption
145 and smoking showed no clear trend with parity. Age at first birth showed a very strong
146 gradient – women with higher parity have their first child at a younger age.

Age-specific incidence rates for BC in the NOWAC study increased from age 35-39 up to 60-64 years, decreased until 75-79 years, with another increase in the oldest age group (Figure 1a).

3.2 Parity-specific CIR of BC

The CIR for all age groups were 12.6% for nullipara. For women who gave birth to one child, CIR was 13.3%, and for those who gave birth to 2, 3, 4, and 5-6 children, CIR was 11.8%, 10.5 %, 9.6% and 8.7%, respectively (Figure 2a).

When using parity categorized into groups, CIR for 1-2 children was 12.1%, and for 3-4 children CIR was 10.2% (Figure 2b).

3.3 Parity-specific CIR of BC, stratified by others BC risk factors

In general, CIR of BC for women with BMI >25, women who smoked or used alcohol, was higher than CIR for women with BMI<25, who did not smoke or use alcohol.

For all analyses of parity-specific CIR of BC stratified by others BC risk factors, parity was categorized into groups. Overall, CIR of BC declines with increasing number of children. Within the same parity group, women exposed to others BC risk factors had higher CIR. With some exceptions, differences in CIR between parity groups were statistically significant (Table 2).

Age at first birth

The CIR for mothers who had their first child before the age of 25 was lower in all three parity groups compared to mothers who had their first child after the age of 25 (Figure 3).

Women who had their first child before the age of 25, had a CIR 1% lower than those who had their first child after 25 (parity groups 1-2 and 5-6). For women in parity group 3-4 the difference was 0.2% depending on age at first child (Table 2).

CIR for BC decreased with increasing number of children independent of age at first birth.

Breastfeeding

Due to few cases in each group, analyses of breastfeeding were divided into two parity groups, 1-3 and 4-6 children. The duration of breastfeeding was defined as below or above 6 months per child. Only small differences were found in both parity groups regarding duration of breastfeeding (Figure 4). Women with 1-3 children and shorter breastfeeding duration, had

176 a CIR of 9.9 % compared to 9.5% for those with longer duration (Table 2). For women with
177 4-6 children, CIR decreased with increasing number of children, independent of breastfeeding
178 duration.

179 ## BMI

180 In the total study population, CIR of women with BMI <25 was 10.8% compared to CIR 12.7
181 % for women with BMI ≥25 (Figure 5a, Table 2).

182 For nulliparous women with BMI <25, CIR was 11.2% compared with CIR 12.0% for
183 nulliparous with BMI ≥25. In each parity group, CIR of BC was higher for women with BMI
184 ≥25. CIR decreased in both BMI groups with increased number of children (Figure 5b, Table
185 2).

186 ## Alcohol use

187 In the total study population, CIR of women who did not consume alcohol was 11.0% and
188 12% for alcohol consumers (Figure 6a, Table 2).

189 For nulliparous women CIR was higher in alcohol consumers compared to non-consumers
190 (12.1% vs 10.8%).

191 In all parity groups, CIR was higher in alcohol consumers (Figure 6b, Table 2). CIR
192 decreased with increasing number of children in both consumers and non-consumers of
193 alcohol.

194 ## Smoking

195 CIR in the total population, was 12.4% for ever smokers and 13.6% for never smokers (Figure
196 7a, Table 2).

197 For nulliparous women, CIR was higher in non-smokers compared with smokers (13.9%
198 vs.12.7%). In all three parity groups, CIR was higher in ever smokers compared with never
199 smokers (Figure 7b, Table 2). As for analyses of the others BC risk factors, CIR decreased in
200 both smokers and non-smokers with increasing number of children.

201

4 Discussion

In this study, we describe the risk of BC in women 40 tears and older, in relation to parity. Our results show an average decrease in CIR of 0.8% per child. The parity-specific CIR had the same pattern of decrease according to parity, independent of other risk factors.

The age at first birth in this cohort was 23.8 years, lower than the average of 30 years in Norway today (20). The average number of children per woman was 2.2.

The age at first birth in Norwegian women has changed substantially over the last decades. In year 2000, women aged 35-39 years gave birth to 14.2% of all born that year. In 2018 this number had increased to 20.4.% (21). From 40 years these numbers were reduced to less than 2%. Consequently, most analyses started at age 40 years to capture the effect of childbirths taking place later in life.

Reproductive history is the strongest known modifier of a woman's breast cancer risk. Full-term pregnancy at early age and increasing number of children both lower the BC risk in postmenopausal women (9).

The association between parity and risk of BC is complex, but mainly determined by age at first birth and number of full-time pregnancies (9). The associations between the parity and risk of BC found in this study are consistent with existing research. Studies have found that multiple pregnancies reduce a woman's risk of developing BC by 8% for each additional pregnancy (8, 9)

There are few studies using CIR as risk estimate for BC according to parity. Colditz et al., used data from the Nurses' Health Study to calculate the cumulative risk of BC up to the age of 70 years. They showed that women with multiple births and a first birth at an early age, had a reduced risk of BC at or after menopause, relative to nulliparous women. However, they did not stratify for parity (16). A study by The Collaborative group on hormonal factors, estimated a 50% reduction of cumulative incidence of BC in developed countries, if women had the same average number of births as women in developing countries. They emphasized breastfeeding as a factor that could account for almost two-thirds of this estimated difference (13).

In our study, breastfeeding showed an effect on risk reduction of BC only among women with 5-6 children, or three years or more of lactation. This is in contrast with the Oxford

collaborative group which showed a relative risk reduction of BC for every 12 months of breastfeeding. Breastfeeding rates in Norway were low at the end of the 1960s and the 1970. Bottle feeding was modern and promoted by formula companies. At that time, exclusive breastfeeding, beyond 4 months was not advocated (22). This could be part of the explanation of the different risk estimates, as well as use of a different cutoff for breastfeeding time, compared to the Collaborative group results.

Potential mechanisms behind early full-term pregnancy induced protection against BC are not fully understood. Several studies, including gene expression and epigenetic studies, explored the hypothesis that first full-term pregnancy affects remodeling in mammary tissue by inducing differentiation of breast cells that make them less sensitive to carcinogenic influence (23, 24, 25) In our study, each additional childbirth reduces BC risk in a linear fashion starting from the first birth. Hence, our findings do not support the hypothesis that it is the first pregnancy that has a decisive role in the protection against postmenopausal BC.

Husby et al showed that pregnancies lasting 34 weeks or longer were associated with considerable risk reduction, compared with pregnancies lasting less than 34 weeks (26). In the same study they showed that both live and stillbirths were associated with risk reduction given that the pregnancy lasted 34 weeks or longer. These overall findings strengthen the hypothesis that parity, and not breastfeeding, is the main driver of the risk reduction. The distinctive difference in risk of BC by pregnancies lasting longer or shorter than 34 weeks, suggests that biological processes that take place towards the end of the pregnancy may change the BC susceptibility of the mother. Lund et al highlighted involvement of the immune system in parity-associated BC protection, with the use of a systems epidemiology approach (8). In sum, the protective effect of parity can be due to processes that take place locally in the breast tissue, or systemic, involving the immune system.

Analyses for selected risk factors showed consistently increased CIR of BC for those exposed to selected risk factors. However, exposure to risk factors did not affect protective effect of increasing parity. The reduction in CIR with increasing parity in most sub-strata demonstrates that the protective effect of parity is mainly independent of other risk factors.

In our study we find no difference in CIR of BC in women < 65 years of age in relation to BMI. Women aged 65 and older, with BMI ≥ 25, had a higher risk of BC compared with women in the same age group with BMI <25. Previous studies showed that higher BMI in

263 premenopausal age is protective for BC but in postmenopausal age, higher BMI increases the
264 risk of BC (27). Our results show that BMI had the same effect on risk of BC in all parity
265 groups, with the largest difference in CIR in the highest parity group of 5-6 children. BMI did
266 not change the effect of parity on BC risk.

267 Alcohol consumption increased CIR of BC by 1% in our population. CIR decreased with
268 increased parity up to the age of 74. This is in line with other studies that show an association
269 between alcohol consumption and BC risk (28, 29).

270 Overall, smokers in our study had higher CIR than non-smokers. Stratified by parity,
271 nulliparous women had slightly lower CIR than women with one child.

272 Smoking is linked to a higher risk of BC in women who start smoking at a younger age,
273 before they give birth to their first child (30). Higher risk for uniparous smokers then
274 nulliparous smokers is also described by Andersen at al. (31). One explanation for this may be
275 that women who smoke reach menopause earlier (32).

276 5 Strengths and limitation

277 The strengths of the current study are prospective design and random selection of female
278 participants. To ensure a representative sample, each participant was randomly selected from
279 the national population register, with birth year serving as the sole criterion for sample
280 differentiation. Previous analyses have confirmed that the participants in the NOWAC study
281 exhibit the same level of fertility as the general population (23). In addition, the external
282 validation of the CIR using data from the National Cancer Registry for the same period
283 demonstrated a high degree of representativeness. The study is prospective with information
284 on parity and other risk factors collected before start of follow-up. While the statistical power
285 of the total study was high, the stratification left several analyses with few cases, especially
286 below 40 and above 80 years of age. Hence, in the stratified analyses these age-groups were
287 removed, and our results cannot be extrapolated to these age groups.

288 This study has some limitations related to the questionnaire information. Self-reported height
289 and weight for BMI calculations may be biased. However, in a validation study in NOWAC,
290 self-reported weight and height provided a valid ranking of BMI for middle-aged Norwegian
291 women (33). Information on smoking and alcohol consumption was self-reported and did not
292 include additional information on history or intensity of exposure. In addition, smoking and

alcohol are well known confounders. Variables were not validated in NOWAC but smoking status at baseline and follow-up questionnaires were compared in 2018 (34); only 1.8% of participants reported that they were ever smokers at baseline and never smokers at follow-up. A low degree of non- differential misclassification is therefore expected. Parity status was not updated after the baseline questionnaire and is a potential source of differential misclassification.

Most epidemiological studies use RR for calculating risk for BC. In this study we use CIR as risk estimate. CIR estimates the risk of developing BC within a given age-interval. The comparison of CIR between two strata is comparable to a standardization to a square population with the same number of individuals in each age group (35). Risk factors for BC change throughout women's life and affect risk of BC differently depending on the age; using CIR as measure of risk we can attenuate this problem.

Further, our study lacks data on hormone receptor status of each woman's BC. Recent studies show that parity increases the risk of hormone receptor negative BC, while breastfeeding modifies this risk. Future studies should aim to include this variable so that CIR for both hormone receptor positive and negative BC can be calculated separately by parity.

6 Conclusion

Our findings suggest that parity is a protective factor for the development of breast cancer, irrespective of other established risk factors, including those related to pregnancy and other lifestyle factors. Notably, each childbirth, rather than only the first, is associated with a reduction in breast cancer risk among postmenopausal women.

Declarations

Ethical approval and consent to participate.

The NOWAC study was approved by the Norwegian Data Inspectorate and the Regional Ethical Committee of North Norway/ The Regional Committee for Medical and Health research Ethics. The study was conducted in compliance with the Declaration of Helsinki, and all participants gave written informed consent. The linkages of the NOWAC database to national registries such as the Cancer Registry of Norway and registries on death and emigration were approved by the Directorate of Health. The women were informed about these linkages.

Availability of data and material

The analyses were based on anonymous group data. The data is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by the University in Tromsø, The Arctic University of Norway.

Authors contributions

Study design: EL. Data analysis: EL, ABW and SKH. Data/results interpretation: EL and SKH. Funding acquisition: EL. Writing, review and/or revision of the manuscript: SKH, EL, ABW, RHP, ML and KSO. All authors reviewed and approved the final manuscript.

Acknowledgments

We would like to thank all participants and staff in the NOWAC study for their contributions.

Some of the data in this article are from the Cancer Registry of Norway. The Cancer Registry of Norway is not responsible for the analysis or interpretation of the data presented.

Abbreviations

NOWAC: Norwegian Women and Cancer Study

BC: Breast cancer

BMI: Body mass index

343 HT: hormone therapy

344 RR: relative risk

345 ## Additional files

346 ## References

347

348 1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global
349 Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36
350 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.
351 2. Sharma R. Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR)
352 are associated with human development, 1990-2016: evidence from Global Burden of Disease
353 Study 2016. Breast Cancer. 2019;26(4):428-45.
354 3. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al.
355 Cancer statistics for the year 2020: An overview. Int J Cancer. 2021.
356 4. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast
357 Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current
358 Treatment Strategies—An Updated Review. Cancers (Basel). 2021;13(17):4287.
359 5. Horn J, Vatten LJ. Reproductive and hormonal risk factors of breast cancer: a
360 historical perspective. Int J Womens Health. 2017;9:265-72.
361 6. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data
362 on 53 297 women with breast cancer and 100 239 women without breast cancer from 54
363 epidemiological studies. Lancet. 1996;347(9017):1713-27.
364 7. Narod SA. Hormone replacement therapy and the risk of breast cancer. Nature
365 Reviews Clinical Oncology. 2011;8(11):669-76.
366 8. Lund E, Nakamura A, Snapkov I, Thalabard JC, Olsen KS, Holden L, et al. Each
367 pregnancy linearly changes immune gene expression in the blood of healthy women
368 compared with breast cancer patients. Clin Epidemiol. 2018;10:931-40.
369 9. Ewertz M, Duffy SW, Adami HO, Kvåle G, Lund E, Meirik O, et al. Age at first birth,
370 parity and risk of breast cancer: a meta-analysis of 8 studies from the Nordic countries. Int J
371 Cancer. 1990;46(4):597-603.

372 10. Clavel-Chapelon F. Differential effects of reproductive factors on the risk of pre- and
373 postmenopausal breast cancer. Results from a large cohort of French women. Br J Cancer.
374 2002;86(5):723-7.
375 11. Ritte R, Lukanova A, Berrino F, Dossus L, Tjonneland A, Olsen A, et al. Adiposity,
376 hormone replacement therapy use and breast cancer risk by age and hormone receptor status:
377 a large prospective cohort study. Breast Cancer Res. 2012;14:R76.
378 12. Lund E, Arnesen E, Borgan JK. Pattern of childbearing and mortality in married
379 women--a national prospective study from Norway. J Epidemiol Community Health.
380 1990;44(3):237-40.
381 13. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47
382 epidemiological studies in 30 countries, including 50302 women with breast cancer and
383 96973 women without the disease. Lancet. 2002;360:187 - 95.
384 14. Zhou Y, Chen J, Li Q, Huang W, Lan H, Jiang H. Association between breastfeeding
385 and breast cancer risk: evidence from a meta-analysis. Breastfeed Med. 2015;10(3):175-82.
386 15. Brenner H, Stegmaier C. Projected impact of the trend toward delayed childbearing on
387 breast cancer incidence in the Saarland/FRG. Soz Praventivmed. 1990;35(1):1-4.
388 16. Colditz GA, Rosner B. Cumulative Risk of Breast Cancer to Age 70 Years According
389 to Risk Factor Status: Data from the Nurses' Health Study. Am J Epidemiol.
390 2000;152(10):950-64.
391 17. Lund E, Dumeaux V, Braaten T, Hjartåker A, Engeset D, Skeie G, et al. Cohort
392 Profile: The Norwegian Women and Cancer Study—NOWAC—Kvinner og kreft. Int J
393 Epidemiol. 2007;37(1):36-41.
394 18. Bland JM, Altman DG. The logrank test. BMJ. 2004;328(7447):1073.
395 19. Cox DR. Regression Models and Life-Tables. In: Kotz S, Johnson NL, editors.
396 Breakthroughs in Statistics: Methodology and Distribution. New York, NY: Springer New
397 York; 1992. p. 527-41.
398 20. Statistics Norway [Available from: www.ssb.no.
399 21. Medical Birth Registry of Norway: Norwegian Institute of Public Health; [Available
400 from: https://www.fhi.no/hn/helseregistre-og-registre/mfr/.
401 22. Njå A. Veiledning i spedbarnsernæring o spedbarnsstell. In: Helsedirektøren, editor.
402 Bergen1965.

403 23. Peri S, de Cicco R, Santucci-Pereira J, Slifker M, Ross E, Russo I, et al. Defining the
404 genomic signature of the parous breast. BMC Med Genomics. 2012;5(1):46.
405 24. Russo I, Koszalka M, Russo J. Comparative study of the influence of pregnancy and
406 hormonal treatment on mammary carcinogenesis. Br J Cancer. 1991;64:481 - 4.
407 25. Russo J, Moral R, Balogh G, Mailo D, Russo I. The protective role of pregnancy in
408 breast cancer. Breast Cancer Res. 2005;7:131 - 42.
409 26. Husby A, Wohlfahrt J, Øyen N, Melbye M. Pregnancy duration and breast cancer risk.
410 Nature Communications. 2018;9(1):4255.
411 27. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from
412 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without
413 breast cancer. Lancet. 1997;350.
414 28. Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and
415 experimental evidence and potential mechanisms. JAMA. 2001;286(17):2143-51.
416 29. Heath AK, Muller DC, van den Brandt PA, Papadimitriou N, Critselis E, Gunter M, et
417 al. Nutrient-wide association study of 92 foods and nutrients and breast cancer risk. Breast
418 Cancer Res. 2020;22(1):5.
419 30. Bjerkaas E, Parajuli R, Weiderpass E, Engeland A, Maskarinec G, Selmer R, et al.
420 Smoking duration before first childbirth: an emerging risk factor for breast cancer? Results
421 from 302,865 Norwegian women. Cancer Causes Control. 2013;24(7):1347-56.
422 31. Andersen ZJ, Jørgensen JT, Grøn R, Brauner EV, Lynge E. Active smoking and risk
423 of breast cancer in a Danish nurse cohort study. BMC Cancer. 2017;17(1):556.
424 32. Whitcomb BW, Purdue-Smithe AC, Szegda KL, Boutot ME, Hankinson SE, Manson
425 JE, et al. Cigarette Smoking and Risk of Early Natural Menopause. Am J Epidemiol.
426 2017;187(4):696-704.
427 33. Skeie G, Mode N, Henningsen M, Borch KB. Validity of self-reported body mass
428 index among middle-aged participants in the Norwegian Women and Cancer study. Clin
429 Epidemiol. 2015;7:313-23.
430 34. Lukic M, Nilsson LM, Skeie G, Lindahl B, Braaten T. Coffee consumption and risk of
431 rare cancers in Scandinavian countries. Eur J Epidemiol. 2018;33(3):287-302.
432 35. Hakulinen T. Cancer Incidence in Five Continents Vol. VII. D. M. Parkin, S. L.
433 Whelan, J. Ferlay, L. Raymond and J. Young (eds), IARC Scientific Publications No. 143,

434 Lyon, 1997. No. of pages:1240. Price: £129. ISBN 92 832 2143 5. Stat Med.
435 2000;19(9):1261-3.
436

437 Figure 1a. Age-specific incidence rates in NOWAC 2000-2018, by age groups.

438 Figure 1b. Cumulative incidence of breast cancer in the NOWAC cohort compared to national
439 figures for the same years based on the Cancer Registry of Norway, 40-89 years.

440 Figure 2a. Cumulative incidence rates according to parity in NOWAC 2000-2018

441 Figure 2b. Cumulative incidence rates according to parity groups: 0, 1-2, 3-4, 5-6.

442 Figure 3. Cumulative incidence rates for breast cancer after parity stratified by age at first
443 birth less than 25 years, or 25+.

444 Figure 4. Cumulative incidence rates for breast cancer among women with more than 6
445 months breastfeeding (BF+) for each child compared to women with shorter breastfeeding
446 (BF-).

447 Figure 5a. Cumulative incidence rates for breast cancer according to body mass index (BMI).
448 BMI less than 25 (25-) and BMI 25 or more (25+).

449 Figure 5b. Cumulative incidence rates for breast cancer according to body mass index (BMI)
450 by parity. BMI less than 25 (25-) and BMI 25 or more (25+).

451 Figure 6a. Cumulative incidence rates for breast cancer stratified by alcohol consumption.
452 Never drinking (non-d) versus alcohol consumption (drink).

453 Figure 6b. Cumulative incidence rates for breast cancer for parity stratified by alcohol
454 consumption. Never drinking (non-d) versus alcohol consumption (drink).

455 Figure 7a. Cumulative incidence rates for breast cancer stratified by smoking (smk) versus
456 non-smokers (non-smk).

457 Figure 7b. Cumulative incidence rates for breast cancer for parity stratified by smoking (smk)
458 versus non-smokers (non-smk).

459

460

461

462 Figure 1a. Age-specific incidence rates in NOWAC 2000-2018.

463

464

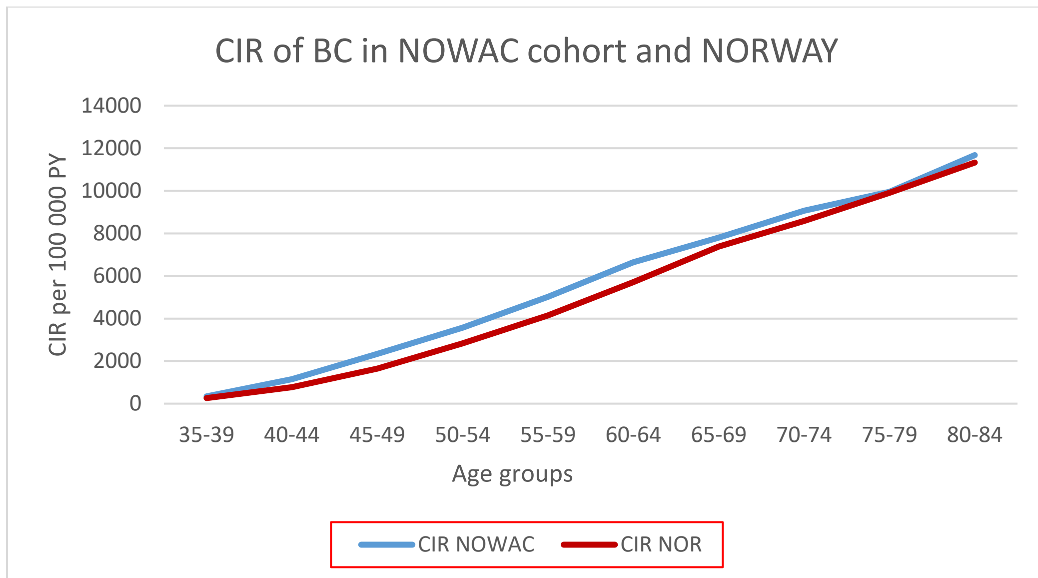
465

466

467

468 Figure 1b. Cumulative incidence of breast cancer in the NOWAC cohort compared to national
469 figures for the same years based on the Cancer Registry of Norway

470



471

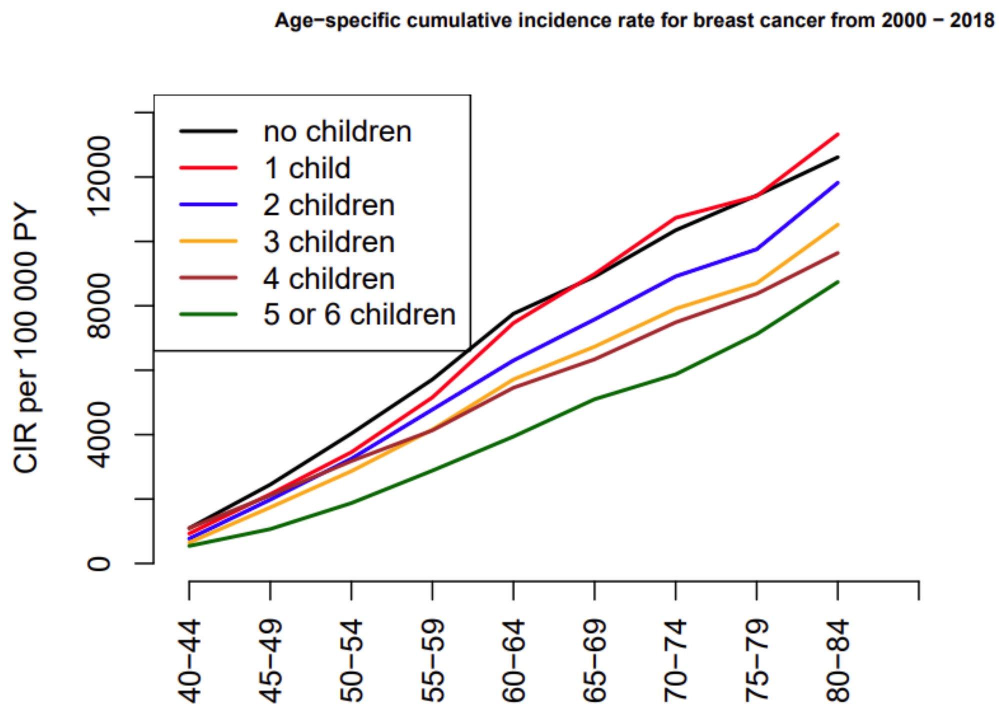
472

473

474

475

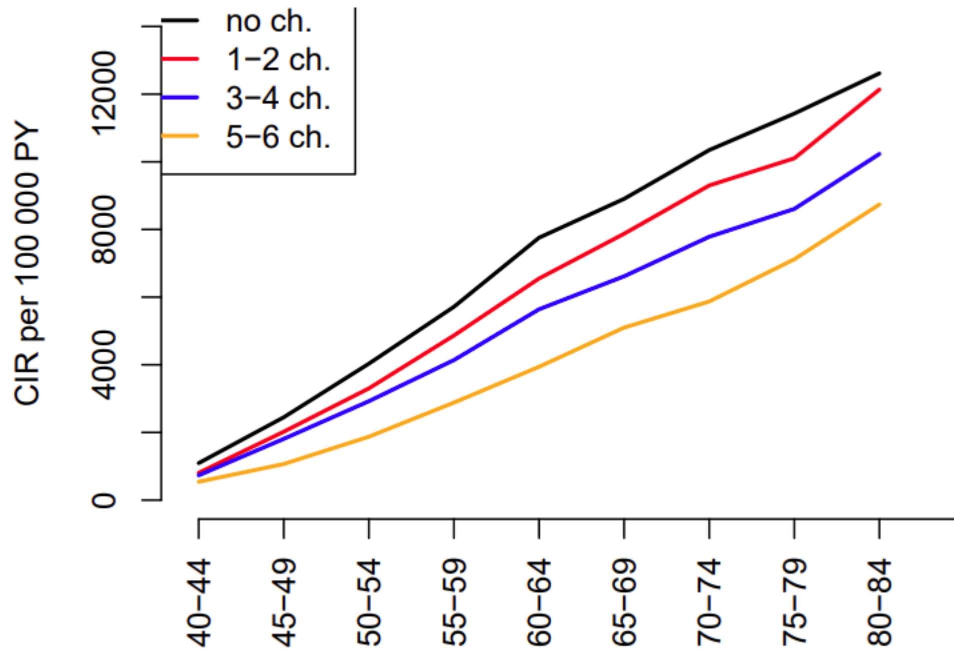
476 Figure 2a and 2b. Cumulative incidence rates according to parity and parity cohorts in
477 NOWAC 2000-2018



478

479

480



481

482

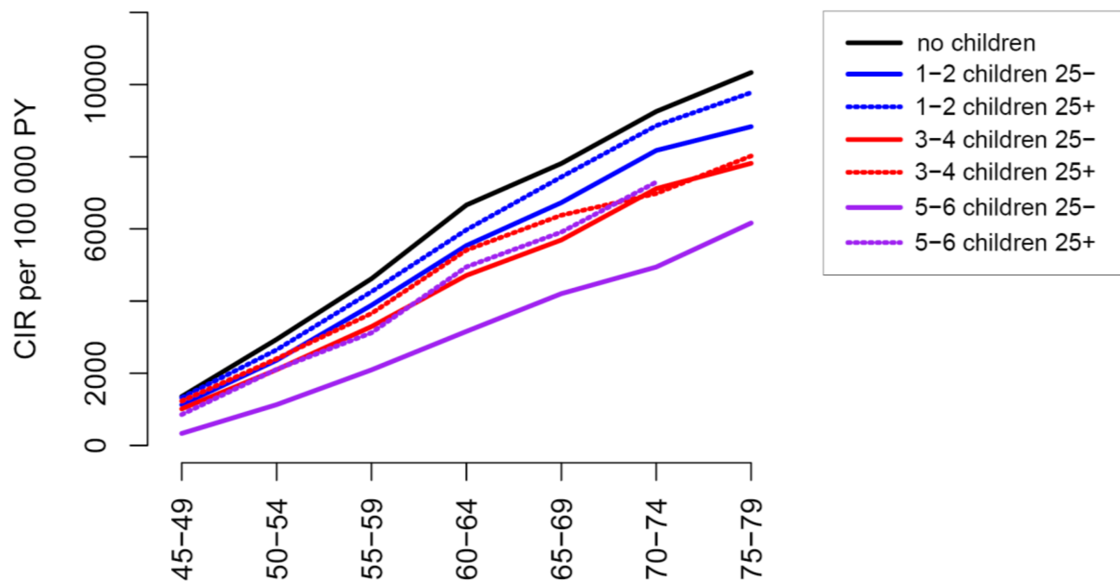
483

484

485 Figure 3. Cumulative incidence rates for breast cancer by parity, stratified by age at first birth
486 less than 25 years, or ≥25 years.

487

488



489

490

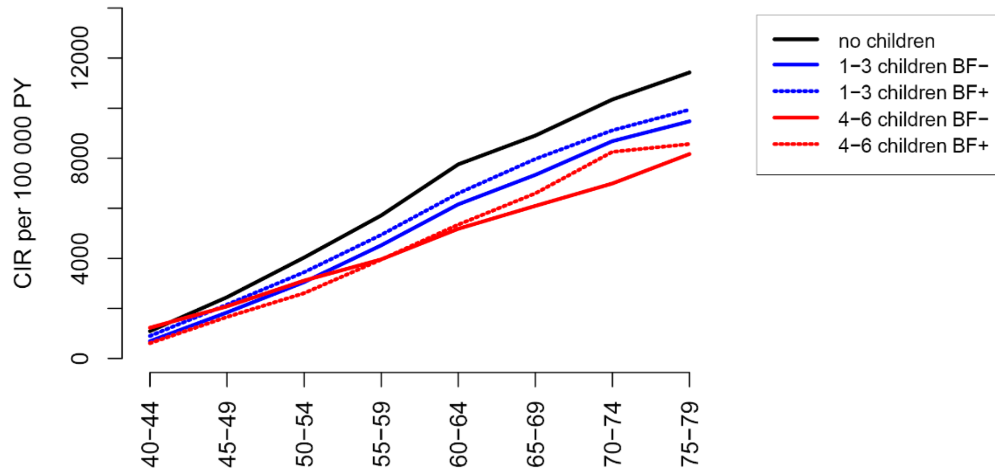
491

492

493

494 Figure 4. Cumulative incidence rates for breast cancer among women with more than 6
495 months breastfeeding (BF+) for each child compared to women with shorter breastfeeding
496 (BF-).

497

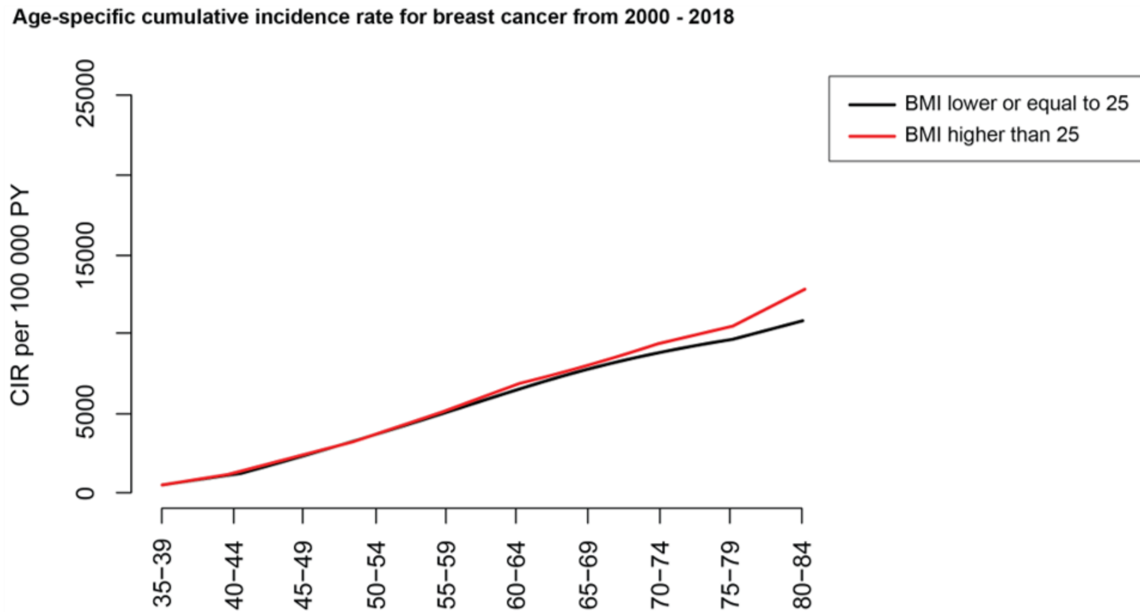


498

499

500 Figure 5a. Cumulative incidence rates for breast cancer according to body mass index (BMI).
501 BMI less than 25 (25-) and BMI 25 or more (25+).

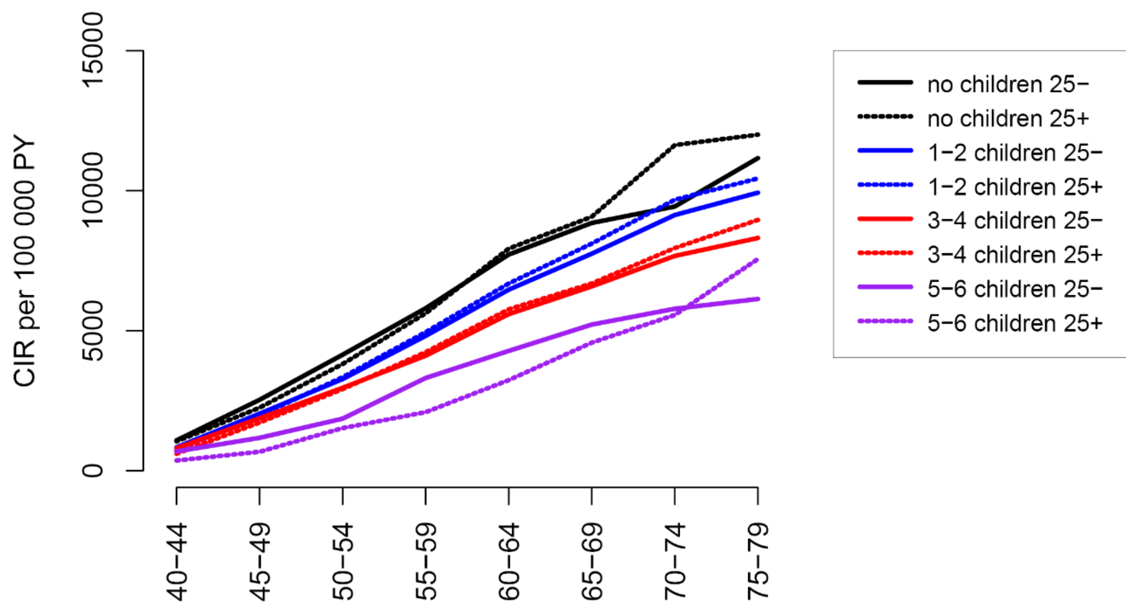
502



503

504

505 Figure 5b. Cumulative incidence rates for breast cancer according to body mass index (BMI)
506 by parity. BMI less than 25 (25-) and BMI 25 or more (25+).

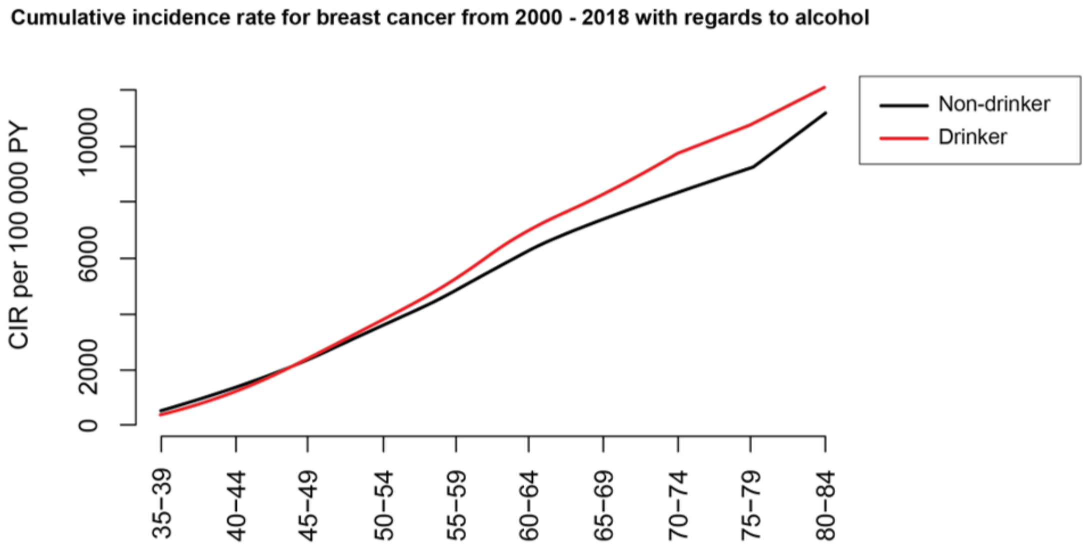


507

508

509 Figure 6a. Cumulative incidence rates for breast cancer stratified by alcohol consumption.
510 Never drinking (non-d) versus alcohol consumption (drink).

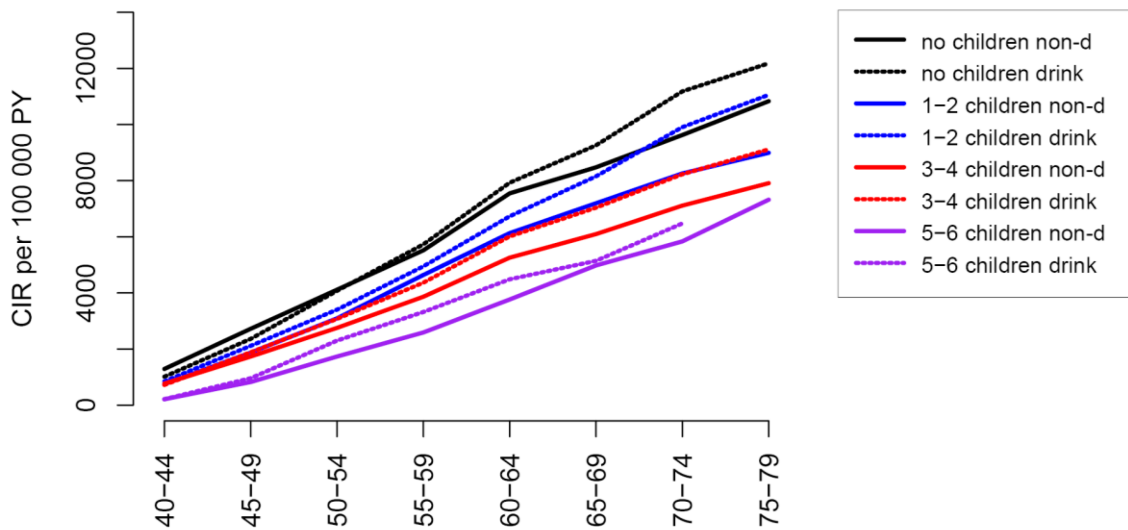
511



512

513

514 Figure 6b. Cumulative incidence rates for breast cancer for parity stratified by alcohol
515 consumption. Never drinking (non-d) versus alcohol consumption (drink).



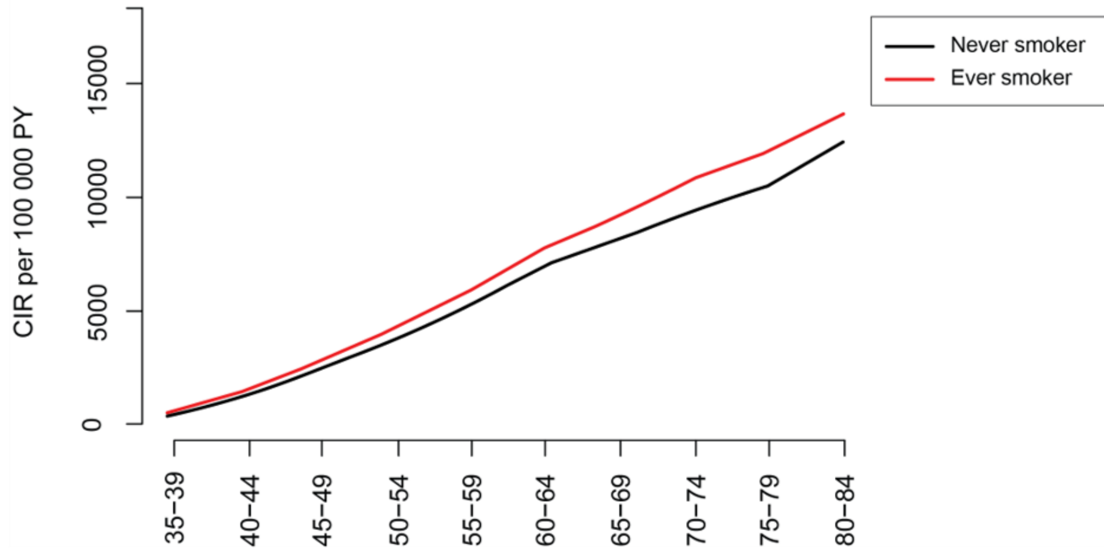
516

517

518

519 Figure 7a. Cumulative incidence rates for breast cancer stratified by smoking (smk) versus
520 non-smokers (non-smk).

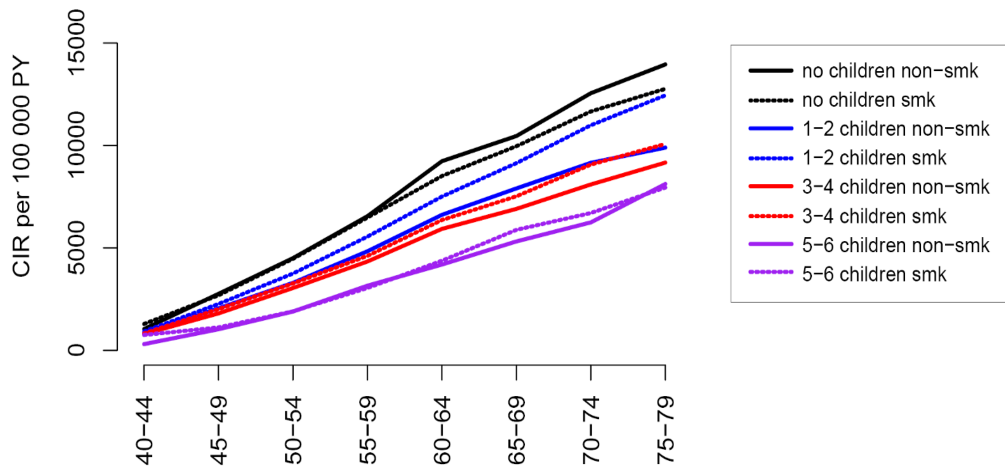
521



522

523

524 Figure 7b. Cumulative incidence rates for breast cancer for parity stratified by smoking (smk)
525 versus non-smokers (non-smk).



526

527

528

529

530 Table 1: Baseline characteristics of study population according to parity

531

Characteristics of study population	Total	Parity	Parity	Parity	Parity	Missing
	N %	0	1-2 children	3-4 children	5-6 children	0
Number of women	164105	16124	88243	55280	4458	0
Age at 1st birth, <25 n, % >=25 n, %	 89945 54.81% 57792 35.22%	N.a. N.a.	 46643 52.86% 41420 46.94%	 39604 71,64% 15621 28,26%	 3698 82.95% 751 16.85%	244 0.15%
BMI <=25 n, % > 25 n, %	 104529 63.70% 55501 33.82%	 10448 64.80% 5676 35.20%	 58374 66.15% 29869 33.85%	 33606 60.79% 21674 39.21%	 2101 47.13% 2357 52.87%	4075 2.48%
Alcohol use Yes n, % No n, %	 93215 56.8% 62051 37.81%	 9374 58.14% 5792 35.2%	 54721 62.01% 29244 33.14%	 27851 50.38% 24205 43.79%	 1269 28.47% 2810 63.03%	8839 5.39%
Smoking status Ever n, % Never n, %	 105912 64.54% 55985 34.12%	 9016 55.92% 6016 37.31%	 59825 67.8% 27340 30.98%	 33906 61.34% 20613 37.29%	 2350 52.71% 2016 45.22%	2208 1.35%

532

533

534 7 Table 2: Percentage difference in CIR between the two substrata for each risk factor.
535

Risk factor (parity groups) **whole population*	*CIR*	*Difference*	*p-value from log-rank test*
BMI (<=25/>25) *	10.8%/12.7%	1.9%	
BMI (<=25/>25(0)	11.2%/12.0%	0.8%	0.0092
BMI <=25/>25(1-2)	9.9%/10.4%	0.5%	<0.001
BMI <=25/>25(3-4)	8.3%/8.9%	0.5%	<0.001
BMI <=25/>25(5-6)	6.1/7.5%	1.4%	0.0109
Age at 1st birth <25/>25 (1-2)	8.8%/9.8%	1%	0.1065
Age at 1st birth <25/>25 (3-4)	7.8% / 8%	0.2%	<0.001
Age at 1st birth <25/>25 (5-6)	6.1%/7.3%	0.8%	0.0008
Smoker: Never vs Ever *	12.4% vs 13.6%	1.2%	
Smoker: Never vs Ever (0)	13.9% vs 12.7%	-0.2%	<0.001
Smoker: Never vs Ever (1-2)	9.8% vs 12.4%	2.6%	<0.001
Smoker: Never vs Ever (3-4)	9.1% vs 10%	0.8%	0.0020
Smoker: Never vs Ever (5-6)	8.1% vs 7.9%	-0.2%	0.4258
Alcohol use: no/yes/ *	11%/12%	1%	
Alcohol use: no/yes/ (0)	10.8%/12.1%	1.3%	0.0033
Alcohol use: no/yes/ (1-2)	8.9%/11.0%	1.1%	<0.001
Alcohol use: no/yes/ (3-4)	7.9%/9.1%	1.2%	0.0001
Alcohol use: no/yes/ (5-6)	5.8%/6.5%	0.7%	0.6341
Breastfeeding >6mnt/child vs.<6mnt/ child (1-3)	9.9% vs 9.5%	-0.4%	<0.001
Breastfeeding >6mnt/child vs.<6mnt/child (4-6)	8.6% vs 8.2%	-0.4%	<0.001

536
537
538

Paper II

Krum-Hansen, S. & Olsen, K.S. (2020).

Chapter 4. Practical and Ethical Issues in Establishing a Collection of Normal Breast Tissue Biopsies— Part of the NOWAC Post-Genome Cohort

In: Lund, E. (Ed.), *Advancing Systems Epidemiology in Cancer. Exploring Trajectories of Gene Expression* (pp 63-76). Scandinavian University Press.

Lund, E. (Ed.) (2020). *Advancing Systems Epidemiology in Cancer. Exploring Trajectories of Gene Expression.*
Oslo: Scandinavian University Press.
DOI: https://doi.org/10.18261/9788215041193-2020-04

4. Practical and Ethical Issues in Establishing a Collection of Normal Breast Tissue Biopsies—Part of the NOWAC Post-Genome Cohort

Sanda Krum-Hansen and Karina Standahl Olsen

Abstract For tissue-based studies of breast cancer, getting access to truly normal, well-annotated tissue can be a challenge. To address that need, we collected 368 breast tissue biopsies and buffered blood samples from healthy postmenopausal women. Volunteers were part of the Norwegian Women and Cancer (NOWAC) Post-genome cohort, recruited through the national mammography screening program. The NOWAC normal breast tissue biobank for gene expression analysis will provide a correct basis for comparison in case-control studies.

Keywords normal breast tissue | biobank | breast cancer

BACKGROUND

Epidemiology and risk factors of breast cancer

Breast cancer is the most frequent type of cancer among females worldwide. The latest GLOBOCAN report estimated approximately 2.1 million newly diagnosed breast cancers in 2018 (Bray et al. 2018).The incidence of breast cancer varies significantly around the world, but is increasing in most countries (Bray et al. 2018). The high incidence in developed countries has to some extent been counterbalanced by a reduction in mortality. Early diagnosis due to mammographic screening, improved treatment, secondary prophylaxis and follow-up have improved the outcome for breast cancer patients. The 5-year survival rate in Norway is 90.4%—yet breast cancer is the leading course of cancer-related deaths among females

This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0).
To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

(Cancer Registry of Norway 2017). The increasing incidence and improved survival rate results in high prevalence of the disease. Since the treatment is associated with severe side effects over a long period, the burden of the disease is large.

The current body of evidence suggests that genetic structure and internal and external risk factors, as well as their interactions, combine to constitute the causes of breast cancer. Two major risk factors are gender and age. Other causal factors relate to the levels of endogenous hormones determined by age at the first menstruation, age at menopause, age at first birth, and number of births, as well as use of oral contraceptives and hormone therapy (HT) (Kaminska et al. 2015). Lifestyle factors regarded as risk factors include lack of physical activity, obesity, alcohol consumption, smoking, night shift work, exposure to radiation, and possibly diet (Sun et al. 2017). Hereditary breast cancer accounts for 5–10% of cases (Apostolou and Fostira 2013), making non-hereditary risk factors the major drivers of incidences of breast cancer.

Breast cancer characteristics

Breast cancer is a heterogeneous disease both etiologically and genetically. It consists of several sub-types with different molecular profiles, and biological and clinical behavior. Different sub-groups are associated with different risk profiles and present a big challenge for clinical management. In clinical practice, an array of methods is used to determine which sub-type the patient has: tumor-node-metastasis (TNM) staging, histological sub-typing, tumor grade, tumor invasion in lymphatic and vascular tissue, axillary lymph node status, immune-histochemical staging providing estrogen and progesterone receptor status, presence of human epidermal growth factor receptor 2 (HER2) receptor, and Ki67 marker. These factors describe the tumor biology regarding hormone sensitivity and tumor aggressiveness, guide decision-making for treatment, and predict the prognosis.

Today there are efficient surgical and medical treatments available, but we are unable to determine specifically which type of treatment the individual patient needs, often implying overtreatment. There is a need for better prognostic and predictive markers to individualize the treatment in order to provide the best treatment for patients with high-risk profiles, and to avoid overtreatment of patients with a low risk profile.

Normal breast tissue histology and development

The human breast is an apocrine gland designed to produce milk, and breast tissue is heterogeneous and complex in composition. The breast consists of three main

components: the skin, containing areola and nipple, the subcutaneous adipose tissue (white fat tissue), and the glandular tissue (functional tissue of the breast) including both parenchyma and stroma. The parenchyma is divided into 15–25 lobes, each made up of 20–40 lobules. The structure is based on a branching duct system that leads from the collecting ducts to the terminal duct-lobular units (TLUs). The TLUs are the functional unit of the breast tissue and sites of milk production. The terminal collecting ducts drain the milk from TDLUs into 4–18 lactiferous ducts, which drain to the nipple. The inter- and perilobular connective tissue surrounding the TLUs and lobules contain fibrovascular tissue and white adipose tissue. Fibrous stroma provides the background architecture for the glandular tissue, as well as nutrition and protection. The proportion of adipose and fibrous tissues varies from one woman to another and changes in the same person over time.

Breast tissue development occurs in defined stages: embryonic, pre-pubertal, pubertal, pregnancy, lactation and involution. The tissue only reaches its final level of development during the last stages of pregnancy, and if pregnancy does not occur, it is never reached. During menopause, the glandular tissue is progressively atrophied. The lobules decrease in size and number, mainly through progressive involution of the milk-producing acini. Fibrous tissue is also replaced by adipocytes. However, the extensive use of hormonal replacement therapy has considerably changed the appearance of this postmenopausal breast tissue.

Biobanking of normal breast tissue for research

Tissue-based studies of breast carcinogenesis utilize breast cancer tissue and different types of non-cancerous breast tissue, sometimes called normal breast tissue, as control for comparison. Most commonly used non-cancerous breast tissue is derived from reduction mammoplasty either from breast cancer patients, of unaffected breast for symmetry in breast cancer patients, or from healthy women operated for cosmetic purposes. Other sources of non-malignant breast tissue used in research include tissue from prophylactic mastectomy, neighboring breast tissue from women with benign breast lesions, excess tissues with benign histological appearance collected from surgical procedures, or unaffected ipsilateral or contralateral breast tissue from patients with breast cancer.

Usually there is a medical reason to surgically remove tissue—for example in prophylactic mastectomy for high risk of breast cancer due to gene mutations, or removal of benign lesions due to pathological features. Therefore, this type of tissue is not suitable for use as “normal” tissue. Breast tissue collected by reduction

mammoplasty, selected on the basis of convenience, may be the best representative of normal tissue. It is plentiful and removed for cosmetic reasons, not because of clinical abnormalities or high-risk profiles. However, none of these tissues have been found suitable as a substitute for truly normal breast tissue in studies of breast cancer carcinogenesis (Ambaye et al. 2009, Graham et al 2010, Degnim et al. 2012, Tadler et al. 2014, Acevedo et al. 2019).

Today there are several breast cancer tissue biobanks around Europe, North and South America, Asia and Australia, but to our knowledge the only biobank that collects truly normal breast tissue is the Susan G Komen for the Cure Tissue Bank (KTB) at Indiana University Simon Comprehensive Cancer Center in the USA (Sherman et al. 2012). There, tissue has been collected from volunteers of all ethnicities aged 18 and upward. Several articles have been published using this material. Radovic et al. 2014 concluded that breast tissue from healthy volunteers acts as a superior normal breast tissue control. The same source of tissue has been used in Pardo et al. 2014, where the author analyzed the transcriptome of normal, healthy, pre-menopausal breast tissue using next-generation sequencing.

In order to move breast cancer research forward, there is a need for well-annotated collections of breast tissue from healthy women (Thompson et al. 2008, Eccles et al. 2013). Adequate control tissue will help shed light on pre-clinical molecular events, and provide the correct basis for comparison in case-control studies. The overall goal of this study was to establish a biobank of normal breast tissue biopsies. The biobank was established for the purpose of describing baseline gene expression patterns in normal breast tissue of postmenopausal women. We will also explore the variation of gene expression in normal breast tissue following exposure to known breast cancer risk factors (smoking, alcohol consumption, HT use, obesity and parity), and finally, we will use the normal breast tissue in future case-control studies.

METHODS

The normal breast tissue biopsy study, part of the NOWAC Postgenome cohort

This study is part of the Norwegian Women and Cancer (NOWAC) Postgenome cohort. NOWAC is a national, prospective study started in 1991, where breast cancer is the most important endpoint (Lund et al. 2008). The study included 150 000 women born 1943–1957, who to date have answered between one and three questionnaires. During the period 2003–2006 we built a unique biobank by collecting

blood samples, buffered to protect the mRNA gene expression profile, from 50 000 NOWAC participants. These samples constitute the major part of the NOWAC Postgenome cohort. Furthermore, starting in 2006 and in collaboration with 11 Norwegian hospitals, we collected buffered blood samples and tissue samples from 400 women with breast cancer tumors at the time of diagnosis. These women were also participants in NOWAC, they were born between 1943–1957, and were diagnosed with breast cancer during the period 2006–2011. Until that time, there was no suitable tissue material available that expressed the normal pattern of variation in gene expression in the relevant age group. To address that need, during the period 2010–2012 we collected breast tissue and buffered blood samples from 368 healthy women. Volunteers for this part of the study were recruited from the NOWAC cohort through the national mammography screening program, which they were participating in at the time.

Recruitment of study participants

Recruitment to the study and the tissue collection took place at the Breast Diagnostic Center at the University Hospital of Northern Norway (UNN), Tromsø, Norway. Inclusion criteria were as follows: enrolled in the NOWAC cohort, born between 1943 and 1957, and consent given. The radiographer (not affiliated with the NOWAC study) asked women, when presenting at the mammography screening unit, if they would consider participating in this study. If answering positively, the candidate would meet after the screening procedure for written and oral information and to get answers to any questions they may have had. The women who agreed to participate were asked to sign a written, informed consent form. All participants completed a two-page questionnaire regarding menopausal status, weight and height, exposure to smoking and alcohol consumption, use of HT and other medication. Exclusion criteria included previous history of breast cancer, positive mammogram, other relevant malignant diseases, and use of anticoagulation therapy with Coumadin (Marevan), Heparin, Persantine, or Plavix. Use of acetylsalicylic acid was not an exclusion criterion.

Procedures for tissue and blood sampling

Core biopsies of normal breast tissue were obtained immediately after mammography, from the gland tissue of the upper lateral quadrant of the left breast. The tissue biopsy was taken with the women in declined position on the examination bed. The skin was disinfected with chlorhexidine solution in alcohol prior to incision.

Intradermal local anesthesia was applied using 2 ccl of 1% Lidocaine. A 3 mm skin incision was performed with a scalpel. With ultrasound guidance, a cylinder biopsy was taken with a needle size 14 gauge in a biopsy pistol, by an experienced radiologist. Compression bandage was placed at the biopsy site, which was to be kept in place until the next day. No further activity restriction was advised. During the study, no systematic follow-up has been undertaken. The biopsy was immediately placed in RNAlater for RNA stabilization (Qiagen, Hilden, Germany), and kept at room temperature for <24 hours until storage in a freezer at –70˚C.

Two vials of blood were taken by standard venipuncture (phlebotomy) with hypodermic butterfly needle on a closed system to the vacuum test tubes. One of the blood samples was taken using the PAXgene Blood RNA collection system (Preanalytix/Qiagen, Hombrechtikon, Switzerland), which contains a buffer for stabilizing the mRNA gene expression profile during long-term storage. The other blood sample was mixed with standard citrate solution. Blood samples were kept at –70˚C until further use. The blood sampling was performed before the tissue sampling.

RESULTS

We collected 368 biopsies of normal breast tissue from postmenopausal women. The rate of inclusion of all women invited to participate was 64%. A linkage to the Norwegian Cancer Registry 3 years after the sampling period ended resulted in five biopsies being excluded due to breast cancer diagnosis within 3 years after the biopsy was taken, and one due to a prior lymphoma diagnosis with unknown treatment. We used 16 biopsies for testing of different laboratory methods. A total of 311 biopsies were included for further analysis, which matched the number of cancer biopsies in our biobank collected for a comparative study.

All participants were advised to contact a physician in case of any suspicion of adverse reaction or complication such as hematoma, infection, or pain. No case of allergic reaction to the local anesthesia was registered. One participant directly reported a hematoma at the biopsy site. She was examined by a surgeon, who found a 3 cm hematoma, but no treatment or follow-up was considered necessary.

Characteristics of women included in this study

Characteristics of the 311 women included in the final study sample is summarized in Table 4.1. All participants were post-menopausal, and the average age was 60 years. The population, as a whole, were slightly overweight after WHO standard, with average BMI 26,2. Most of the women had given birth (have completed

full term pregnancy), and the average number of children was 1,9. The highest number of children was 8. A majority of the women (79%) had consumed alcohol during the week before sampling, and 21 % had been smoking during the week prior to biopsy sampling. Very few participants (8,4 %) used HT for menopausal symptoms. The majority of participants (70 %) used different types of medication in the week prior to blood sampling, either alone or in combination. The most frequent types were blood pressure medication, anti-cholesterol drugs, and synthetic thyroid hormone, followed by ASA (aspirin) and NSAIDs.

Table 4.1. Characteristics of the study population (n=311)

Characteristics	Mean/Frequency	Missing
Age, mean (SD)	60,1 (3,9)	0
BMI, mean (SD)	26,2 (4,5)	4
Parity (n, %)		0
Yes	256 (82,3)	
No	55 (17,7)	
N children (mean, SD)	1,9 (1,2)	0
Smoking (n, %)		0
Yes	66 (21,2)	
No	245 (78,8)	
HT use (n, %)		1
Yes	26 (8,4)	
No	284 (91,6)	
Alcohol (n, %)		6
Yes	241 (79)	
No	64 (21)	
Medication use (n, %)		
Any medication	216 (70,8)	6
Blood pressure alone or in comb. with antiarrhythmic	56 (18,4)	
Anti-cholesterol	36 (11,8	
Levaxin (synthetic thyroid medications)	30 (9,8)	
Asthma/allergy	23 (7,5)	
NSAIDs alone or in combination with Paracetamol	22 (7,2)	
Albyl (acetylsalicylic acid)	19 (6,2)	
Other	30 (9,8)	

Abbreviations: BMI, body mass index; HT, hormone therapy; NSAID, non-steroidal anti-inflammatory drugs; SD, standard deviation.

DISCUSSION

Above we have described the process of establishing a biobank of normal breast tissue biopsies from 311 postmenopausal women. In the following we discuss practical aspects of establishing the biobank, as well as ethical considerations, and highlight some factors that enabled the successful establishment of the NOWAC normal breast tissue biobank.

Where to find volunteers and how to recruit them?

The process of recruiting healthy volunteers for an invasive procedure may, if not planned properly, render the final study sample heavily affected by selection bias, subsequently reducing the generalizability of any findings. To reduce selection bias, our starting point was the nationally representative NOWAC study, as well as the national mammography screening program. The screening program invites all Norwegian women aged 50–69 years to mammography every other year, free of charge. Hence, an important success factor for this study was the use of the local screening facility, which enabled us to contact all eligible women in the region.

Prior to our work, the same facility had completed two small surveys (unpublished) to start the process of assessing the feasibility of collecting tissue biopsies from healthy women. The first was conducted to register discomfort and possible complications associated with the biopsy procedure and was based on interviews with 100 women who had undergone this procedure. The women were asked about pain, bleeding, hematoma, and infections. The result was consistent with the impression from the clinical work that biopsy taking is virtually painless and there is a very low risk for complications associated with the procedure. The second survey aimed to determine whether it would be possible to collect breast tissue biopsies from healthy women. We asked 81 women who participated in the mammography screening program if, hypothetically, they would be willing to have a breast biopsy taken to be used for research purposes. After receiving written and oral information, 12% answered no, 14% needed more information, and 74% answered yes. These results gave important cues on feasibility.

Collaboration with clinicians

The local mammography screening facility handles about 40 invitations every day. The NOWAC study has been collaborating with the facility since March 2002, when approximately 2 000 blood samples were collected for a different NOWAC project. The facility also played an active role in recruiting partners for a cancer

biopsy study at eleven of the country's hospitals. This close and long-standing collaboration is another important success factor for the present project. The screening facility already had valuable experience in contributing to research during their clinical everyday setting. Though the environment was familiar with research, it was necessary to make a detailed plan and spend time to figure out the most feasible way to complete all the steps with the clinical personnel involved. This included having the same person involved every day, who was familiar with the hospital environment and the department's work, as well as being involved in the research project.

The biopsy procedure involved is virtually painless, with a very low complication rate, and was performed by an experienced radiologist within the well-established framework of the screening facility, minimizing the risk of unforeseen incidents. All women were given information on actions to be taken in the case of complications. Since the procedures took place in the hospital setting, any complication or injury would be reported as a patient injury according to established national guidelines. Women were encouraged to contact the screening facility if a suspicion of a complication should arise after leaving the department. Complications requiring immediate treatment outside opening hours would be attended by the staff in the emergency room. These actions were largely comparable to actions to be taken in case of complications after any breast tissue biopsy procedure, and put no extra burden on the clinical staff.

Ethical aspects

In accordance with legal requirements for research on human biological material and personal data (The Health Research Act, Chapters 3-7), the Regional Committee for Medical and Health Research Ethics of Northern Norway (REC North) approved the protocol for the present study, and the Data Protection Authority granted a license for the use of health-related data. However, the project was planned some years ago, before the European Union issued the new General Data Protection Regulation (GDPR) in 2018. In Norway, GDPR was implemented at the national level through a new Personal Data Act, also in 2018. The risk of misuse of personal information, or the risk of loss of control of the personal information, is present in the current project, but this risk is by no means greater here than in comparable projects. These aforementioned risks are the focus of GDPR, and after its implementation, data-handling procedures have also been improved for the NOWAC project.

The need for close regulation of biomedical research dates back to atrocities during the Second World War, which led to the emphasis on human rights in the

Nuremberg Code of 1947. A main point in the Code stated that participation in research must be voluntary. Furthermore, the World Medical Association's Declaration of Helsinki (1964) focused on obligations of the researchers and the research institutions, and stressed the concept of informed consent (Fisher 2006). That the consent must be voluntary or free means that the individual included in the research shall not decide his/her position through a process characterized by coercion or pressure. Likewise, situations that do not include direct coercion can mean an unacceptable weakening of the consent that was given. Our participants were already part of the NOWAC study when they were invited for the biopsy study. Potentially, this could contribute to a feeling of pressure to participate in the biopsy study. We, the researchers, regarded this project as a continuation of the ongoing NOWAC study, and this backdrop may have put an indirect pressure on the women at the point of invitation. Still, the option to decline participation was always clearly communicated, both orally and in writing, hence we conclude that the principle of voluntary participation was never challenged.

The principle of informed consent entails that the individual being subjected to research must be aware of the study's methodology/procedures, purposes, and the type of results expected. The information given to participants must include a description of any expected inconvenience, discomfort, or risk that may be inflicted. This principle may be regarded as particularly important when performing an invasive procedure on healthy volunteers who would not otherwise undergo such procedures. Further, as the material collected in our study will be used for genomic profiling (mRNA gene expression analysis and potentially DNA profiling), care must be taken to ensure that participants understand the information that was given. The participants may have different experiences and assumptions when they internalize and interpret the information. We did not undertake any evaluation of the participants' understanding of the scientific content of the project, but each woman spoke personally to our radiologist, with ample opportunity to ask questions. Legislation on this topic focuses only on groups of people that may be noncompetent to give consent (e.g. persons under the age of 18, or for medical reasons). Hence, some questions may be ethically interesting, but will not have any practical consequences for our project. As examples, one might ask if it would be ethically acceptable to include participants if we discovered that they had not understood the information correctly. In addition, what about individuals who did not want to read the information that was given, but nonetheless wished to participate in the project?

One of our pre-study surveys assessed the healthy women's willingness to donate a breast tissue biopsy. The majority (74%) were willing to donate, and many women expressed a high degree of motivation to continue contributing to research

on breast cancer. Contributing biological sample material to research may be viewed in different ways. The biopsy may be a viewed as a gift or a donation, with no expectation of receiving anything in return. It may also be viewed as a transaction. It that case, the regional ethical committee would act as the real estate agent, looking out for the donors rights, and the consent form may be regarded as the contract between the two parties in the transaction. Viewed as a transaction, there is an expectation of receiving something in return, in this case somewhat distant "payments" such as knowledge of breast cancer, and better treatment. Another option for how to view the act of contributing a biopsy would be as an act of reciprocity. Modern-day medicine is an empirical science which has been built on the knowledge generated from the general population and from patients. Patients today expect to receive the latest treatments that are developed on the basis of this knowledge, and as such, they are morally obliged to contribute to that same knowledge base. In this normative ethics setting, the consent may be viewed as an expression of gratitude toward previous sample donors, of acknowledgment of the moral obligation to contribute, of the will to contribute, and of trust in that the donated material will be used as intended.

We do not have information on each woman's motivation to contribute to the study, but some external factors may also be at play. The city of Tromsø is small, with only 72 000 inhabitants. The city's one university is young and was founded in 1968 during a period of strong growth for the city, and, naturally, its foundation contributed to this growth. Today, the university is one of the city's two largest workplaces, along with the university hospital. These aspects contribute to the fact that the university is a strong part of the city's identity and the inhabitants are well known for contributing to research (Jacobsen et al. 2012). Hence, the feeling of reciprocity, grounded in normative ethics, and supported by favorable local conditions, may be important aspects for the high participation rates in the present study.

There is an ongoing debate on whether researchers should be obliged to return information on health-related aspects to research participants (Klingstrom et al. 2018). However, the present study and its analytical methodology is purely explorative in nature. No clinical relevance of potential findings based on our chosen analytical methods has been established (low clinical validity), and any findings would be non-actionable (i.e. the participant or clinicians could not take action to improve the risk or progression of a potential disease) (Klingstrom et al. 2018). Based on the limited clinical relevance of any findings in this project, any results were unlikely to affect the patient's need for further information, or for their consent. Hence, in this project giving feedback to participants was not considered as relevant, and this was stated in the information given to participants.

Strength and weaknesses

Firstly, the women were recruited from the mammography screening program, not referred from a physician due to symptoms or suspicion of breast pathology. Their biopsies are therefore representative of truly normal breast tissue, and the women have the same risk of developing breast cancer as any other women in the same age group. Since all women were NOWAC participants, extensive information on exposures in the past can be retrieved from questionnaires answered prior to the initiation of the biopsy study. Further strengths of the study include the high inclusion rate (64%) and the high number (368) of biopsies sampled via a standard procedure, which ensures low technical variability. The blood samples were taken at the same time as the biopsies, enabling a valid comparison of gene expression profiles in two different tissues.

One weakness of the study pertains to the risk of selection bias. Our participants were recruited at the mammography screening facility in Tromsø, hence, at one single location. As a consequence, there is the possibility of geographical differences compared to the average Norwegian population regarding the gene expression in relation to different types of exposures. It should be mentioned that the blood and tissue samples were collected by random and continuous invitation during the whole 2-year period, so we expect minimal influence of seasonal biorhythms.

Due to heterogeneity of breast tissue, one single biopsy is not representative of the entire breast. Studies have shown intra-individual variability in composition of breast biopsies, and its impact on gene expression (Chollet-Hinton et al. 2018). This fact has important implications for studies based on normal breast tissue, including our own study. Since our inclusion rate was high and the complication rate turned out to be almost nil, we could have chosen to sample several biopsies from different areas of the same breast via the same skin incision. This can be considered for future trials, taking the varying biopsy composition into account. On the same note, our biopsies are whole tissue biopsies containing multiple cell types which may confound gene expression results. The biopsies were not histologically controlled/evaluated, so we do not have information on the ratio between different cell types. The biopsies were taken from the upper lateral area of the breast, known for a higher density of glandular tissue, in order to reduce the amount of adipocytes and increase mRNA output amounts. However, the biopsies were collected from postmenopausal women. The quantity of glandular tissue decreases with age, and our biopsies likely contain a higher proportion of fat and less glandular tissue compared to samples taken from younger women.

CONCLUSION

The work presented shows that establishing a collection of normal breast tissue samples is feasible and doable. Enabling factors for the present study included largely unbiased access to eligible participants, and close collaboration with clinicians during all steps of the sampling procedures. Furthermore, the source population of the present study has a high degree of health literacy and willingness to participate in research, which contributes to a high participation rate. The NOWAC normal breast tissue biobank for gene expression analysis will provide much-needed information on pre-clinical molecular events and a correct basis for comparison in case-control studies.

REFERENCES

Acevedo F, Armengol VD, Deng Z, Tang R, Coopey SB, Braun D et al. Pathologic findings in reduction mammoplasty specimens: a surrogate for the population prevalence of breast cancer and high-risk lesions. Breast Cancer Res Treat. 2019 Jan; 173(1): 201–207. Available from: https://link.springer.com/article/10.1007%2Fs10549-018-4962-0

Ambaye AB, MacLennan SE, Goodwin AJ, Suppan T, Naud S, Weaver DL. Carcinoma and atypical hyperplasia in reduction mammaplasty: increased sampling leads to increased detection. A prospective study. Plast Reconstr Surg. 2009 Nov; 124(5): 1386–1392. Available from: https://insights.ovid.com/pubmed?pmid=20009822

Apostolou P, Fostira F. Hereditary breast cancer: the era of new susceptibility genes. Biomed Res Int. 2013; 2013: 747318. Available from: https://www.hindawi.com/journals/bmri/2013/747318/

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018 Nov; 68(6): 394–424. Available from: https://onlinelibrary.wiley.com/doi/full/10.3322/caac.21492

Cancer Registry of Norway. Cancer in Norway 2017 – Cancer incidence, mortality, survival and prevalence in Norway. Oslo: Cancer Registry of Norway, 2018

Chollet-Hinton L, Puvanesarajah S, Sandhu R, Kirk EL, Midkiff BR, Ghosh K et al. Stroma modifies relationships between risk factor exposure and age-related epithelial involution in benign breast. Mod Pathol. 2018 Jul; 31(7): 1085–1096. Available from: https://www.nature.com/articles/s41379-018-0033-7

Degnim AC, Visscher DW, Hoskin TL, Frost MH, Vierkant RA, Vachon CM et al. Histologic Findings in Normal Breast Tissues: Comparison to Reduction Mammaplasty and Benign Breast Disease Tissues. Breast Cancer Res Treat. 2012 May; 133(1): 169–177. Available from: https://link.springer.com/article/10.1007%2Fs10549-011-1746-1

Eccles SA, Aboagye EO, Ali S, Anderson AS, Armes J, Berditchevski F et al. Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. Breast Cancer Res. 2013 Oct 1; 15(5): R92. Available from: https://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr3493

Fischer BA 4th. A summary of important documents in the field of research ethics. Schizophr Bull. 2006 Jan; 32(1): 69–80. Available from: https://academic.oup.com/schizophreniabulletin/article/32/1/69/2888604

Graham K, de las Morenas A, Tripathi A, King C, Kavanah M, Mendez J et al. Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. Br J Cancer. 2010 Apr 13; 102(8): 1284–1293. Available from: https://www.nature.com/articles/6605576

Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: the Tromso Study. Int J Epidemiol. 2012 Aug; 41(4): 961–967. Available from: https://academic.oup.com/ije/article/41/4/961/683871

Kamińska M, Ciszewski T, Łopacka-Szatan K, Miotła P, Starosławska E. Breast cancer risk factors. Prz Menopauzalny. 2015 Sep; 14(3): 196–202. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4612558/

Klingstrom T, Bongcam-Rudloff E, Reichel J. Legal & ethical compliance when sharing biospecimen. Brief Funct Genomics. 2018 Jan 1; 17(1): 1–7. Available from: https://academic.oup.com/bfg/article/17/1/1/3782585

Lund E, Dumeaux V, Braaten T, Hjartåker A, Engeset D, Skeie G et al. Cohort profile: The Norwegian Women and Cancer Study – NOWAC – Kvinner og kreft. Int J Epidemiol. 2008 Feb; 37(1): 36–41. Available from: https://academic.oup.com/ije/article/37/1/36/763947

Pardo I, Lillemoe HA, Blosser RJ, Choi M, Sauder CAM, Doxey DK et al. Next-generation transcriptome sequencing of the premenopausal breast epithelium using specimens from a normal human breast tissue bank. Breast Cancer Res. 2014 Mar 17; 16(2): R26. Available from: https://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr3627

Radovich M, Clare SE, Atale R, Pardo I, Hancock BA, Solzak JP et al. Characterizing the heterogeneity of triple-negative breast cancers using microdissected normal ductal epithelium and RNA-sequencing. Breast Cancer Res Treat. 2014 Jan; 143(1): 57–68. Available from: https://link.springer.com/article/10.1007%2Fs10549-013-2780-y

Sherman ME, Figueroa JD, Henry JE, Clare SE, Rufenbarger C, Storniolo AM. The Susan G. Komen for the Cure Tissue Bank at the IU Simon Cancer Center: a unique resource for defining the “molecular histology” of the breast. Cancer Prev Res (Phila). 2012 Apr; 5(4): 528–535. Available from: https://cancerpreventionresearch.aacrjournals.org/content/5/4/528.long

Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY et al. Risk Factors and Preventions of Breast Cancer. Int J Biol Sci. 2017 Nov 1; 13(11): 1387–1397. Available from: https://www.ijbs.com/v13p1387.htm

Tadler M, Vlastos G, Pelte MF, Tille JC, Bouchardy C, Usel M et al. Breast lesions in reduction mammaplasty specimens: a histopathological pattern in 534 patients. Br J Cancer. 2014 Feb 4; 110(3): 788–791. Available from: https://www.nature.com/articles/bjc2013708

Thompson A, Brennan K, Cox A, Gee J, Harcourt D, Harris A et al. Evaluation of the current knowledge limitations in breast cancer research: a gap analysis. Breast Cancer Res. 2008; 10(2): R26. Available from: https://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr1983

Paper III

Krum-Hansen, S., Olsen, K.S., Anderssen, E., Frantzen, J.O., Lund, E. & Paulssen, R.H. (2023).

Associations of breast cancer related exposures and gene expression profiles in normal breast tissue – The Norwegian Women and Cancer (NOWAC) normal breast tissue study

Cancer Reports, 6(4), e1777.

Received: 3 February 2022 | Revised: 11 November 2022 | Accepted: 12 December 2022

DOI: 10.1002/cnr2.1777

ORIGINAL ARTICLE

Cancer Reports WILEY

Associations of breast cancer related exposures and gene expression profiles in normal breast tissue—The Norwegian Women and Cancer normal breast tissue study

Sanda Krum-Hansen[1,2] | Karina Standahl Olsen[1] | Endre Anderssen[3] | Jan Ole Frantzen[4] | Eiliv Lund[1] | Ruth H. Paulssen[3,5]

[1]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway

[2]Department of Hematology and Oncology, Stavanger University Hospital, Stavanger, Norway

[3]Genomics Support Center Tromsø (GSCT), UiT The Arctic University of Norway, Tromsø, Norway

[4]Narvik Hospital, University Hospital of North Norway, Narvik, Norway

[5]Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway

Correspondence
Sanda Krum-Hansen and Karina Standahl Olsen, Department of Community Medicine, UiT The Arctic University of Norway, N-9037 Tromsø, Norway.
Email: sanda.krum-hansen@uit.no and karina.s.olsen@uit.no

Funding information
Northern Norway Regional Health Authority (Helse Nord); Universitetet i Tromsø The Arctic University of Norway

Abstract

Background: Normal breast tissue is utilized in tissue-based studies of breast carcinogenesis. While gene expression in breast tumor tissue is well explored, our knowledge of transcriptomic signatures in normal breast tissue is still incomplete. The aim of this study was to investigate variability of gene expression in a large sample of normal breast tissue biopsies, according to breast cancer related exposures (obesity, smoking, alcohol, hormone therapy, and parity).

Methods: We analyzed gene expression profiles from 311 normal breast tissue biopsies from cancer-free, post-menopausal women, using Illumina bead chip arrays. Principal component analysis and *K*-means clustering was used for initial analysis of the dataset. The association of exposures and covariates with gene expression was determined using linear models for microarrays.

Results: Heterogeneity of the breast tissue and cell composition had the strongest influence on gene expression profiles. After adjusting for cell composition, obesity, smoking, and alcohol showed the highest numbers of associated genes and pathways, whereas hormone therapy and parity were associated with negligible gene expression differences.

Conclusion: Our results provide insight into associations between major exposures and gene expression profiles and provide an informative baseline for improved understanding of exposure-related molecular events in normal breast tissue of cancer-free, post-menopausal women.

KEYWORDS

alcohol consumption, breast cancer, breast tissue, gene expression, hormone therapy, microarray, normal tissue, obesity, parity, smoking

1 | BACKGROUND

Breast cancer is now the most frequently diagnosed cancer worldwide, with more than two million new cases per year.[1] The main risk

Sanda Krum-Hansen and Karina Standahl Olsen contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Cancer Reports* published by Wiley Periodicals LLC.

factors for breast cancer, other than sex and age, include overweight/obesity, alcohol consumption, family history of breast cancer, reproductive history, postmenopausal hormone therapy (HT), as well as smoking for pre-menopausal breast cancer.[2,3] With growing incidence rates, increased understanding of the mechanisms of cancer development is needed for preventive measures and early detection.

Despite growing understanding of breast cancer development at the molecular level, our knowledge of transcriptomic signatures of normal breast tissue is still incomplete. Tissue samples of normal breast have been widely used in breast cancer research serving as control tissue.[4–7] However, these tissue samples often originate from reduction mammoplasty, benign breast lesions, or histologically normal tissue adjacent to breast cancer.[4,5,8] Such tissue samples often show more histological abnormalities when compared to tissue obtained from healthy donors,[9–11] and using different sources of control breast tissue in different studies make comparisons between studies difficult. In addition, most studies on gene expression profiles were generated from a small set of samples that were likely not representative of the general population.[8,12] Finally, a better understanding of the natural variability of gene expression in normal breast tissue would represent a significant step forward in our understanding of early disease-related mechanisms.

With this study of a large, random sample of cancer-free, post-menopausal Norwegian women, we investigated the variability of gene expression in normal breast tissue. In particular, we explored gene expression patterns associated with exposure to known risk factors for breast cancer, such as obesity, parity, alcohol consumption, and use of menopausal HT. We also examined if smoking was associated with gene expression. The generated data represent a baseline of gene expression patterns in normal breast tissue from cancer-free, post-menopausal women, and can potentially play an important role in the feasibility, design, and analysis of future tissue-based studies investigating biomarkers of exposure, as well as breast cancer development.

2 | MATERIAL AND METHODS

2.1 | Study population

A detailed description of the recruitment proses and study population as well as ethical aspects of genetic research in healthy populations are presented and discussed in our previously published article.[13] Briefly, study participants were recruited through the national mammography screening program at the Breast Diagnostic Center at the University Hospital of North Norway (UNN), Tromsø, Norway, during the years 2010–2011. They were not referred due to pathological clinical findings or irregularities on previous mammograms but attended a scheduled routine mammogram. Eligible women were age between 53 and 67 years, were post-menopausal, and were already participating in the nationally representative Norwegian Women and Cancer study (NOWAC).[14] Exclusion criteria for the present study included self-reported previous history of breast cancer, positive mammogram, other current malignant diseases and use of anticoagulation therapy with warfarin, heparin, dipyridamole, or clopidogrel. Eligible women who agreed to participate received written and oral information, signed an informed consent form, and answered a two-page questionnaire regarding menopausal status, weight and height, smoking and alcohol consumption, use of HT and other medication. The number of included participants was 317. Three years after inclusion, data was linked to the Cancer Registry of Norway, using the unique personal identification number. This resulted in the exclusion of five participants who developed breast cancer within 3 years after the biopsy was taken, and one participant due to prior lymphoma diagnosis with unknown treatment. Thus, the final number of women included for statistical analysis was 311. The North Norway Regional Committee for Medical and Health Research Ethics (REK-Nord case no # 200603551) approved the study.

2.2 | Definition of exposures

Information on year of birth, menopausal status, current height and weight and exposures (HT use, smoking and alcohol consumption) was extracted from the two-page questionnaires answered at the time of inclusion.

Body mass index (BMI) was calculated, and obesity was defined according to the definition of the World Health Organization (WHO, BMI > 30). Women were considered post-menopausal if they reported that menstruation had ceased. In case of incomplete information, women were defined as post-menopausal if they were older than 53 years. Women who had consumed alcohol during the week prior to the biopsy, regardless of the type or amount, were defined as alcohol consumers. Similarly, women who had smoked during the week prior to biopsy were defined as smokers. Only women who were current users of systemic HT (tablets or patch) were defined as HT users. Data on parity was retrieved from the NOWAC database, and the variable was dichotomized into parous versus non-parous for analyses of gene expression.

We also collected data on smoking status and alcohol consumption (g/day) from the more comprehensive eight-page questionnaire answered by the participants as part of the prospective data collection in NOWAC. Participants of the biopsy study answered the eight-page questionnaire 0–20 years prior to donating a biopsy (1991–2011). These data were used for a sensitivity analysis.

2.3 | Tissue samples

An experienced radiologist obtained tissue samples after mammography, by ultrasound-guided needle-biopsy (14 gauges) from the gland tissue of the upper lateral quadrant of the left breast. The procedure was standardized. We collected one biopsy from every participant. In case of macroscopically sparse material, a second biopsy was obtained. The biopsies were kept at room temperature in RNA Later (Qiagen) for <24 h until storage at -70°C.

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.4 | RNA preparation

RNA preparation was conducted at The Department of Cancer Genetics, Oslo University Hospital, Oslo, Norway. Tissue biopsies were homogenized using TissueLyser LT (Qiagen, Hilden, Germany) and 5 mm steel beads. Total RNA and genomic DNA were isolated using the AllPrep DNA/RNA Mini Kit from Qiagen (Cat. No. 80204) and the QIAcube instrument (Qiagen, Hilden, Germany). A custom protocol was followed for the extraction (RNA_AllPrepDNARNA_AnimalCells_AllPrep350_ID2481). RNA was stored at −80°C. RNA quantity was measured by NanoDrop (Thermo Fisher Scientific, Wilmington, Delaware, USA). The BioAnalyzer 2100 and Agilent RNA 6000 Nano kit (cat. No. 5067–1511) were used to evaluate RNA integrity (Agilent, Santa Clara, US).

2.5 | Gene expression analysis

mRNA gene expression was analyzed at a certified Illumina service provider (NTNU Genomics Core Facility, Trondheim, Norway). Briefly, RNA was amplified with Ambion's Illumina® TotalPrep RNA amplification kit (Cat #AMIL 1791) using 400 ng of total RNA as input material. Incorporation of biotin-labeled nucleotides was performed overnight (14 h) at 37°C in vitro transcription (IVT). cRNA was quantified using the NanoDrop ND-1000 (NanoDrop, Wilmington, USA), and cRNA integrity was determined by electrophoresis using the Experion Bioanalyzer (BioRad). A total of 750 ng of biotin-labeled cRNA was hybridized to IlluminaHumanHT-12 v.4 expression bead chip (Illumina®). Beadchips were scanned with Illumina BeadArray Reader. Numerical results were extracted with Bead Studio v3.0.19.0 without any normalization or background subtraction.

2.6 | Statistical analysis

Data analysis was done using R (r-project.org). Raw files were quantile normalized using the Bioconductor lumi package.[15] Principal component analysis (PCA) and clustering was used for initial analysis of the dataset. The PCA was computed with all genes included. In order to obtain distinct clusters that correlate with the PCA scores to simplify interpretation we clustered the genes with the most variability (inter quantile range [IQR] > 1 log2 unit).

The three gene clusters identified were analyzed for overrepresented gene ontology (GO) terms using the clusterProfiler package.[16] This analysis highlighted cell composition as a potentially important co variate and non-negative matrix factorization (NMF) was used to obtain improved cellular composition estimates.[17] NMF was run with the "nsNMF" method and initialized with non-negative singular value decomposition to get a sparser estimate for the gene profiles of the cell types. The association of exposures and covariates on gene expression was determined using linear models for microarrays (LIMMA), with the scores on principal component one and two included as covariates to correct for bias due to the cellular composition of the biopsies. *p* Values from the linear models were corrected for multiple testing using the method of Benjamin and Hochberg.[18] Finally Camera[19] was used to identify pathways and GO terms that were related to the exposure variables. The Camera analysis was carried out with all genes using the same model as for the limma analysis, that is, PC1 and PC2 were included to correct for cell type composition.

STATA (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.) was used for descriptive statistics. *T* tests were used for BMI and age as continuous variables, and chi-square tests were used for the categorical variables.

3 | RESULTS

3.1 | Characteristics of study participants

For this study, 311 cancer-free, post-menopausal women were included. Average age of the study population was 60 years. More than half of the participants (54.7%) were classified as overweight according to WHO (BMI > 25), with an average BMI for the whole study populations of 26.2. More than 79% of the women had consumed alcohol, and 21% had been smoking prior to biopsy sampling. Very few participants used HT (8.7%), and most of the women had completed at least one full term pregnancy (Table 1).

3.2 | Gene expression in normal breast tissue

3.2.1 | Unsupervised clustering

After normalization of data, the initial analyses identified 607 genes with high level of variance with IQR larger than one log2 unit. These 607 genes were analyzed by *K*-means clustering and three dominating cluster were identified (Figure 1). These clusters appear unrelated to either of the exposures. PCA analysis with the exposure variables illustrated are shown in Figure S1.

Genes from the three clusters were analyzed using cluster profiler to identify GO categories that describe the functionality of the clusters (Figure 2).

Cluster 1 is dominated by processes related to epithelial cells, both differentiation processes, cell functions, and proliferation. Cluster 2 is dominated by genes involved in immune system processes and immune cell specific processes such as leucocyte chemotaxis, regulation of leucocytes proliferation, and interleukin production. Metabolic genes specifically related to lipid metabolism and fat cell differentiation dominate the third cluster.

PCA of the full dataset showed that 47% of the total variability of the gene expression data was captured by the two first principal components. Principal component one therefore reflects the balance between fatty tissue and epithelial tissue in the biopsy and principal component two reflects the fraction of immune cells included (Figure 3). NMF identified three factors that correlate with well-known cell type markers (Supporting Information S1). However, as the

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Descriptive characteristics of study population ($n = 311$)

		n (%)	Age, mean	BMI, mean	Current smokers, *n* (%)	Alcohol users, *n* (%)	HT users, *n* (%)	Parous, yes (%)	*n* Children, mean
Total		311	60.0	26.2	66 (21.2)	241 (79.0)	27 (8.7)	256 (82.3)	2.1
Obese	No (BMI < 30)	261 (83.9)	60.0	24.8	52 (19.9)	208 (80.9)	25 (9.6)	219 (83.9)	2.1
	Yes (BMI > 30)	50 (16.0)	60.7	33.4	14 (28)	33 (68.8)	2 (4.0)	37 (74.0)	2.2
	p, <30 vs. >30		.20		.20	.06	.27	.09	.60
Current smokers	No	245 (78.7)	60.3	26.1		195 (81.2)	25 (10.2)	201 (82.0)	2
	Yes	66 (21.2)	59.3	26.9		46 (70.7)	2 (3.0)	55 (83.3)	2.2
	p, no vs. yes		.07	.32		.07	.08	.80	.16
Alcohol users	No	64 (20.5)	61.2	27.4	19 (29.9)		5 (7.8)	53 (82.8)	2.2
	Yes	241 (77.5)	59.8	25.9	46 (19.0)		21 (8.7)	199 (82.6)	2.1
	p, no vs. yes		.01	.02	.06		.80	.96	.60
HT users	No	283 (90.9)	60.1	26.3	64 (22.6)	220 (78.8)		237 (83.7)	2.2
	Yes	27 (8.7)	59.6	25.7	2 (7.4)	21 (80.7)		18 (66.6)	1.6
	p, no vs. yes		.55	.33	.08	.80		.03	.04
Parous	No	55 (17.7)	60.4	26.7	11 (20.0)	42 (79.2)	9 (16.4)		
	Yes	256 (82.3)	60.0	26.11	55 (21.5)	199 (78.9)	18 (7.0)		2.3
	p, no vs. yes		.45	.34	.80	.90	.03		

Abbreviations: BMI, body mass index; HT, hormone therapy.

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

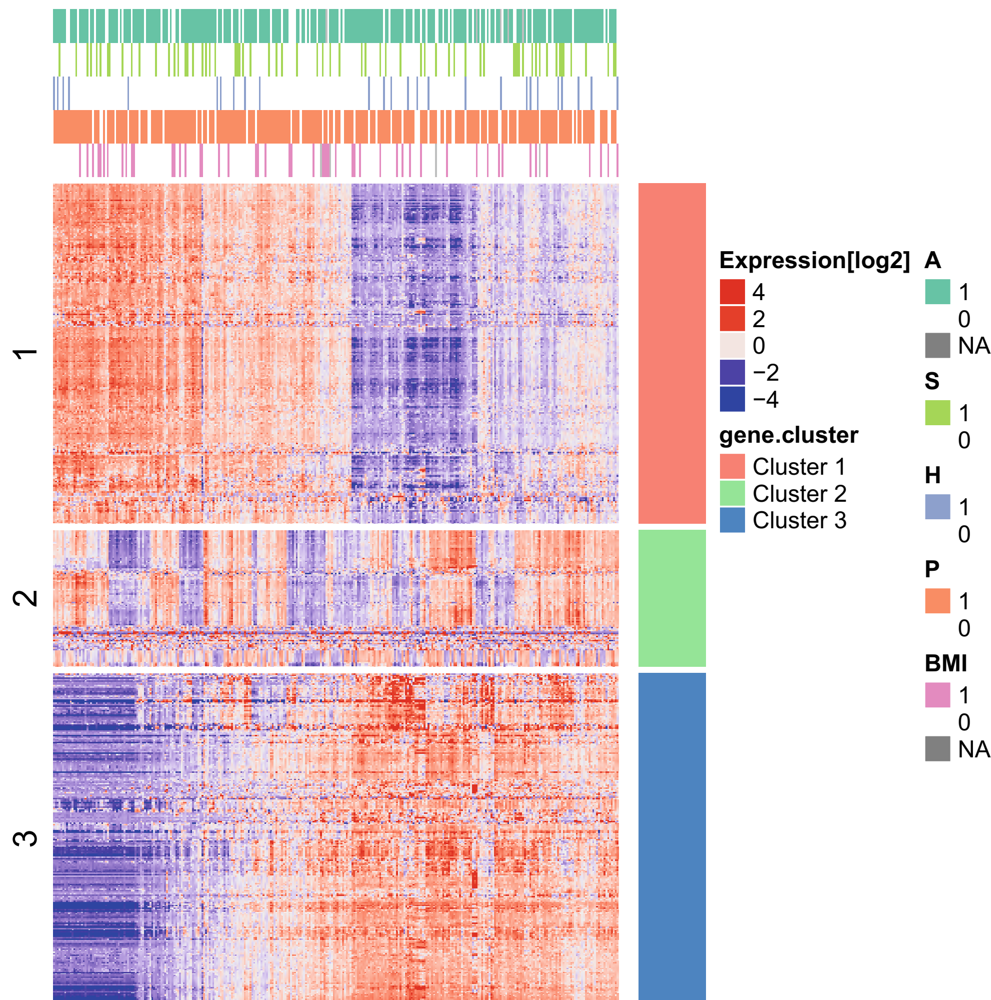


FIGURE 1 High-variance genes in breast tissue from healthy women represent three main clusters unrelated to major breast cancer risk factors. We analyzed 607 high-variance genes (inter quantile range larger than 1 log2 unit) by *K*-means clustering, and identified three dominating clusters (1-red, 2-green, and 3-blue). Distribution of the exposures are shown in the top pane (in color, legend to the right). A, alcohol; BMI, body mass index; H, HT use; P, parity; S, smoking

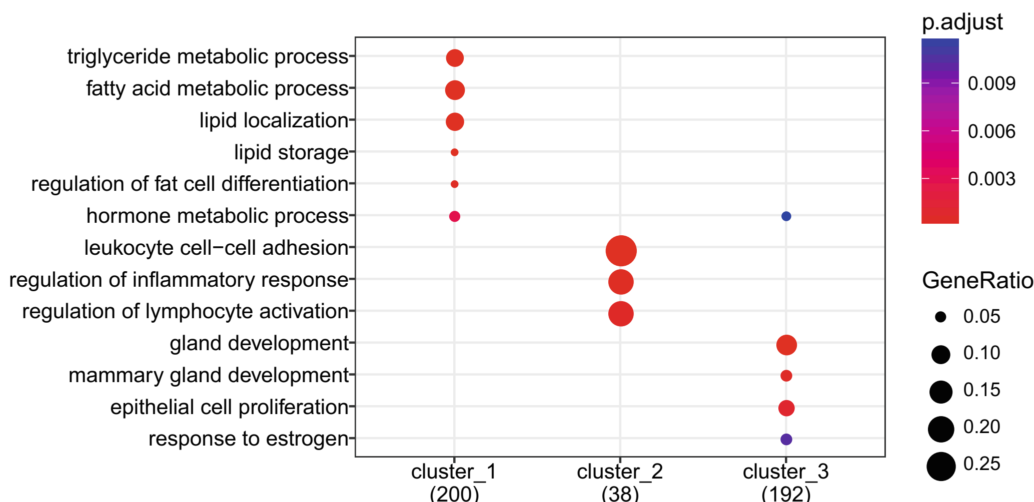


FIGURE 2 The three main gene expression clusters identified in breast tissue from healthy women likely reflect biopsy composition. Genes from the three clusters identified using *K*-means clustering (Figure 1) were analyzed using clusterProfiler, to identify overrepresented gene ontology categories that describe the functionality of the clusters

dominating cell types are epithelial and adipose cells with only a small fraction of immune cells, epithelial and adipose markers correlate strongly (negatively and positively respectively) with the first principal component. This enables the use of a single factor (PC1) to represent epithelial versus adipose tissue which reduces collinearity problems in the linear modeling. Hence, NMF was not included in further analyses.

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

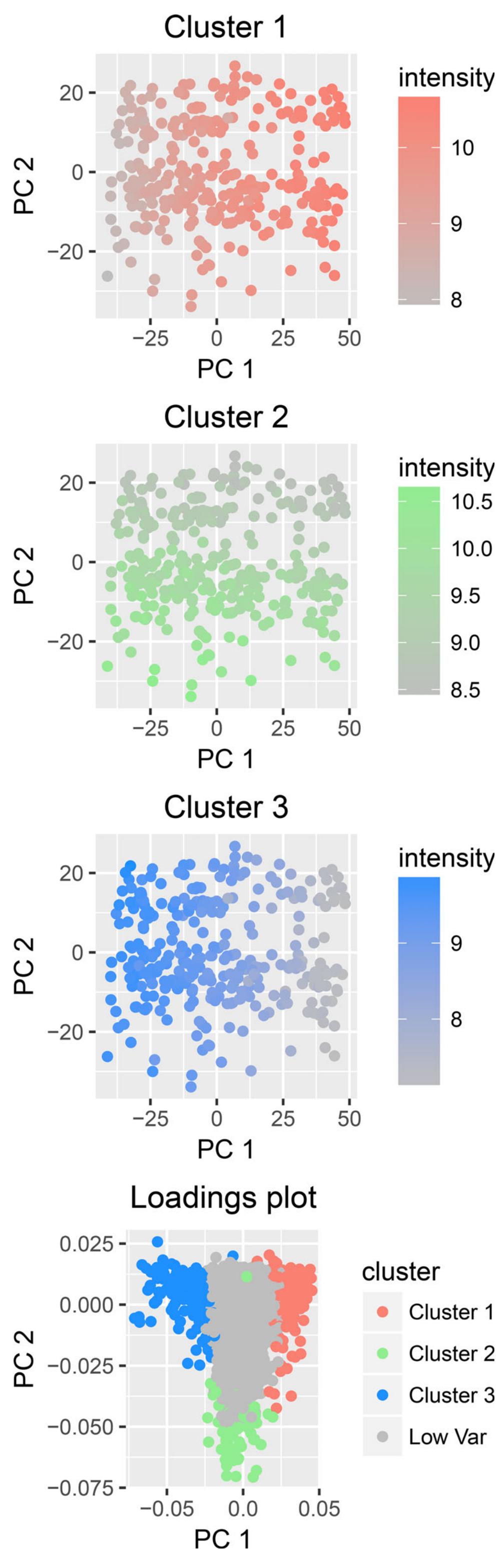


FIGURE 3 The total variability in gene expression data from breast tissue from healthy women is dominated by the balance of fatty tissue and epithelial tissue. Principal component analysis of the gene expression data was performed. Based on the overrepresented gene ontology terms of each cluster (Figure 2), the principal components of the loadings plot reflect biopsy composition represented by cluster 1 (fatty tissue), cluster 2 (immune cells), and cluster 3 (epithelial tissue)

3.2.2 | Associations between exposures and gene expression profiles

To analyze associations of selected exposures with gene expression profiles, we used LIMMA. An overview of the results is presented in Table 2. The 20 most significant genes and pathways associated with each exposure are presented in Tables 3 and 4 for obesity, Tables 5 and 6 for smoking, Tables 7 and 8 for alcohol, Tables 9 and 10 for HT, and Tables 11 and 12 for parity. The list of differentially expressed genes and gene sets are provided in Supporting Information S2 and S3.

When comparing gene expression profiles from breast tissue biopsies from women with BMI of 30 and above to those with BMI below 30, we identified 1577 significantly differentially expressed genes (Top 20 genes in Table 3). The differentially expressed genes included three alcohol dehydrogenases. There were more than 600 differentially expressed gene sets from GO and Kyoto Encyclopedia of Genes and Genomes, the majority of which were up-regulated in women with obesity (Top 20 pathways in Table 4). The up-regulated gene sets were dominated by immune-related processes, with both innate and adaptive immunology represented. The list of down-regulated gene sets included processes related to aerobic oxidation, fatty acid metabolism, amino acid metabolism, and protein translation in the mitochondria, which were all present among the top 20 gene sets, when sorted by *p* value.

Ten genes and 19 gene sets were statistically associated with smoking (Table 5). The genes CYP1B1, CYP1A1, F2RL3, CYTL1, TMEM178A, STAB1, NEURL1B, and EDC3 were significantly upregulated, whereas SPARC was downregulated. All the significant pathways were upregulated in smokers (Table 6).

Nine genes were statistically associated with alcohol exposure (Table 7). Eight of these were upregulated (MAMDC4, ISCA2, FAM171A2, BCDIN3D, SMIM20, RIT1, DHRS4-AS1, and UNC50) and one, EPB42, was downregulated. Pathway analysis revealed 80 alcohol-associated gene sets, and the 30 downregulated gene sets were all related to immunological processes (Top 20 pathways in Table 8). The 50 upregulated pathways were related to aerobic oxidation and fatty acid metabolism, and these were all among the top 20 gene sets when sorted by *p* value.

Two genes (ZCCHC12 and SEL1L2) were associated with HT use, both upregulated, but no pathways were identified (Tables 9 and 10). Finally, when comparing parous versus non-parous women, we found no associated genes or pathways at our chosen level of statistical significance ($p < .05$), Tables 11 and 12.

We carried out a sensitivity analysis combining two sources of exposure data for smoking and alcohol: the detailed, eight-page questionnaire answered 0–20 years prior to the biopsy, and the two-page questionnaire answered at the time of the biopsy. Being classified as a current smoker when combining data from the two time points was associated with the same top five genes as having smoked during the last week before the biopsy (data not shown). Being classified as a former smoker when combining data from the two timepoints was not associated with any differentially expressed genes (data not shown). Assessed in the eight-page questionnaire, the median amount of

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Overview of results: Genes and pathways associated with exposures

	Single genes (*n*)			Pathways (*n*)		
Exposure	Total	Upreg.	Downreg	Total	Upreg.	Downreg.
Obesity	1577	812	765	606	527	79
Smoking	10	9	1	19	19	0
Alcohol	9	8	1	80	50	30
HT	2	2	0	0	0	0
Parity	0	0	0	0	0	0

Abbreviations: downreg., downregulated; HT, hormone therapy; upreg., upregulated.

TABLE 3 Top 20 genes associated with obesity

Entrez ID	Gene symbol	Fold change	Adj. *p* value
124	ADH1A	−0.4839	<.001
125	ADH1B	−0.4241	<.001
137872	ADHFE1	−0.2960	<.001
122622	ADSSL1	−0.2956	<.001
283	ANG	−0.2377	<.001
398	ARHGDIG	0.1761	<.001
618	BCYRN1	0.5317	<.001
80763	SPX	−0.5217	<.001
78995	C17orf53	−0.1484	<.001
717	C2	0.2267	<.001
54976	C20orf27	−0.2612	<.001
719	C3AR1	0.1931	<.001
728	C5AR1	0.2035	<.001
85027	SMIM3	−0.1984	<.001
154791	C7orf55	−0.2249	<.001
56997	ADCK3	−0.3315	<.001
1230	CCR1	0.1737	<.001
8832	CD84	0.1881	<.001
1066	CES1	0.6174	<.001
1149	CIDEA	−0.6599	<.001

alcohol consumed was 3.08 g/day, and only four participants reported consuming more than 20 g/day. Information on alcohol intake in g/day collected 0–20 years prior to the biopsy was not associated with any differentially expressed genes.

4 | DISCUSSION

In this study, we explored the association of known risk factors for breast cancer with gene expression profiles in 311 biopsies from normal breast tissue. The number of associated genes and pathways were the highest for obesity, followed by smoking and alcohol. HT use versus non-use, and parity were associated with negligible differences in gene expression. The expression profiles of the biopsies where most likely influenced by the balance of cell types presented in the biopsy, as expected from this heterogeneous sample type.

4.1 | Obesity

BMI was the most influential risk factor in this gene expression study, with 1577 differentially expressed genes, and more than 600 differentially expressed gene sets. Overall, processes like immunology, estrogen metabolism and energy metabolism dominated the BMI-related results. In a wide perspective, our results reflect current hypotheses related to the causal association between obesity and breast cancer risk.

Our results indicate that immune-related processes are activated in the breast tissue of women with obesity. We identified five toll-like receptor (TLR)-related gene sets, 34 interleukin-related gene sets (including IL-1, -1β, -2, -6, -8, and -17), 12 interferon-related gene sets (IFN-α, -β, -γ), and three NF-κB gene sets. All of the mentioned gene sets were up-regulated. This finding is in line with established hypotheses on adipose tissue as a mediator for establishment of chronic inflammation, which is ultimately linked to increased risk of cancer.[20] During obesity, macrophages, putatively of the M1 type, accumulate in the adipose tissue, serving as a rich source of cytokines.[20,21] In obese breast tissue, inflammatory foci with dead adipocytes circled by macrophages, have been observed.[22,23] In the breast tissue, macrophages are exposed to saturated fatty acids from lipolysis leading to TLR 4 signaling via NFκ-b, culminating in increased expression of pro-inflammatory genes like COX-2, IL-1β, IL-6, and TNF-α.[20] These obesity-linked pro-inflammatory mediators may have local, pro-neoplastic effects, but also contribute to diminished overall health in obesity. However, our study cannot distinguish between breast tissue transcriptomic patterns associated with local inflammation, and transcriptomic patterns associated with systemic, circulating inflammatory factors. Biologically, however, this distinction is somewhat artificial, as the local and systemic effects of obesity are closely interrelated.[20]

Estrogen receptor signaling, increased by obesity and inflammation, is a key contributor to increased risk of hormone-positive breast cancer. In our data, five prostaglandin-related gene sets were identified. Prostaglandin contributes to increased aromatase expression in breast tissue, which is the rate-limiting enzyme in estrogen biosynthesis.[24,25] Hence, our results support the "obesity-inflammation-aromatase" axis[23] that leads to elevated estrogen levels in obese, postmenopausal women, resulting in increased breast cancer risk. The down-regulated gene sets in our results were dominated by processes such as aerobic oxidation and fatty acid metabolism. Imbalances in these energy metabolism pathways are closely linked to accumulation

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Top 20 gene sets associated with obesity

Pathway ID	Number of genes	Direction	Adj. *p* value	Category name
GO:0033108	73	Down	3.10E-13	Mitochondrial respiratory chain complex assembly
GO:0010257	56	Down	7.23E-12	NADH dehydrogenase complex assembly
GO:0032981	56	Down	7.23E-12	Mitochondrial respiratory chain complex I assembly
GO:0097031	56	Down	7.23E-12	Mitochondrial respiratory chain complex I biogenesis
GO:0022904	126	Down	1.40E-11	Respiratory electron transport chain
GO:0022900	128	Down	3.81E-11	Electron transport chain
GO:0045333	193	Down	1.12E-10	Cellular respiration
GO:0042775	62	Down	2.48E-09	Mitochondrial ATP synthesis coupled electron transport
GO:0042773	63	Down	3.81E-09	ATP synthesis coupled electron transport
GO:0006415	178	Down	5.15E-09	Translational termination
GO:0070124	84	Down	5.90E-09	Mitochondrial translational initiation
GO:0070125	84	Down	7.03E-09	Mitochondrial translational elongation
GO:0070126	86	Down	9.14E-09	Mitochondrial translational termination
hsa00190	119	Down	1.36E-08	Oxidative phosphorylation
GO:0006119	84	Down	1.80E-08	Oxidative phosphorylation
GO:0032543	117	Down	6.16E-08	Mitochondrial translation
GO:0006414	207	Down	6.45E-08	Translational elongation
hsa05012	117	Down	9.37E-08	Parkinson's disease
GO:0006103	20	Down	1.08E-07	2-oxoglutarate metabolic process
GO:0006120	41	Down	3.95E-07	Mitochondrial electron transport. NADH to ubiquinone

TABLE 5 Top 20 genes associated with smoking

Entrez ID	Gene symbol	Fold change	Adj. *p* value
1545	CYP1B1	0.4135	<.001
1543	CYP1A1	0.8881	<.001
9002	F2RL3	0.1422	<.001
54360	CYTL1	0.2314	5.00 E-04
130733	TMEM178A	0.2648	7.00 E-04
23166	STAB1	0.1677	.0147
54492	NEURL1B	0.1658	.0303
6678	SPARC	−0.2189	.0306
80153	EDC3	0.0662	.0306
3992	FADS1	0.3898	.0998
4689	NCF4	0.1332	.1226
84215	ZNF541	−0.1279	.1226
26011	TENM4	−0.1598	.2758
23743	BHMT2	−0.1358	.4033
10494	STK25	0.0674	.4876
10243	GPHN	0.0762	.5127
158800	RHOXF1	−0.0662	.5127
10003	NAALAD2	0.1332	.5164
5604	MAP2K1	−0.1012	.5164
3421	IDH3G	0.0645	.5164

of body fat leading to obesity.[26] In the breast, adipocytes are involved in normal tissue development, but there is also close interaction between stromal adipocytes and tumor cells.[27] Our results are in line with the finding of gene expression related to lipogenesis and fatty acid oxidation being downregulated in subcutaneous fat of both moderately and morbidly obese women, potentially as a mechanism for limiting further development of fat mass.[28,29] Of note, it has been suggested that this profile may be reversed by tumor cells, allowing adipocytes to provide lipids for the growing tumor.[27]

We also identified several down-regulated gene sets related to protein translation in the mitochondria. Metabolic imbalance is closely related to mitochondrial function, as they in addition to ATP production are involved in production and elimination of reactive oxygen species (ROS).[30] Obesity causes increased inflammation and oxidative stress through ROS production, which in turn may lead to mitochondrial dysfunction.[30] In adipose tissue and skeletal muscle, numbers of mitochondria and rate of mitochondria biogenesis may decrease during obesity.[30] Although these processes have not been described in breast tissue, our gene expression data from normal breast tissue supports this overall concept.

4.2 | Smoking

Our results revealed several genes and pathways significantly associated with smoking (Tables 5 and 6). Eight genes were up-regulated, one gene was down-regulated, and 19 gene sets were associated with smoking. Cytochrome P-450 1A1 and -1B1 (CYP1A1 and CYP1B1) were among the top up-regulated genes. These genes are involved in

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Top 20 gene sets associated with smoking

Pathway ID	Number of genes	Direction	Adj. *p* value	Category name
hsa00280	44	Up	<0.001	Valine. leucine and isoleucine degradation
GO:0009404	12	Up	<0.001	Toxin metabolic process
GO:0050665	12	Up	9.00 E-04	Hydrogen peroxide biosynthetic process
hsa00640	32	Up	0.0016	Propanoate metabolism
hsa00380	42	Up	0.0016	Tryptophan metabolism
GO:0009083	24	Up	0.0016	Branched-chain amino acid catabolic process
GO:0006084	32	Up	0.0022	Acetyl-CoA metabolic process
GO:0035383	91	Up	0.0043	Thioester metabolic process
GO:0009081	27	Up	0.0043	Branched-chain amino acid metabolic process
GO:0006637	91	Up	0.0043	Acyl-CoA metabolic process
GO:0006635	69	Up	0.0089	Fatty acid beta-oxidation
hsa00071	43	Up	0.0148	Fatty acid degradation
hsa00020	30	Up	0.0148	Citrate cycle (TCA cycle)
GO:0009062	86	Up	0.0148	Fatty acid catabolic process
GO:0046395	227	Up	0.0248	Carboxylic acid catabolic process
GO:0016054	227	Up	0.0248	Organic acid catabolic process
GO:0019395	94	Up	0.0274	Fatty acid oxidation
hsa05310	30	Up	0.0287	Asthma
GO:0034440	96	Up	0.0306	Lipid oxidation
GO:0006085	18	Up	0.0787	Acetyl-CoA biosynthetic process

TABLE 7 Top 20 genes associated with alcohol consumption

Entrez ID	Gene symbol	Fold change	Adj. *p* value
158056	MAMDC4	0.0664	.0043
122961	ISCA2	0.0775	.0043
284069	FAM171A2	0.0445	.0208
144233	BCDIN3D	0.0845	.0208
389203	SMIM20	0.0636	.0244
6016	RIT1	0.082	.0284
55449	DHRS4-AS1	0.1512	.0297
2038	EPB42	−0.1687	.0415
25972	UNC50	0.0911	.0428
128977	C22orf39	0.0621	.1285
554	AVPR2	0.0475	.1375
54496	PRMT7	0.0609	.1375
5162	PDHB	0.0842	.1542
5224	PGAM2	0.1471	.1565
51522	TMEM14C	0.0722	.1565
3654	IRAK1	−0.1114	.1746
23034	SAMD4A	−0.0528	.1746
283927	NUDT7	0.1107	.1792
128346	C1orf162	−0.2169	.1820
7345	UCHL1	−0.2757	.1898

metabolism of carcinogens including combustion products like polycyclic aromatic hydrocarbons found in cigarette smoke, but they are also involved in estrogen metabolism,[31] as well as breast cancer proliferation and survival.[32] Several aspects of CYP gene biology (expression levels, methylation levels, gene function) are related to smoking. Tsai et al.[33] investigated smoking-associated DNA methylation and gene expression variation in adipose tissue biopsies. Five of the identified genes in that study (AHRR, CYP1A1, CYP1B1, CYTL1, F2RL3) were both hypo-methylated and upregulated in current smokers. Four of those five genes were identified in our study (CYP1A1, CYP1B1, CYTL1, and F2RL3). Furthermore, CYP gene biology was associated with smoking in studies of breast cancer patient survival,[34] lung tissue,[35,36] prostate cancer cells,[37] and fetal placenta and livers.[38] Complementing these previous findings with our smoking-related gene expression data in normal breast tissue, gives important clues to the biological effects of smoking, that may contribute to increased breast cancer risk.

Similar to our findings on CYP1A1 and –B1, coagulation factor II receptor-like 3 (F2RL3, coding for the proteinase-activated receptor 4 protein) was up-regulated in our data, in accordance with previous reports. In an epigenome-wide study of DNA from pre-diagnostic blood samples, F2RL3 hypo-methylation strongly correlated with smoking.[39] Further, F2RL3 methylation was suggested as a biomarker of smoking,[40] and over-expression and hypo-methylation were associated with higher risk of lung cancer,[41] and with tumor aggressiveness and poor survival in renal cancer.[42]

In the smoking group, stabilin-1 (STAB1) was up-regulated, and secreted protein acidic and rich in cysteine (SPARC) was down-regulated. STAB1 is a scavenger receptor mediating both phagocytosis of

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

unwanted self-components, intracellular sorting, and endocytosis of extracellular ligands such as the extracellular matrix component SPARC.[43] STAB1 was up-regulated in smokers and COPD patients compared to non-smokers, although non-significantly.[44] STAB1 is expressed on tumor-associated macrophages in several cancers, and in human breast cancer STAB1 was found in stage I and IV disease, suggesting a role in early primary tumor growth and progression.[43] Studies have reported that SPARC induction inhibits breast cancer cell proliferation,[45] and down-regulated expression of SPARC correlated with poor breast cancer prognosis.[46] However, the role of SPARC may be highly dependent on context.[47]

Among the 19 pathways associated with smoking, the most prominent feature were energy metabolism and fatty acid metabolism pathways, including fatty acid degradation processes, acetyl-CoA metabolism, and the citrate cycle. Smoking has well-established effects on adipose tissue, termed smoking-induced dyslipidemia. In this state, lipolysis and free fatty acids are increased, involving hormone sensitive lipase and adipocyte differentiation.[48] Systemically administered nicotine induces lipolysis, in part by activating the classical adrenergic mechanism, and in part by directly activating a nicotinic cholinergic lipolytic receptor located in adipose tissue.[49]

4.3 | Alcohol

Exposure to alcohol was associated with 9 differentially expressed genes, and 80 gene sets (Tables 7 and 8). Overall, the magnitude of the differential gene expression is comparable to previous analyses of alcohol and gene expression in breast tumors.[50] Interestingly, the up-regulated BCDIN3 domain containing RNA methyltransferase gene (BCDIN3D) has been clearly linked to breast cancer progression, via down-regulation of tumor suppressor miRNAs.[51] In a cohort of 227 breast cancer patients, tumor levels of BCDIN3D was associated with lower disease-free survival.[52] Hence, BCDIN3D could serve as a link between alcohol consumption and breast cancer tumorigenesis and survival.

Among the 80 pathways associated with alcohol consumption, 30 were down-regulated. All of these were related to immunological processes, and the majority describe aspects of the innate immune system. Particularly, mast cell mediated immunity was present, including mast cell activation and degranulation. Mast cells have been linked to alcohol consumption, as they may mediate the damaging effects of alcohol by contributing to chronic inflammation, tissue damage, and remodeling, especially in the gastrointestinal tract.[53] Their role in cancer,[54] including breast cancer, is controversial, with conflicting results on the association with disease subtypes and prognosis.[55,56] In sum, these findings warrant further investigation of the effects of alcohol in the pre-cancerous breast tissue environment.

Fifty pathways were up-regulated in alcohol consumers (Table 8). Among those, two related processes were represented: aerobic oxidation, including translation of mitochondrial proteins for oxidative phosphorylation, and fatty acid metabolism. As in the liver, alcohol is metabolized in breast tissue into acetaldehyde, a class 1 carcinogen forming DNA and protein adducts, and further into acetic acid and

TABLE 8 Top 20 gene sets associated with alcohol consumption

Pathway ID	Number of genes	Direction	Adj. *p* value	Category name
GO:0070126	86	Up	<.001	Mitochondrial translational termination
GO:0032543	117	Up	<.001	Mitochondrial translation
GO:0070125	84	Up	<.001	Mitochondrial translational elongation
GO:0070124	84	Up	<.001	Mitochondrial translational initiation
GO:0022904	126	Up	<.001	Respiratory electron transport chain
GO:0022900	128	Up	<.001	Electron transport chain
GO:0033108	73	Up	1.00 E-04	Mitochondrial respiratory chain complex assembly
GO:0097031	56	Up	1.00 E-04	Mitochondrial respiratory chain complex I biogenesis
GO:0032981	56	Up	1.00 E-04	Mitochondrial respiratory chain complex I assembly
GO:0010257	56	Up	1.00 E-04	NADH dehydrogenase complex assembly
GO:0045333	193	Up	1.00 E-04	Cellular respiration
GO:0031163	18	Up	4.00 E-04	Metallo-sulfur cluster assembly
GO:0016226	18	Up	4.00 E-04	Iron–sulfur cluster assembly
GO:0046487	25	Up	4.00 E-04	Glyoxylate metabolic process
hsa00190	119	Up	5.00 E-04	Oxidative phosphorylation
GO:0006119	84	Up	7.00 E-04	Oxidative phosphorylation
GO:0015936	17	Up	.0011	Coenzyme A metabolic process
GO:0019395	94	Up	.0011	Fatty acid oxidation
GO:0006635	69	Up	.0011	Fatty acid beta-oxidation
hsa05016	177	Up	.0015	Huntington's disease

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

acetyl-CoA, the latter which enters the citric acid cycle.[57,58] This increased acetyl-CoA input may drive energy metabolism and increase the cellular energy state.[59] Furthermore, it has been suggested that acetyl-CoA is not merely a passive metabolite, but rather an important signaling molecule dictating cell function in a variety of settings.[58] Similarly, the various metabolites of the tricarboxylic acid cycle (TCA), increasing in concentrations upon an up-regulation of the cycle, affect intracellular and organismal processes, such as innate immunity, inflammation, and immune effector cells (succinate), and tumor cell growth (fumarate). TCA metabolite release from the mitochondria are one of the main processes by which the mitochondria influence cell function.[58] With the upregulation of aerobic oxidation pathways in our data, perhaps as a response to increased levels of acetyl-CoA from ethanol, it is also evident that the genotoxic acetaldehyde may be present in the breast tissue. Taken together, our data suggests that alcohol consumption may influence gene expression related to breast tissue physiology and metabolism.

4.4 | Hormone therapy

A current exposure to HT in our study was associated with two upregulated single genes, but no significant pathways (Tables 9 and 10). Prolonged, systemic use of all, yet especially combined HT is associated with increased risk for breast cancer.[60] Hall et al. found a distinct gene expression profile in breast cancer tissue associated with HT use, and linked HT use to better recurrence-free survival.[61] Changes in gene expression patterns in normal breast tissue after treatment with HT was observed in one experimental study,[62] and a recent study on DNA methylation showed association between HT use and epigenetic changes in normal breast tissue.[63] We did not observe similar changes in our data. Few participants exposed and lack of information on prior HT use, as well as duration of use, could partially explain these results.

TABLE 9 Top 20 genes associated with HT use

Entrez ID	Gene symbol	Fold change	Adj. *p* value
170261	ZCCHC12	0.2202	<.001
80343	SEL1L2	0.4523	5.00 E-04
400120	SERTM1	0.1426	.2945
9423	NTN1	0.1640	.4350
25884	CHRDL2	0.1356	.4903
140730	RIMS4	0.0701	.6069
6424	SFRP4	0.6362	.6069
8974	P4HA2	0.1608	.6493
26585	GREM1	0.3627	.7361
89876	MAATS1	−0.0950	.7730
220963	SLC16A9	0.1590	.7730
51081	MRPS7	0.0950	.7768
25878	MXRA5	0.2552	.7909
5387	PMS2P3	−0.0826	.7926
9201	DCLK1	0.2094	.7926
9540	TP53I3	0.1139	.7962
1136	CHRNA3	0.0695	.7962
4973	OLR1	−0.1745	.8054
6774	STAT3	0.1131	.8054
170261	ZCCHC12	0.2202	.8054

Abbreviation: HT, hormone therapy.

4.5 | Parity

In our study, parity was not associated with any significant single genes or pathways (Tables 11 and 12). Epidemiological studies have shown that a first full time pregnancy at an early age, as well as multiple pregnancies, are associated with long-term risk reduction for breast cancer.[64] Several studies found genomic signature of pregnancy in the breast tissue by comparing gene expression profiles of parous and non-parous postmenopausal women.[65,66] We did not reproduce these findings, perhaps related to a low number of non-parous women in our cohort (55 women, 17% of the study sample).

4.6 | Sensitivity analysis

As a sensitivity analysis, we combined two sources of exposure data for smoking and alcohol: the detailed, eight-page questionnaire answered 0–20 years prior to the biopsy, and the two-page questionnaire answered at the time of the biopsy. The combined information on smoking exposure provided no further insight. This was also true for participants being classified as former smokers. Similarly, combining information on alcohol intake during the week before the biopsy with the data on alcohol intake 0–20 years prior to the biopsy provided no further insight. The average alcohol consumption of our participants was low (median: 3.08 g/day), which limits our ability to discern effects of higher alcohol consumption.

For these sensitivity analyses of smoking and alcohol, the detailed exposure information was separated with up to 20 years in time from the more limited two-page questionnaire answered at the time of the biopsy. As the additional data did to add much, we chose to present the most recent exposure information as our main result, even though it was less detailed compared to the eight-page questionnaire. The lack of additional findings ties in with the understanding of gene expression being a highly dynamic and responsive biological process, which is likely to reflect recent exposures rather than exposure history. The results are in line with findings on gene expression in blood related to smoking history and current smoking.[67] In comparison, DNA methylation patterns may to a larger extent reflect previous exposure.[68]

5 | STRENGTH AND LIMITATIONS

The main strength of this study is the analysis of normal breast tissue samples from cancer-free women. With our choice of sample material,

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 10 Top 20 gene sets associated with HT use

Pathway ID	Number of genes	Direction	Adj. *p* value	Category name
GO:0006614	108	Up	.4887	SRP-dependent cotranslational protein targeting to membrane
GO:0006613	110	Up	.4887	Cotranslational protein targeting to membrane
GO:0045047	112	Up	.4887	Protein targeting to ER
GO:0045058	43	Down	.4887	T cell selection
GO:0050855	15	Down	.4887	Regulation of B cell receptor signaling pathway
GO:0006415	178	Up	.4887	Translational termination
GO:0006414	207	Up	.4887	Translational elongation
hsa00640	32	Up	.4887	Propanoate metabolism
hsa05012	117	Up	.4887	Parkinson's disease
hsa00280	44	Up	.4887	Valine. leucine and isoleucine degradation
hsa00190	119	Up	.4887	Oxidative phosphorylation
GO:0070125	84	Up	.4898	Mitochondrial translational elongation
GO:0001887	115	Up	.4977	Selenium compound metabolic process
GO:0050853	46	Down	.4977	B cell receptor signaling pathway
hsa05010	161	Up	.4977	Alzheimer's disease
GO:0042775	62	Up	.4977	Mitochondrial ATP synthesis coupled electron transport
GO:0097296	20	Down	.5193	Activation of cysteine-type endopeptidase activity involved in apoptotic signaling pathway
GO:0022904	126	Up	.5193	Respiratory electron transport chain
GO:0022900	128	Up	.5193	Electron transport chain
GO:0072599	116	Up	.5548	Establishment of protein localization to endoplasmic reticulum

Abbreviation: HT, hormone therapy.

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 11 Top 20 genes associated with parity, sorted by fold change

	Gene symbol	Fold change	Adj. *p* value
7018	TF	0.0704	.9997
4246	SCGB2A1	0.0676	.9997
26353	HSPB8	0.0534	.9997
8483	CILP	0.0534	.9997
1396	CRIP1	0.0471	.9997
5519	PPP2R1B	0.0471	.9997
51716	CES1P1	0.0428	.9997
4653	MYOC	0.0426	.9997
399888	FAM180B	0.0383	.9997
6289	SAA2	0.0325	.9997
5918	RARRES1	−0.0648	.9997
26585	GREM1	−0.0568	.9997
1545	CYP1B1	−0.0529	.9997
347733	TUBB2B	−0.0495	.9997
220	ALDH1A3	−0.0447	.9997
54360	CYTL1	−0.0351	.9997
972	CD74	−0.0309	.9997
29990	PILRB	−0.0305	.9997
5414	SEPTIN4	−0.0304	.9997
338328	GPIHBP1	−0.0303	.9997

Note: As the parity variable gave a uniform distribution of *p* values, we present the genes with the 10 highest and lowest fold changes.

we were able to describe the variability of gene expression according to established breast cancer risk factors in women with no clinically detected breast abnormalities. Our relatively large sample size also improves the generalizability of the findings, compared to many other studies. Further, placing the biopsies directly in an RNA stabilizing agent after surgical removal diminishes the risk of ex vivo expression changes and RNA degradation during storage. Information on current exposures is important for gene expression profiling, and these were collected at the time of the biopsy sampling.

Several limitations must be considered. We collected whole biopsies, and no histological assessment of the tissue composition was performed. In general, using self-report as the data collection method may introduce information bias, and unmeasured confounding factors may influence our findings. Twenty-one percent of our study population were defined as smokers. This number is comparable to smoking prevalence for adult females in Norway in 2010, at the time of our sample collection.[69] We did not assess smoking beyond current smoking status in our main analysis. Hence, a certain degree of misclassification is expected, for example, in categorizing former smokers as non-smokers, which may drive our results toward the null. However, our sensitivity analysis supports that smoking history does not have effect on gene expression in the breast tissue. Further, alcohol consumption is associated with smoking and is itself a known risk factor for BC. We adjusted for alcohol intake in the smoking analyses. Nonetheless, statistical adjustment using self-reported alcohol consumption may not be adequate to control fully for confounding by alcohol. Our alcohol-related analysis also has a few limitations. There was a high

TABLE 12 Top 20 gene sets associated with parity

Pathway ID	Number of genes	Direction	Adj. *p* value	Category name
GO:0070126	86	Up	.8550	Mitochondrial translational termination
GO:0070124	84	Up	.8550	Mitochondrial translational initiation
GO:0070125	84	Up	.8550	Mitochondrial translational elongation
hsa00280	44	Up	.8550	Valine. leucine and isoleucine degradation
hsa05144	51	Up	.9994	Malaria
GO:0030316	87	Up	.9994	Osteoclast differentiation
GO:2000482	19	Up	.9994	Regulation of interleukin-8 secretion
GO:1903034	403	Down	.9994	Regulation of response to wounding
GO:0002544	24	Up	.9994	Chronic inflammatory response
GO:0042554	23	Up	.9994	Superoxide anion generation
GO:0050777	120	Up	.9994	Negative regulation of immune response
GO:0032700	12	Up	.9994	Negative regulation of interleukin-17 production
GO:0046849	77	Up	.9994	Bone remodeling
GO:0002712	42	Down	.9994	Regulation of B cell mediated immunity
GO:0002889	41	Down	.9994	Regulation of immunoglobulin mediated immune response
GO:0048771	150	Down	.9994	Tissue remodeling
GO:0006959	180	Down	.9994	Humoral immune response
GO:2000641	16	Up	.9994	Regulation of early endosome to late endosome transport
GO:0045124	33	Up	.9994	Regulation of bone resorption
GO:0033005	17	Down	.9994	Positive regulation of mast cell activation

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

percentage of alcohol consumers in this study, but the proportion is comparable to the whole NOWAC study.[70] Additionally, we only have data on alcohol consumption during the previous week before biopsy in our main analysis. Potential effects of alcohol dose were addressed in the sensitivity analysis, although, these data were separated by up to several years from the biopsy. Finally, we did not include any analyses stratified by type and amount of alcohol, due to loss of power in such subgroup analyses.

The descriptive, cross-sectional design of this study provides a snapshot in time of gene expression profiles and does not allow any discussion of causality. By its nature, gene expression analysis is hypothesis generating. Testing the identified gene expression associations by using other study designs such as randomized controlled trials, or in an experimental, in vitro setting was beyond the scope of the present study.

6 | CONCLUSION

To our knowledge, this is the first study describing associations of breast cancer related exposures and gene expression profiles, in normal breast tissue from cancer-free, post-menopausal women. Obesity, smoking, and alcohol had the highest numbers of associated genes and pathways, whereas HT use and parity were associated with negligible gene expression differences in our data. Our results provide both confirmation of some previously reported findings, but also new hypotheses for further exploration. We conclude that our data provide an informative baseline for improved understanding of exposure-related molecular events in normal breast tissue.

AUTHOR CONTRIBUTIONS

Sanda Krum-Hansen: Conceptualization (equal); data curation (supporting); formal analysis (supporting); investigation (lead); methodology (equal); writing – original draft (equal); writing – review and editing (equal). **Karina Standahl Olsen:** Supervision (supporting); visualization (supporting); writing – original draft (equal); writing – review and editing (equal). **Endre Anderssen:** Data curation (lead); formal analysis (lead); methodology (supporting); software (lead); visualization (lead); writing – original draft (supporting); writing – review and editing (supporting). **Jan Ole Frantzen:** Conceptualization (supporting); funding acquisition (supporting); investigation (supporting). **Eiliv Lund:** Funding acquisition (lead); project administration (lead); resources (equal); supervision (supporting). **Ruth H. Paulssen:** Conceptualization (equal); funding acquisition (supporting); methodology (equal); project administration (supporting); resources (equal); supervision (lead); writing – original draft (supporting); writing – review and editing (supporting).

ACKNOWLEDGMENTS

We would like to thank all the women who participated and donated their time and breast tissue biopsy for this study. Thanks to all personnel working at the Breast Diagnostic Center at the University Hospital of North Norway (UNN), Tromsø, who made this collection possible. We thank Inger Riise Bergheim, Anita Halvei and Eldri Undlien Due at the Department of Cancer Genetics, Oslo University Hospital for isolating the RNA. The microarray service was provided by the Genomics Core Facility (GCF) at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. Bente Augdal and Marita Melhus has been responsible for the administration of the data collection and biobank. Finally, we are grateful to Dr Marko Lukic for data management services.

CONFLICT OF INTEREST

The authors have stated explicitly that there are no conflicts of interest in connection with this article.

DATA AVAILABILITY STATEMENT

Due to ethical restrictions on this dataset, which contains potentially sensitive information, the data will be made available upon request. Please contact the authors (karina.s.olsen@uit.no).

DISCLAIMER

The gene expression laboratory analyses were provided by the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and Central Norway Regional Health Authority.

ETHICS STATEMENT

The Regional Committee for Medical and Health research Ethics (case no # 200603551) has approved this study. Participants in this study signed an informed consent form.

ORCID

Sanda Krum-Hansen https://orcid.org/0000-0001-9090-035X

REFERENCES

1. WHO. Global Cancer Observatory 2022. https://gco.iarc.fr/.
2. WHO. Breast Cancer Fact Sheet 2022. https://www.who.int/news-room/fact-sheets/detail/breast-cancer.
3. WCRF. Diet and Cancer Report 2022. www.wcrf.org/dietandcancer/contents.
4. Schummer M, Green A, Beatty JD, et al. Comparison of breast cancer to healthy control tissue discovers novel markers with potential for prognosis and early detection. *PLoS One*. 2010;5(2):e9122.
5. Graham K, de Las M, Tripathi A, et al. Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *Br J Cancer*. 2010;102:1284-1293.
6. Pirone JR, D'Arcy M, Stewart DA, et al. Age-associated gene expression in normal breast tissue mirrors qualitative age-at-incidence patterns for breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2012;21(10):1735-1744.
7. Sun X, Shan Y, Li Q, et al. Intra-individual gene expression variability of histologically normal breast tissue. *Sci Rep*. 2018;8(1):9137.
8. Haakensen VD, Lingjaerde OC, Luders T, et al. Gene expression profiles of breast biopsies from healthy women identify a group with claudin-low features. *BMC Med Genomics*. 2011;4:77.
9. Degnim AC, Visscher DW, Hoskin TL, et al. Histologic findings in normal breast tissues: comparison to reduction mammaplasty and benign breast disease tissues. *Breast Cancer Res Treat*. 2012;133(1):169-177.

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10. Ambaye AB, MacLennan SE, Goodwin AJ, Suppan T, Naud S, Weaver DL. Carcinoma and atypical hyperplasia in reduction mammaplasty: increased sampling leads to increased detection. A prospective study. *Plast Reconstr Surg*. 2009;124(5):1386-1392.
11. Acevedo F, Armengol VD, Deng Z, et al. Pathologic findings in reduction mammoplasty specimens: a surrogate for the population prevalence of breast cancer and high-risk lesions. *Breast Cancer Res Treat*. 2019;173(1):201-207.
12. Bergholtz H, Lien TG, Ursin G, et al. A longitudinal study of the association between mammographic density and gene expression in normal breast tissue. *J Mammary Gland Biol Neoplasia*. 2019;24(2): 163-175.
13. Olsen KS, Krum-Hansen S. 4. Practical and ethical issues in establishing a collection of normal breast tissue biopsies—part of the NOWAC post-genome cohort. In: Lund E, ed. *Advancing Systems Epidemiology in Cancer*. Scandinavian University Press; 2020:63-76.
14. Lund E, Dumeaux V, Braaten T, et al. Cohort profile: the Norwegian Women and Cancer study—NOWAC—Kvinner og kreft. *Int J Epidemiol*. 2008;37:36-41.
15. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing illumina microarray. *Bioinformatics*. 2008;24(13):1547-1548.
16. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS J Integr Biol*. 2012;16(5):284-287.
17. Gaujoux R, Seoighe C. Semi-supervised nonnegative matrix factorization for gene expression deconvolution: a case study. *Infect Genet Evol*. 2012;12(5):913-921.
18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 1995;57(1):289-300.
19. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res*. 2012;40(17):e133.
20. Howe LR, Subbaramaiah K, Hudis CA, Dannenberg AJ. Molecular pathways: adipose inflammation as a mediator of obesity-associated cancer. *Clin Cancer Res*. 2013;19(22):6074-6083.
21. Agurs-Collins T, Ross SA, Dunn BK. The many faces of obesity and its influence on breast cancer risk. *Front Oncol*. 2019;9:765.
22. Sun X, Casbas-Hernandez P, Bigelow C, et al. Normal breast tissue of obese women is enriched for macrophage markers and macrophage-associated gene expression. *Breast Cancer Res Treat*. 2012;131(3): 1003-1012.
23. Morris PG, Hudis CA, Giri D, et al. Inflammation and increased aromatase expression occur in the breast tissue of obese women with breast cancer. *Cancer Prev Res (Phila)*. 2011;4(7):1021-1029.
24. Subbaramaiah K, Morris PG, Zhou XK, et al. Increased levels of COX-2 and prostaglandin E2 contribute to elevated aromatase expression in inflamed breast tissue of obese women. *Cancer Discov*. 2012;2(4):356-365.
25. Subbaramaiah K, Iyengar NM, Morrow M, Elemento O, Zhou XK, Dannenberg AJ. Prostaglandin E2 down-regulates sirtuin 1 (SIRT1), leading to elevated levels of aromatase, providing insights into the obesity-breast cancer connection. *J Biol Chem*. 2019;294(1):361-371.
26. Serra D, Mera P, Malandrino MI, Mir JF, Herrero L. Mitochondrial fatty acid oxidation in obesity. *Antioxid Redox Signal*. 2013;19(3): 269-284.
27. Zwick RK, Guerrero-Juarez CF, Horsley V, Plikus MV. Anatomical, physiological, and functional diversity of adipose tissue. *Cell Metab*. 2018;27(1):68-83.
28. Auguet T, Guiu-Jurado E, Berlanga A, et al. Downregulation of lipogenesis and fatty acid oxidation in the subcutaneous adipose tissue of morbidly obese women. *Obesity (Silver Spring)*. 2014;22(9):2032-2038.
29. Guiu-Jurado E, Auguet T, Berlanga A, et al. Downregulation of de novo fatty acid synthesis in subcutaneous adipose tissue of moderately obese women. *Int J Mol Sci*. 2015;16(12):29911-29922.
30. de Mello AH, Costa AB, Engel JDG, Rezin GT. Mitochondrial dysfunction in obesity. *Life Sci*. 2018;192:26-32.
31. Blackburn HL, Ellsworth DL, Shriver CD, Ellsworth RE. Role of cytochrome P450 genes in breast cancer etiology and treatment: effects on estrogen biosynthesis, metabolism, and response to endocrine therapy. *Cancer Causes Control*. 2015;26(3):319-332.
32. Rodriguez M, Potter DA. CYP1A1 regulates breast cancer proliferation and survival. *Mol Cancer Res*. 2013;11(7):780-792.
33. Tsai P-C, Glastonbury CA, Eliot MN, et al. Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *Clin Epigenetics*. 2018; 10(1):126.
34. Andres SA, Bickett KE, Alatoum MA, Kalbfleisch TS, Brock GN, Wittliff JL. Interaction between smoking history and gene expression levels impacts survival of breast cancer patients. *Breast Cancer Res Treat*. 2015;152(3):545-556.
35. Tekpli X, Zienolddiny S, Skaug V, Stangeland L, Haugen A, Mollerup S. DNA methylation of the CYP1A1 enhancer is associated with smoking-induced genetic alterations in human lung. *Int J Cancer*. 2012;131(7):1509-1516.
36. Kim JH, Sherman ME, Curriero FC, Guengerich FP, Strickland PT, Sutter TR. Expression of cytochromes P450 1A1 and 1B1 in human lung from smokers, non-smokers, and ex-smokers. *Toxicol Appl Pharmacol*. 2004;199(3):210-219.
37. Mitsui Y, Chang I, Kato T, et al. Functional role and tobacco smoking effects on methylation of CYP1A1 gene in prostate cancer. *Oncotarget*. 2016;7(31):49107-49121.
38. Fa S, Larsen TV, Bilde K, et al. Changes in first trimester fetal CYP1A1 and AHRR DNA methylation and mRNA expression in response to exposure to maternal cigarette smoking. *Environ Toxicol Pharmacol*. 2018;57:19-27.
39. Guida F, Sandanger TM, Castagne R, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum Mol Genet*. 2015;24:2349-2359.
40. Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect*. 2014;122(2):131-137.
41. Fasanelli F, Baglietto L, Ponzi E, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in three prospective cohorts. *Nat Commun*. 2015;6(10192):1-9.
42. Cao D, Qu Y, Zhang X, et al. High expression of F2RL3 correlates with aggressive features and poor survival in clear cell renal cell carcinoma. *J Cancer*. 2018;9(18):3400-3406.
43. Riabov V, Yin S, Song B, et al. Stabilin-1 is expressed in human breast cancer and supports tumor growth in mammary adenocarcinoma mouse model. *Oncotarget*. 2016;7(21):31097-31110.
44. Shaykhiev R, Krause A, Salit J, et al. Smoking-dependent reprogramming of alveolar macrophage polarization: implication for pathogenesis of chronic obstructive pulmonary disease. *J Immunol*. 2009;183(4): 2867-2883.
45. Dhanesuan N, Sharp JA, Blick T, Price JT, Thompson EW. Doxycycline-inducible expression of SPARC/Osteonectin/BM40 in MDA-MB-231 human breast cancer cells results in growth inhibition. *Breast Cancer Res Treat*. 2002;75(1):73-85.
46. Nagai MA, Gerhard R, Fregnani JH, et al. Prognostic value of NDRG1 and SPARC protein expression in breast cancer patients. *Breast Cancer Res Treat*. 2011;126(1):1-14.
47. Said N. Role of SPARC in cancer; friend or foe. *Ann Carcinog*. 2016;1(1).
48. Wang Z, Wang D, Wang Y. Cigarette smoking and adipose tissue: the emerging role in progression of atherosclerosis. *Mediators Inflamm*. 2017;2017:3102737-3102711.
49. Andersson K, Arner P. Systemic nicotine stimulates human adipose tissue lipolysis through local cholinergic and catecholaminergic receptors. *Int J Obes Relat Metab Disord*. 2001;25(8):1225-1232.

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

50. Wang J, Heng YJ, Eliassen AH, et al. Alcohol consumption and breast tumor gene expression. *Breast Cancer Res*. 2017;19(1):108.
51. Xhemalce B, Robson Samuel C, Kouzarides T. Human RNA methyltransferase BCDIN3D regulates microRNA processing. *Cell*. 2012;151(2):278-288.
52. Yao L, Chi Y, Hu X, et al. Elevated expression of RNA methyltransferase BCDIN3D predicts poor prognosis in breast cancer. *Oncotarget*. 2016;7(33):53895-53902.
53. Charity Fix BL. A role for mast cells in alcohol-induced tissue damage and remodeling. *J Clin Exp Pathol*. 2015;5(2).
54. Varricchi G, Galdiero MR, Loffredo S, et al. Are mast cells MASTers in cancer? *Front Immunol*. 2017;8(424).
55. Glajcar A, Szpor J, Pacek A, et al. The relationship between breast cancer molecular subtypes and mast cell populations in tumor microenvironment. *Virchows Arch*. 2017;470(5):505-515.
56. Carpenco E, Ceausu RA, Cimpean AM, et al. Mast cells as an indicator and prognostic marker in molecular subtypes of breast cancer. *In Vivo*. 2019;33(3):743-748.
57. Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: a look at the evidence. *Womens Health (Lond Engl)*. 2015;11(1):65-77.
58. Martínez-Reyes I, Chandel NS. Mitochondrial TCA cycle metabolites control physiology and disease. *Nat Commun*. 2020;11(1):102.
59. Wilson DF, Matschinsky FM. Ethanol metabolism: the good, the bad, and the ugly. *Med Hypotheses*. 2020;140:109638.
60. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. *Lancet*. 2019;394(10204):1159-1168.
61. Hall P, Ploner A, Bjohle J, et al. Hormone-replacement therapy influences gene expression profiles and is associated with breast-cancer prognosis: a cohort study. *BMC Med*. 2006;4:16.
62. Sieuwerts AM, De Napoli G, van Galen A, et al. Hormone replacement therapy dependent changes in breast cancer-related gene expression in breast tissue of healthy postmenopausal women. *Mol Oncol*. 2011;5(6):504-516.
63. Harlid S, Xu Z, Kirk E, Wilson LE, Troester MA, Taylor JA. Hormone therapy use and breast tissue DNA methylation: analysis of epigenome wide data from the normal breast study. *Epigenetics*. 2019;14(2):146-157.
64. Husby A, Wohlfahrt J, Øyen N, Melbye M. Pregnancy duration and breast cancer risk. *Nat Commun*. 2018;9(1):4255.
65. Peri S, de Cicco RL, Santucci-Pereira J, et al. Defining the genomic signature of the parous breast. *BMC Med Genomics*. 2012;5(1):1-11.
66. Belitskaya-Lévy I, Zeleniuch-Jacquotte A, Russo J, et al. Characterization of a genomic signature of pregnancy identified in the breast. *Cancer Prev Res*. 2011;4(9):1457-1464.
67. Baiju N, Sandanger TM, Sætrom P, Nøst TH. Gene expression in blood reflects smoking exposure among cancer-free women in the Norwegian Women and Cancer (NOWAC) postgenome cohort. *Sci Rep*. 2021;11(1):680.
68. Fasanelli F, Baglietto L, Ponzi E, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat Commun*. 2015;6:10192.
69. StatisticsNorway. *Statbank: Tobacco, Alcohol and Other Drugs*. Statistics Norway; 2020 https://www.ssb.no/en/helse/statistikker/royk
70. Gram IT, Little MA, Lund E, Braaten T. The fraction of breast cancer attributable to smoking: the Norwegian women and cancer study 1991–2012. *Br J Cancer*. 2016;115(5):616-623.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Krum-Hansen S, Standahl Olsen K, Anderssen E, Frantzen JO, Lund E, Paulssen RH. Associations of breast cancer related exposures and gene expression profiles in normal breast tissue—The Norwegian Women and Cancer normal breast tissue study. *Cancer Reports*. 2023;6(4):e1777. doi:10.1002/cnr2.1777

25738348, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cnr2.1777 by Norwegian Institute Of Public Healt Invoice Receipt DFO, Wiley Online Library on [19/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Paper IV

Lund, E., Krum-Hansen, S., Olsen, K.S., Shvetsov, N., Snapkow, I., Gavriluk, O., Busund, L.T.R., Frantzen, J.O., Holden, M. & Holden, L.

No impact of parity on global gene expression levels in breast cancer tissue and normal breast tissue – a nested case-control study in the NOWA Postgenomic biobank

Submitted manuscript

4.12

No impact of parity on global gene expression levels in breast cancer tissue and normal breast tissue – a nested case-control study in the Norwegian Women and Cancer study

Eiliv Lund[1], Sanda Krum-Hansen[1,2,3], Karina Standahl Olsen[1], Nikita Shvetsov[4], Igor Snapkow[5], Oxana Gavriluk[6,7], Lill-Tove Rasmussen Busund[8,9], Jan Ole Frantzen[10], Marit Holden[11], Lars Holden[11]

1. Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway
2. Department of Hematology and Oncology, Stavanger University Hospital, Stavanger, Norway
3. Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway
4. Department of Computer Science, Faculty of Science and Technology, UiT The Arctic University of Norway
5. Department of Chemical Toxicology, Norwegian Institute of Public Health, Oslo, Norway
6. Research Centre for Obstetrics and Gynaecology (RESCOG), Oslo University Hospital, Oslo, Norway
7. Department of Oncology, Akershus University Hospital, Lørenskog, Norway
8. Department of Medical Biology, UiT The Arctic University of Norway, Tromsø, Norway
9. Department of Clinical Pathology, University Hospital of North Norway, Tromsø, Norway
10. Department of Radiology, University Hospital of North Norway, Narvik, Norway
11. Norwegian Computing Center, Oslo, Norway

Correspondence: Igor Snapkow, Division of Climate and Environmental Health, Department of Chemical Toxicology, Norwegian Institute of Public Health, Lovisenberggata 8, 0456 Oslo, Norway,Email igor.snapkow@fhi.no;

Eiliv Lund, Department of Community Medicine, UiT The Arctic University of Norway, Hansine Hansens veg 18, 9019 Tromsø, Norway, Email eiliv.lund@uit.no

Keywords: full-term pregnancy, breast cancer, NOWAC postgenome biobank, systems epidemiology, parity

ABSTRACT

Background: We have previously shown that the linear decrease in post-menopausal breast cancer incidence with increasing parity is associated with similar changes in hundreds of genes in peripheral blood cells of healthy women, but not in cancer patients' blood. Here, we explore the linear relationship between parity and genome gene expression in breast cancer tissue versus normal breast tissue. Both studies were based on the The Norwegian Women and Cancer (NOWAC) study.

Methods: NOWAC is a prospective cohort study (N=172 000) built 1991-2007. In the years 2006-2010 eleven Norwegian hospitals participated in the NOWAC biopsy study. Women with a diagnostic biopsy of breast cancer were asked to give a second biopsy for research. Normal breast tissue was collected at the national mammographic screening program among NOWAC women in North Norway 2010-2011. The final analyses consisted of 279 age-matched case-control pairs. The same protocol for biopsy taking was used throughout. The paired matched design was kept in all laboratory analyses. Differences in gene expression (Illumina microarray) between breast cancer cases and controls were identified using the Bioconductor R-package limma. The PAM-50 gene expression was used for the classification of intrinsic subtypes in both tissues.

Results: Almost all genes were significantly differentially expressed between cases and controls; 10013 out of 11308 (FDR q-values 0.05), 5768 upregulated and 4245 downregulated. With mostly significant genes no gene set enrichment analyses were performed. When parity was included, significant changes were found only in three genes in normal tissue and none in breast cancer tissue. Classification by PAM-50 showed no relationship between parity and the different intrinsic subtypes. Control tissues classified as luminal A had increased risk of developing breast cancer compared to normal-like (p=0.02).

Conclusions: We found no differences in gene expression between normal and breast cancer tissue dependent on number of full-term pregnancies or parity.

65 **Introduction**

66 Three centuries ago, it was suggested that nulliparity, or the lack of childbirth, implied a
67 higher risk of breast cancer among nuns.[1] Subsequent studies conducted up until the
68 1960s confirmed the protective effect of parity on breast cancer, which was partially
69 attributed to lactation (1,2). However, in 1970, a large hospital-based case-control study
70 spanning seven countries shifted the focus from parity to age at first full-term pregnancy
71 (3). The results of this study indicated that women who had their first full-term
72 pregnancy at an early age experienced a strong breast cancer risk reduction. Multiparity
73 was found to offer no added risk reduction. As a result, the biological rationale behind
74 the protective effect of parity against the development of breast cancer was simplified,
75 leading to its consideration as a dichotomous variable based on age at first full-term
76 pregnancy. However, in the mid-1980s epidemiologists began to realize that hospital-
77 based case-control studies were vulnerable to both selection and recall bias (4). It was
78 also recognized that matched cases and controls should be kept together through the
79 analyses (5). Later, large epidemiological prospective studies have consistently shown a
80 linear decrease in risk of 7– 8% for each child, at least up to 6 children, when adjusted
81 for age at first birth (6-14). Studies with high-parity women, those with seven or more
82 children, have shown a further protective effect of childbirths, but the statistical power
83 of these studies has been too weak to draw firm conclusions (15-20).

84 With the development of technologies for mRNA analyses, gene expression studies
85 opened new possibilities for testing biological hypotheses concerning the impact of
86 parity and age at first birth on breast cancer development. One of the leading hypotheses
87 of breast cancer has been that the first full-term pregnancy protects against breast cancer
88 through a specific genomic signature in human breast tissue (21,22) but the theory does
89 not clearly explain the protective effect of each additional pregnancy shown in almost
90 all studies of parity and breast cancer (6-20). An important methodological issue is the
91 validity of studies comparing breast cancer tissues with adjacent normal tissue in the
92 same breast with the use of control tissue samples taken from normal breast diseases
93 like hyperplasia and benign tumors. Few large studies have collected breast tissue from
94 healthy women as controls (23).

95 A previous epidemiological analysis in NOWAC showed that each additional full-term
96 pregnancy reduced the risk of breast cancer with 8% (24). This linear change was tested
97 against changes in global gene expression in the NOWAC biobank. For hundreds of
98 genes the same changes were found according to parity in blood from healthy women,
99 but no changes found in breast cancer patients. Here we will expand this finding by
100 looking on changes in global gene expression in buffered tissue samples in a linear
101 model of parity. We have collected tumor tissue at time of diagnoses of breast cancer
102 patients and among healthy postmenopausal women in NOWAC using a matched case-
103 control design. The objective was to compare the impact of each additional full-term
104 pregnancy on gene expression in breast cancer tissue compared to normal breast tissue.

105

106 ## Material and methods

107 *The Norwegian Women and Cancer study*

108 The NOWAC study has a prospective design and recruited 172,000 women randomly
109 sampled from the National Population Register in Norway, beginning in 1991 (25).
110 Invitations and questionnaires were sent to these women by mail, with additional
111 questionnaires sent at various intervals. All participating women were notified that
112 follow-up regarding cancer incidence and mortality would be conducted through linkage
113 to national registries based on unique national identification numbers assigned to all
114 Norway residents alive during the 1960 census or born afterward. The information on
115 parity, as displayed in Table 1, was obtained from the last questionnaire completed
116 before the biopsies of normal or malignant breast tissues were performed.

117 *Collection of breast cancer tissue samples*

118 Between 2006 and 2010, the NOWAC study involved the collection of tumor biopsies
119 from participants across eleven different hospitals in Norway (26). During that time,
120 around one-third of women born between 1943-1957 in Norway had enrolled in the
121 NOWAC study by completing one or more questionnaires. To increase the probability
122 of finding NOWAC participants among women who have undergone breast biopsy,
123 additional sampling was conducted in certain counties. In collaboration with the
124 Norwegian Breast Cancer Group, participating hospitals asked women to donate an
125 additional biopsy after the diagnostic biopsy. After the biopsy, a blood sample and a
126 one-page questionnaire were also collected. In total, 316 biopsies were obtained
127 (Supplementary table 1) and 311 cases were deemed eligible for this case-control study.

128 *Collection of normal breast tissue*

129 From October 2010 to May of 2011, a sample collection was conducted at the
130 University Hospital of North Norway in Tromsø, where breast tissue biopsies, buffered
131 blood samples, and a two-page questionnaire were obtained from women with a normal
132 mammogram who participated in the national screening program (27). Women were
133 asked if they had previously taken part in the NOWAC study. The biopsies were
134 obtained after written informed consent, and the response rate was 64%. An experienced
135 radiologist obtained tissue samples after mammography, using ultrasound-guided needle
136 biopsy from the gland tissue of the upper lateral quadrant of the left breast. The
137 procedure was standardized, and one biopsy was collected from each participant. If the
138 material was macroscopically sparse, a second biopsy was obtained. A total of 480
139 biopsies were taken, with either 14- or 16-gauge needles, but only gauge 14 samples
140 were used for this comparison. In a later linkage to the Norwegian Cancer Registry,
141 conducted three years after the sampling period, five control women were diagnosed

142 with breast cancer and were therefore excluded from this study. In the end, 311 women
143 were matched with cancer biopsies (cases) comprising the study population.

144 *Laboratory procedures*

145 Samples obtained from breast cancer and normal tissue were immediately placed in a
146 PAX-gene blood RNA collection kit (Preanalytix, Qiagen, Hilden, Germany) and stored
147 at room temperature for 24 hours. The tube in the kit contained a protective buffer,
148 RNAlater, which preserves mRNA in blood, allowing for long-term frozen storage at -
149 70°C.

150 RNA preparation was conducted at The Department of Cancer Genetics, Oslo
151 University Hospital, Oslo, Norway. Tissue biopsies were homogenized using
152 TissueLyser LT (Qiagen, Hilden, Germany) and 5 mm steel beads. Total RNA and
153 genomic DNA were isolated using the AllPrep DNA/RNA Mini Kit from Qiagen (Cat.
154 No. 80204) and the QIAcube instrument (Qiagen, Hilden, Germany). A custom protocol
155 was followed for the extraction
156 (RNA_AllPrepDNARNA_AnimalCells_AllPrep350_ID2481). RNA was stored at
157 −80°C. RNA quantity was measured by NanoDrop (Thermo Fisher Scientific,
158 Wilmington, Delaware, USA). The BioAnalyzer 2100 and Agilent RNA 6000 Nano kit
159 (cat. No. 5067–1511) were used to evaluate RNA integrity (Agilent, Santa Clara, US).

160 mRNA gene expression was analyzed at a certified Illumina service provider (NTNU
161 Genomics Core Facility, Trondheim, Norway). Briefly, RNA was amplified with
162 Ambion's Illumina® TotalPrep RNA amplification kit (Cat #AMIL 1791) using 400 ng
163 of total RNA as input material. Incorporation of biotin-labeled nucleotides was
164 performed overnight (14 h) at 37°C in vitro transcription (IVT). cRNA was quantified
165 using the NanoDrop ND-1000 (NanoDrop, Wilmington, USA), and cRNA integrity was
166 determined by electrophoresis using the Experion Bioanalyzer (BioRad). A total of
167 750 ng of biotin-labeled cRNA was hybridized to IlluminaHumanHT-12 v.4 expression
168 bead chip (Illumina®). Beadchips were scanned with Illumina BeadArray Reader.
169 Numerical results were extracted with Bead Studio v3.0.19.0 without any normalization
170 or background subtraction.

171 To minimize technical noise and batch effects, the cases and controls within each pair
172 were kept together throughout all laboratory procedures. All laboratory work was
173 performed for all case-control pairs in 2012.

174 *Dataset assembly and outlier removal*

175 The combined case-control dataset was selected based on inclusion criteria of
176 individuals aged between 50-69 years, as defined by the mammography screening
177 program. This dataset comprised 622 samples or 311 pairs with 47323 probes. Initially,
178 10 samples were removed due to their gene expression values being 0. Individuals who
179 were considered borderline outliers were excluded if their laboratory quality measures

180 were below given thresholds (RNA integrity number value < 7, 260/280 ratio < 2,
181 260/230 ratio < 1.7, and 50 < RNA < 500). To identify outlier samples, we used the
182 nowaclean package (28) and considered a sample as an outlier if it met at least two of
183 the three outlier criteria. In addition, we examined four plots (PCA plot, boxplot,
184 density plot, and MA-plot) for each sample defined as a potential outlier by at least one
185 outlier criterion. Based on these plots, we identified one more sample as outlier. In total,
186 we identified 16 technical outliers using the nowaclean package, and these were
187 removed along with their matching controls or cases. Finally, controls who developed
188 cancer within three years after recruitment, including 5 breast cancer and 1 lymphoma,
189 were withdrawn along with their cases. The resulting dataset consisted of 558 samples,
190 with 279 case-control pairs matched by birth year. After preprocessing the dataset
191 consisted of 11308 genes.

192 In principle component analysis of the entire dataset the data exhibits two distinct
193 clusters, one for the control group and another for the cases (Figure 1A). This outcome
194 is not unexpected, given the marked differences in gene expression between normal and
195 tumor tissue. In turn, the PCA plots for matched case-control pairs did not reveal any
196 clusters (Figure 1B). Moreover, we did not detect significant batch effects in either the
197 cases or controls groups (Figures 1C and D).

198 *Preprocessing the gene expression datasets*

199 Each dataset was background corrected using negative control probes, $\log_2$ transformed
200 using a variance stabilizing technique (29) and quantile normalized. We retained probes
201 present in at least 70% of the samples. If a gene was represented with more than one
202 probe, the average expression level of the probes was used as the expression value for
203 the gene. The probes were translated to genes using lumiHumanIDMapping (30).
204 Finally, the differences in the $\log_2$ gene expression levels for each case-control pair
205 were computed and used in the statistical analyses.

206 *Subclassifications of tissues using the PAM50*

207 Around 2000, the introduction of gene expression analyses in breast cancer tissues
208 resulted in a new classification of breast cancers (31,32). This innovative approach led
209 to the development of a novel classifier, Prediction Analysis of Microarray 50
210 (PAM50), which uses hierarchical clustering of gene expression to enable more
211 personalized treatment of cancer.

212 Using the preprocessed datasets of normal and cancer tissues, we applied the PAM50
213 classifier to all samples. This classifier was constructed based on the hierarchical
214 clustering of genes in tumor tissue compared to normal breast tissue and involved a total
215 of 50 genes that were found sufficient to classify cancer tissue into five distinct
216 subtypes: luminal A, luminal B, HER2-enriched, basal-like, and normal-like. We then

217 examined the relationship between PAM50 subtypes and parity, which was stratified
218 into two categories: 0 versus 1 or more children or 0 versus 0,1–3 and 4–8 children.

219 To evaluate the utility of PAM50 as also a long-term prognostic test, we performed a
220 new linkage to the Norwegian Cancer Registry with follow-up data available up until
221 2018 for women who participated in the normal tissue study.

222 *Statistical methods*

223 To identify differentially expressed genes according to disease state and parity, we used
224 the Bioconductor R-package Limma (Linear models for microarrays) (33). Using gene
225 expression values as responses and paritiesp_c^{case} (p_c^{ctrl}) as explanatory variables we
226 model the gene expression values, denoted as $Y_{g,c} = Y_{g,c}^{case} - Y_{g,c}^{ctrl}$, assuming linearity in
227 parities:

228 $$(1) \qquad Y_{g,c} = \alpha + \beta^{ctrl} p_c^{ctrl} + \beta^{case} p_c^{case} + \varepsilon_{g,c}.$$

229 This model allows us to identify genes that are influenced by parity separately in case
230 and control.

231 For identifying genes that are differentially expressed between cases and controls, we
232 use the simplified model

233 $$(2) \qquad Y_{g,c} = \alpha + \varepsilon_{g,c}.$$

234 We aimed to identify not only individual genes but also sets of genes influenced by
235 parity. To achieve this, we utilized Limma in a way similar to the method used for
236 individual genes. However, instead of gene expression values, we used enrichment
237 scores for gene sets as responses in the model. The enrichment scores were derived
238 from the gene expression values using the Bioconductor R-package, GSVA (gene set
239 variation analysis) (34).

240 We obtained eight different collections of gene sets from the Molecular Signatures
241 Database (35) including C1 – C7 and H. For each collection, we report the number of
242 significant genes after adjusting for multiple testing.

243

244

245 **Results**

246 After stratifying cases and controls based on parity status, the crude odds ratio of
247 developing breast cancer for nulliparous compared to parous women was 1.16 (0.75 –
248 1.78) (Table 1).

249 *Gene expression profiles*

250 Initially, we investigated gene expression profiles without considering any potential
251 variables that could modulate breast cancer risk. Upon comparison of cases and
252 controls, we identified 10013 and 10021 differentially expressed genes at a false
253 discovery rate (FDR) of 5% and 10%, respectively using equation (2). As almost all
254 genes are differentially expressed between cases and controls, we will not do any gene
255 sets enrichment analyses.

256 Next, we examined whether parity, a well-known modulating factor in breast cancer,
257 could influence gene expression. We conducted single gene and gene set analyses
258 separately for cases and controls, with parity categorized into different groups. The
259 results, summarized in Table 2, revealed only a few significant genes in the case group
260 and none in the control group at FDR of 5% and 10%. However, analysis of gene set
261 enrichment in MSigDB collections showed a significant number of enriched gene sets
262 among cases, particularly in groups C2, C3, C5, and C7. In contrast, controls had a low
263 number of significant genes in almost all gene set score groups.

264 Furthermore, we categorized parity into three groups (0, 1 – 3, and 4 – 8) to model gene
265 expression as linear. At an FDR of 5%, only three genes were significantly expressed
266 among cases and none among controls. Increasing the FDR level resulted in an
267 increased number of significantly expressed genes among cases, but no changes were
268 observed in the control group. Table 3 provides a summary of the differentially
269 expressed genes.

270 Among the top 40 most differentially expressed genes, listed in Supplementary table 2
271 for cases and Supplementary table 3 for controls, only three genes were significantly
272 expressed among cases with an FDR q-value of 0.05% (Table 3). These genes included
273 PARP16, MRPL23, and RBMS1, and their log2 gene expression values are
274 demonstrated in Figure 2. No significant genes were observed in the control group, even
275 without FDR adjustment, when the model included grouping of parity (0, 1 – 3, and 4 –
276 8).

277 *PAM50 classification*

278 The PAM50 subclassification for both cancerous and normal tissue can be found in
279 Table 4. Parity was grouped into two categories: nulliparous versus parous, or
280 nulliparous versus parity groups 1 – 3 and 4 – 8. While the number of nulliparous cases

281 was slightly higher than that of parous women, there were no significant differences in
282 subclassification based on parity (p = 0.23 and p = 0.59, respectively).

283 Among the nulliparous cases, only one was classified as normal-like, which accounted
284 for 1.9%, compared to 9.8% of parous women. All control tissue samples, except for 13
285 samples classified as luminal A and one as basal-like, were classified as normal-like. In
286 a subsequent follow-up in 2018, six more women were diagnosed with breast cancer.
287 Among those initially classified as luminal A, two were diagnosed with breast cancer,
288 whereas normal-like classified had six cases of breast cancer (Table 5). Among the three
289 women diagnosed with breast cancer within the first year of follow-up, one had luminal
290 A classification.

291 During follow-up, 3.0% of all normal-like controls developed breast cancer, whereas
292 21.4% of those with luminal A classification developed breast cancer (Fisher's exact
293 test, p = 0.02).

294

295

296

297

298 **Discussion**

299 The key finding of this study is that parity did not result in any changes in global gene
300 expression in either breast cancer tissue or normal breast tissue when analyzed in a
301 linear function. Most genes analyzed in the study showed significant differences in
302 expression between normal breast tissue and breast cancer tissue. However, when parity
303 was included in the analysis, the number of significant genes in the cancer tissues
304 reduced considerably and there were no significant genes observed in the normal tissue
305 samples even with a FDR of 20%. To the best of our knowledge, this work is the first
306 large-scale investigation of gene expression profiles in breast cancer tissues using a
307 matched case-control design with normal breast tissue samples as controls.

308 The findings of a linear change in gene expression in blood in healthy women were not
309 reproduced in normal breast tissue. Thus, the effect of each pregnancy on global gene
310 expression was found in blood but not in tissues.

311 These findings contrast with other studies of gene expression in tissues. In a study by
312 Russo (21,22) many significantly differentially expressed genes were found in normal
313 breast tissue of parous women compared to breast cancer tissue and cancer-free tissue of
314 nulliparous. The Russo study included 25 benign tissue samples with various diagnoses
315 such as fibroadenoma, adenosis, papilloma, and ductal hyperplasia as controls. The
316 present study employed a matched case-control design with population-based sampling
317 of normal breast tissue. The age-matching was maintained throughout the statistical
318 analysis. The present study utilized bulk tissue biopsies for both breast cancer and
319 normal tissue samples, while the Russo study used laser capture microdissection.
320 Although both methods are deemed acceptable, bulk tissue biopsies generally yield a
321 higher proportion of normal-like subtypes. However, this was not observed in the
322 present study where only 8.3% of cancer tissue samples exhibited normal-like subtypes.
323 Another study by Rotunno (36) compared a gene expression signature in 150 paired
324 tumors and adjacent benign breast tissues. The analyses compared nulliparous with
325 parous and found hundreds of significantly upregulated genes, but the signature did not
326 vary by age at first birth. Common to these studies are the simple comparison between
327 nulliparous and all parous. From an epidemiological point of view each pregnancy gives
328 the same protection.

329 As almost all genes were differentially expressed between cases and controls, no further
330 gene set enrichment analyses were performed. Only three genes were found to be
331 affected by changing parity in normal breast tissue, with only one of these genes, RNA
332 binding protein RBMS1, being linked to immune-cold triple-negative breast cancer.
333 None of the top 40 genes in cases and controls were found to be common.

334 The PAM50 test is used for treatment decisions based on the subgroup defined by the
335 test. Notably, the analysis revealed no significant changes in gene expression according

336 to parity in normal tissue samples, regardless of the subgroup. Luminal A classified
337 normal tissue samples may represent false positives or may predict later breast cancer.

338 One important consideration in the context of parity is its close association with
339 lactation. Breastfeeding practices vary widely between countries and even among
340 women within the same cultural context. One large collaborative study found that the
341 risk reduction associated with parity was partly driven by breastfeeding but with no
342 information on receptor status (37). In the Nurses' Health Study breastfeeding was
343 associated with ER- breast cancer, but not with ER+ cancer (38). Another consortium
344 study found a similar protective effect of lactation for ER- cancer (39). In an analysis of
345 breastfeeding in the EPIC study no general associations were found related to breast
346 cancer subtypes (40).

347 For decades, breast cancer has been considered a hormone-dependent cancer (41-42).
348 Increased levels of all endogenous sex hormones have been shown to increase the risk
349 of postmenopausal breast cancer. High levels of endogenous hormones are linked to
350 adiposity, smoking and alcohol consumption, but not to parity-related factors (43). The
351 same pattern was demonstrated in the analyses of gene expression in the healthy,
352 normal controls of NOWAC where gene expression was associated with external risk
353 factors rather than with parity (23). Both combined oral contraceptives and hormonal
354 replacement therapy have been classified as human carcinogens (44).

355 *The pseudo semi-allograft theory.*

356 The results of this analysis should be considered in the context of studies examining
357 gene expression in healthy individuals and breast cancer patients in both blood and
358 tissues. A thorough investigation into risk factors for breast cancer and gene expression
359 patterns in normal tissue found that parity had little impact on gene expression
360 differences. In contrast, obesity, smoking, and alcohol consumption were associated
361 with the greatest number of genes and pathways (23). Previous studies using the
362 NOWAC post genome biobank material revealed that changes in parity were linked to
363 changes in gene expression for hundreds of genes in healthy women's blood, but no
364 changes were observed in the blood of breast cancer patients at diagnosis (24).
365 Additionally, a comparison of gene expression in tumor tissue and blood from the same
366 patient found no strong links between the systemic response and the local tumor
367 response except for tumors with strong immune properties (45).

368 The lack of response to pregnancy in both cancer and healthy tissues raises questions
369 about whether there exist other biological mechanisms underlying the protective effect
370 of parity on breast cancer risk (46).

371 One possible explanation for this protective effect is the long-term impact of
372 immunological changes during each pregnancy (47,48,49). This process, which allows
373 the survival of the fetus while protecting the mother, involves unique and tightly
374 regulated changes in the immune system. The fetus is a semi-allograft or a foreign body

375 for the mother's immune system. Each pregnancy leaves behind many memory cells
376 which could play a role in identifying tumor cells that originate from transformed
377 mother cells, such as breast cancer, acting as a pseudo semi-allograft. These findings
378 support the hypothesis that breast cancer may be a pseudo semi-allograft, as there was
379 no impact on gene expression with increasing parity in normal breast tissue or breast
380 cancer tissue.

381

382 **Strengths**

383 The study employs a nested case-control design with a rigorous methodology to ensure
384 that the cases and control pairs were kept together during laboratory work and statistical
385 analyses, which minimizes noise in the data. Both cancer tissue and normal tissue
386 samples were collected from women who participated in the NOWAC study. The study
387 is representative of the Norwegian female population in terms of parity (50). The
388 external validity of NOWAC strengthens the study's conclusion. In the context of the
389 normal tissue study, a response rate of 64 percent may be considered high.

390 To ensure that all participants had information on parity, invitations were limited to
391 NOWAC participants, and control tissues were collected from only one large screening
392 center. The age of the women in the gene expression analyses was restricted to more
393 than 50 years, or postmenopausal. This age-span made the use of the ongoing screening
394 program natural.

395 The discovery of 13 luminal A cases among the control group suggests a possible lack
396 of sensitivity of the PAM-50 assay in detecting cancer among healthy individuals
397 (Table 5).

398

399 **Limitations.**

400 A potential limitation of the study is the somewhat different methodologies for
401 sampling of the cases and controls. This was due to the collaboration with the
402 Norwegian Breast Cancer Group. Eleven hospitals accepted to participate. They all
403 followed the same strict protocol with a scientific biopsy taken after the diagnostic one
404 by a surgeon or radiologist. The controls were collected in one center in North Norway,
405 UNN the regional university hospital. The same procedures were followed. Both cases
406 and controls belonged to the same source population, participants of the nationwide
407 cohort NOWAC with 172 000 participants.

408 It is important to note that the number of children in our study was limited due to the
409 low fertility rates in Norway, resulting in a relatively small number of women who had
410 given birth to more than four children.

411 In this study material, 8.7% of the samples exhibited a normal-like phenotype,
412 consistent with other published studies (51,52). Our control group comprised healthy
413 women who had been invited for mammography at the screening center and received
414 negative results. Comparisons between tissue biopsies and microdissected cancer cells
415 have been made, with the former showing reasonable comparability to the latter.

416

417 **Conclusion.**

418 Our analysis did not reveal any significant impact of increasing parity on gene
419 expression in either breast cancer or normal breast tissue. These results do not provide
420 support for the theory that the protective effect of pregnancy on breast cancer risk is
421 primarily mediated by tissue-level changes during the first pregnancy. Thus, the long-
422 term protective mechanism of each additional full-term pregnancy may be related to
423 systemic changes in the immune system that occur during pregnancy.

424

425

426 **Abbreviations**

427 FDR, false discovery rate; NOWAC, Norwegian Women and Cancer study; PAM50,
428 Prediction Analysis of Microarray 50; PCA, principal component analysis

429

430 **Ethical approval and informed consent**

431 For the collection of breast cancer and normal tissues, participants were provided with
432 both oral and written information prior to the procedure and were required to give
433 written informed consent.

434 A feasibility study was conducted prior to the collection of biopsies to obtain
435 information on the willingness of healthy women attending the mammographic
436 screening program in Tromsø, Norway to participate in the normal tissue study. A total
437 of 81 women were approached at the screening center and provided with both oral and
438 written information about the study. Of these, 74% responded positively and agreed to
439 donate breast tissue. No response rate was collected for the cancer tissue study.

440 Approval for both the normal and tumor tissue studies was granted by the North
441 Norway Regional Committee for Medical and Health Research Ethics. (P REK NORD
442 20957).

443 The NOWAC study was approved by the Norwegian Data Inspectorate and the
444 Regional Ethical Committee of North Norway (REK). The study was conducted in
445 compliance with the Declaration of Helsinki, and all participants gave written informed
446 consent. The linkages of the NOWAC database to national registries such as the Cancer
447 Registry of Norway and registries on death and emigration were approved by the
448 Directorate of Health. The women were informed about these linkages. Furthermore, the
449 collection and storage of human biological material were approved by the REK in
450 accordance with the Norwegian Biobank Act. Women were informed in the letter of
451 introduction that the blood samples would be used for gene expression analyses.

452

453 **Data availability**

454 Due to ethical restrictions on this dataset, which contains potentially sensitive
455 information, the data will be made available upon reasonable request.

456

457 **Funding**

458 This study was supported by a grant from the European Research Council (ERC-AdG
459 232997 TICE) and a donation from Halfdan Jacobsen og frues legat (The Norwegian
460 Cancer Society) and the Northern Norway Regional Health Authority (Helse Nord).
461 The funders had no role in the design of the study; in the collection, analyses, and
462 interpretation of the data; in the writing of the manuscript; or in the decision to submit
463 for publication.

464
465 **Disclosure**
466 The authors declare that they have no conflicts of interest in this work.
467
468 **Author contributions**
469 EL started the NOWAC Study, initiated this methodological collaboration and raised
470 funding. SKH was responsible for the collection of normal breast tissue. JOF organized
471 the collection of tumor tissues. EL and LH carried out the conception and design of the
472 study. EL, LH, and MH were involved in the development of the methodology. LH and
473 MH performed data analysis (eg, statistical analysis, biostatistics, and computational
474 analysis). EL, SKH, KSO, NS, IS, OG, LTB, JOF, LH and MH wrote, reviewed, and
475 revised the manuscript. EL and LH provided administrative, technical, or material
476 support. EL conducted study supervision. NS had the responsibility for data
477 management. All authors contributed toward data analysis, drafting and critically
478 revising the paper, gave final approval of the version to be published, and agree to be
479 accountable for all aspects of the work.
480
481 **Acknowledgments**
482 We express our gratitude and admiration for the women who generously donated breast
483 tissue for this research project. We acknowledge the invaluable collaboration with the
484 Norwegian Breast Cancer Group, NBCG, that made it possible to collect biopsies from
485 eleven hospitals participating in NBCG as well as the surgeons at each of these clinics
486 who took the biopsies. We also appreciate the support provided by the radiologists and
487 staff at the Bryst Diagnostisk Senter (BDS) at the University Hospital of North Norway.
488 Finally, we are grateful to Bente Augdal, Merete Albertsen, and Knut Hansen for all
489 infrastructure and administrative support.
490
491 **Disclaimer**
492 Some of the data in this article are from the Cancer Registry of Norway. The Cancer
493 Registry of Norway is not responsible for the analysis or interpretation of the data
494 presented.
495
496 **Disclaimer**
497 Microarray service was provided by the Genomics Core Facility, Norwegian University
498 of Science and technology, and NMC – a national technology platform supported by the
499 functional genomics program (FUGE) of the Research Council of Norway.

500 **Figure 1. PCA biplots of the gene expression in tumor and normal tissues in the**
501 **dataset.**

502 Figure 2. **Significantly differentially expressed genes associated with increased**
503 **parity in cases (FDR 5%)**. Boxplots of the log_2 gene expression values for the cases
504 (red, tumor tissue) and controls (white, normal tissue) for the three most significantly
505 differentially expressed genes when modeling gene expression as linear in parity with
506 three groups: 0, 1-3 and 4-8 children.

507

508

509 Table 1. Distribution of parities for the cases (breast cancer tissue) and controls (healthy
510 women tissue)

Parity	0	1	2	3	4	5	6	7	8	**Sum**
Case	54	25	123	61	14	1	1	0	0	**279**
Control	48	37	116	59	16	2	0	0	1	**279**

511

512

513 Table 2. Number of significantly differentially expressed genes and significantly
514 enriched gene sets with parity as explainable variable.

	Parities	0, 1, 2, 3, 4, 5, 6, 8		0, 1, 2, 3 4-8		0, 1-3, 4-8		0, 1-2, 3-4, 5-8	
	FDR	Ctrl	Case	Ctrl	Case	Ctrl	Case	Ctrl	Case
Gene expression data (11308 genes)	5%	0	2	0	7	0	3	0	0
	10%	1	8	2	12	0	19	0	0
	15%	2	12	2	195	0	109	0	10
	20%	3	135	7	417	0	252	0	11
C1 (323 Positional gene sets)	5%	0	0	0	0	0	0	0	0
	10%	0	0	0	0	0	0	0	0
	15%	0	0	0	13	0	0	0	0
	20%	1	18	2	18	0	10	0	0
C2 (4720 Curated gene sets)	5%	0	3	0	78	0	84	0	0
	10%	0	278	0	595	0	604	0	0
	15%	0	674	0	1068	0	1012	0	0
	20%	0	1033	0	1503	0	1398	0	248
C3 (836 Motif gene sets)	5%	0	0	0	72	0	403	0	0
	10%	0	324	0	350	0	525	0	261
	15%	0	435	0	444	0	580	0	377
	20%	0	499	0	500	0	634	0	465
Gene set scores C4 (858 Computational gene sets)	5%	0	0	0	0	0	0	0	0
	10%	0	0	0	0	0	0	0	0
	15%	0	0	0	33	0	15	0	0
	20%	0	0	0	83	0	59	0	0
Gene set scores C5 6160 Gene Ontology gene sets	5%	1	0	1	0	0	0	0	0
	10%	1	76	1	684	0	899	0	0
	15%	1	861	1	1412	0	1508	0	0
	20%	2	1447	1	1886	0	2001	0	0
Gene set scores C6 (189 Oncogenic signatures)	5%	0	4	0	10	0	1	0	0
	10%	0	20	0	41	0	20	0	0
	15%	0	42	0	63	0	43	0	2
	20%	0	67	0	84	0	64	0	30
Gene set scores C7 (4872 Immunologic signatures)	5%	0	0	0	0	0	0	0	0
	10%	0	0	0	353	0	0	0	0
	15%	0	331	0	964	0	674	0	0
	20%	0	887	0	1455	0	1125	0	0
Gene set scores H (50 Hallmark gene sets)	5%	0	1	0	2	0	6	0	0
	10%	0	2	0	5	0	8	0	0
	15%	0	5	0	16	0	9	0	0
	20%	0	13	0	17	0	14	0	5

515
516
517
518

519 Table 3. Number of significantly differentially expressed genes when modeling gene
520 expression as linear in parity and using parity data with three possible values: 0, 1-3 and
521 4-8 with parity as explainable variable.
522

		Upregulated		Downregulated		Sum	
	FDR	Ctrl	Case	Ctrl	Case	Ctrl	Case
Gene expression data (11308 genes)	5%	0	1	0	2	0	3
	10%	0	5	0	14	0	19
	15%	0	24	0	85	0	109
	20%	0	70	0	182	0	252

523

524

525

526 Table 4. Classifications of normal breast tissue (controls) and breast cancer tissue
527 (cases) by PAM50 according to different parity groups.

Controls	PAM50				
Parity	Basal	Her2	LumA	LumB	Normal
0 1+	0 1	0 0	2 11	0 0	46 219
0 1-3 4-8	0 1 0	0 0 0	2 9 2	0 0 0	46 203 16
Cases	PAM50				
Parity	Basal	Her2	LumA	LumB	Normal
0 1+	2 20	6 24	18 69	27 90	1 22
0 1-3 4-8	2 18 2	6 22 2	18 65 4	27 83 7	1 20 2

528
529
530

531 Table 5. PAM50 classification of controls before and after exclusion of individuals who
532 developed breast cancer.
533

Before exclusion				
Basal	Her2	LumA	LumB	Normal
1	0	13	0	265
After exclusion				
Basal	Her2	LumA	LumB	Normal
1	0	11	0	259

534
535
536
537

538 Supplementary table 1 Number of biopsies from breast cancer patients from 11
539 Norwegian hospitals October 2006 – May 2010, in collaboration with Norwegian Breast
540 Cancer Group, NBCG (ref).
541

Hospital	**Number of biopsies**
Molde	2
Stavanger	28
Haukeland	16
Fredrikstad	31
Tromsø	37
Tønsberg	13
Radiumhospitalet	21
Ullevål	86
St.Olavs Hospital - Trondheim	24
Bodø	58
Total	316

542

543

544 Supplementary table 2. The 40 most differentially expressed genes (for the cases)
545 obtained when modeling gene expression as linear in parity and using parity data with
546 three possible values: 0, 1-3 and 4-8 children. LogFC is the estimated log-fold change in
547 gene expression when the parity increases with one.

	Parity for controls			Parity for cases		
Genes	logFC	p-value	FDR q-value	logFC	p-value	FDR q-value
PARP16	0.000346	0.97705	0.999577	-0.05886	1.55E-06	0.017537
MRPL23	0.009869	0.655749	0.999577	-0.10174	6.18E-06	0.034934
RBMS1	0.007934	0.760366	0.999577	0.116426	1.05E-05	0.039765
HIST1H4J	0.007036	0.789189	0.999577	-0.11386	2.01E-05	0.056871
RER1	0.017697	0.302297	0.999577	-0.07232	3.17E-05	0.065854
TARBP2	0.009464	0.413065	0.999577	-0.04849	3.49E-05	0.065854
C15orf29	-0.00721	0.479634	0.999577	-0.0418	5.22E-05	0.084349
TMEM33	0.000416	0.974999	0.999577	-0.05322	7.40E-05	0.087011
TMEM43	-0.02271	0.391178	0.999577	0.10609	7.56E-05	0.087011
C18orf32	0.00317	0.874786	0.999577	-0.08053	7.69E-05	0.087011
EIF4EBP3	-0.01162	0.57955	0.999577	-0.08342	8.55E-05	0.087902
YAP1	0.009504	0.749578	0.999577	0.117652	9.46E-05	0.089133
TMEM14C	0.030751	0.132469	0.999577	-0.07987	0.000109	0.08989
TMEM49	0.001829	0.952668	0.999577	-0.12036	0.000113	0.08989
KENAE	0.008363	0.58367	0.999577	-0.05939	0.000119	0.08989
ZBTB34	-0.01376	0.33413	0.999577	0.055079	0.000131	0.092674
TEX264	-0.00151	0.941318	0.999577	-0.07837	0.000153	0.097098
LRIG3	-0.02131	0.079905	0.999577	0.046313	0.000161	0.097098
SUCLG2	0.024344	0.321373	0.999577	-0.09351	0.000163	0.097098
LSM4	0.010818	0.700596	0.999577	-0.10644	0.000182	0.100263
CYP2R1	0.008734	0.767388	0.999577	-0.1115	0.000188	0.100263
KATNA1	-0.02998	0.081427	0.999577	0.06462	0.000195	0.100263
FAM190B	-0.00486	0.785915	0.999577	0.067176	0.000205	0.100713
LOC728037	0.016393	0.472889	0.999577	-0.08514	0.000224	0.104281
LACTB2	0.025338	0.497009	0.999577	-0.13868	0.000233	0.104281
CLN3	0.000566	0.981208	0.999577	-0.08911	0.000246	0.104281
HIST1H4K	0.0376	0.44918	0.999577	-0.18295	0.000266	0.104281
C14orf112	0.033982	0.222096	0.999577	-0.1024	0.000266	0.104281
FAM173A	0.036046	0.252438	0.999577	-0.11561	0.000275	0.104281
HAGHL	0.047356	0.097462	0.999577	-0.10469	0.000278	0.104281
LRDD	-0.00919	0.35026	0.999577	-0.03603	0.000286	0.104281
PATL1	-0.00654	0.721763	0.999577	0.067	0.000307	0.105305
PSMD13	0.005401	0.592638	0.999577	-0.03679	0.000308	0.105305
MBD5	-0.01147	0.076088	0.999577	0.023446	0.000317	0.105305
POLR3K	0.041684	0.128385	0.999577	-0.09878	0.00035	0.112981
LOC93622	0.004059	0.850309	0.999577	-0.0774	0.000366	0.11486
LOC643438	0.013744	0.607693	0.999577	-0.09607	0.000378	0.115613
C17orf90	0.006036	0.753367	0.999577	-0.06835	0.000424	0.120134
STK16	0.007681	0.483196	0.999577	-0.03883	0.000444	0.120134
AMOT	0.000933	0.97681	0.999577	0.113704	0.000446	0.120134

548

549

550

551 Supplementary table 3. The 40 most differentially expressed genes (for the controls)
552 obtained when modeling gene expression as linear in parity and using parity data with
553 three possible values: 0, 1-3 and 4-8. LogFC is the estimated log-fold change in gene
554 expression when the parity increases with one.

	Parity for controls			Parity for cases		
Genes	logFC	p-value	FDR q-value	logFC	p-value	FDR q-value
FAM105A	0.08340	0.00053	0.99958	-0.03008	0.20625	0.60095
XYLT1	0.11720	0.00070	0.99958	0.03519	0.30357	0.68600
HIST2H2BF	0.02923	0.00074	0.99958	-0.00007	0.99328	0.99826
LOC729905	0.04579	0.00086	0.99958	0.00761	0.57520	0.84747
C12orf5	0.06852	0.00133	0.99958	0.00380	0.85735	0.96187
WDR27	-0.04069	0.00160	0.99958	-0.00553	0.66489	0.88747
HS3ST3A1	0.10382	0.00217	0.99958	0.01811	0.58922	0.85459
SOX15	-0.10990	0.00228	0.99958	0.05172	0.14763	0.54212
C6orf25	0.01925	0.00233	0.99958	-0.00619	0.32338	0.70120
DIAPH1	-0.05517	0.00329	0.99958	-0.02572	0.16728	0.56499
FAM184A	-0.05394	0.00330	0.99958	0.01043	0.56633	0.84264
SEPT8	0.02082	0.00331	0.99958	0.00409	0.56048	0.83871
PCMT1	-0.05772	0.00363	0.99958	0.02155	0.27359	0.66251
LOC642468	0.02170	0.00387	0.99958	-0.00840	0.25956	0.65195
KCTD18	0.03909	0.00398	0.99958	-0.02281	0.09075	0.45811
LOC643668	0.05723	0.00422	0.99958	0.02727	0.16963	0.56537
PIM2	-0.12330	0.00427	0.99958	-0.00292	0.94553	0.98468
PELI2	-0.06879	0.00499	0.99958	0.07768	0.00153	0.15271
ZBTB40	-0.03953	0.00548	0.99958	0.00512	0.71654	0.91289
DMWD	-0.04479	0.00569	0.99958	-0.00129	0.93621	0.98230
RPL37	-0.04152	0.00572	0.99958	0.00828	0.57857	0.84933
LOC652864	0.06754	0.00578	0.99958	-0.02537	0.29607	0.67955
C2orf49	0.05183	0.00588	0.99958	-0.03360	0.07253	0.43009
RCBTB1	0.02158	0.00609	0.99958	-0.00393	0.61456	0.86388
WDR59	-0.04842	0.00611	0.99958	0.01694	0.33379	0.70722
MOBK1B	0.04525	0.00617	0.99958	-0.01107	0.49922	0.81395
IGLL3	-0.21984	0.00681	0.99958	-0.01340	0.86789	0.96381
CNTROB	-0.01846	0.00692	0.99958	0.00157	0.81744	0.94953
DNASE1L1	0.03945	0.00707	0.99958	-0.00385	0.79113	0.93985
OSBPL7	-0.04393	0.00714	0.99958	-0.00813	0.61578	0.86403
BCAP29	0.06982	0.00727	0.99958	-0.00162	0.95007	0.98625
RASL11A	-0.02564	0.00729	0.99958	-0.00397	0.67522	0.89209
ACOT2	0.11363	0.00730	0.99958	0.01087	0.79580	0.94170
MOCOS	0.09848	0.00766	0.99958	-0.00293	0.93627	0.98230
ZNF365	0.06548	0.00779	0.99958	0.03306	0.17630	0.57052
CDC42SE2	-0.06542	0.00794	0.99958	0.01487	0.54300	0.83205
MAP2K5	-0.02463	0.00794	0.99958	0.00876	0.34181	0.71631
TJP1	0.05510	0.00867	0.99958	0.00633	0.76105	0.93098
TBP	-0.03285	0.00882	0.99958	0.00836	0.50198	0.81447
SARS2	-0.05063	0.00895	0.99958	-0.00788	0.68163	0.89581

555

556

557 **References**

558 1. The Search for Causes of Breast Cancer. *The Lancet*. 1965;286(7415):725-726.
559 doi:10.1016/S0140-6736(65)90461-7

560 2. Bogen E. The Cause of Breast Cancer. *Am J Public Health Nations Health*. 1935;25(3):245-
561 250.

562 3. MacMahon B, Cole P, Lin TM, et al. Age at first birth and breast cancer risk. *Bull World*
563 *Health Organ*. 1970;43(2):209-221.

564 4. Lund E. Reproductive Histories and Premenopausal Breast Cancer: Different Estimates
565 Using Population, Neighborhood, or Hospital Controls.

566 5. Pike MC, Casagrande J, Smith PG. Statistical analysis of individually matched case-control
567 studies in epidemiology: factor under study a discrete variable taking multiple values. *Br J*
568 *Prev Soc Med*. 1975;29(3):196-201.

569 6. Beral V. Long term effects of childbearing on health. *J Epidemiol Community Health*.
570 1985;39(4):343-346.

571 7. Kvåle G, Heuch I, Eide GE. A prospective study of reproductive factors and breast cancer. I.
572 Parity. *Am J Epidemiol*. 1987;126(5):831-841. doi:10.1093/oxfordjournals.aje.a114720

573 8. Kvåle G, Heuch I. A prospective study of reproductive factors and breast cancer. II. Age at
574 first and last birth. *Am J Epidemiol*. 1987;126(5):842-850.
575 doi:10.1093/oxfordjournals.aje.a114721

576 9. Lund E. Childbearing in marriage and mortality from breast cancer in Norway. *Int J*
577 *Epidemiol*. 1990;19(3):527-531. doi:10.1093/ije/19.3.527

578 10. Albrektsen G, Heuch I, Tretli S, Kvåle G. Breast cancer incidence before age 55 in relation
579 to parity and age at first and last births: a prospective study of one million Norwegian
580 women. *Epidemiology*. 1994;5(6):604-611. doi:10.1097/00001648-199411000-00008

581 11. Albrektsen G, Heuch I, Kvåle G. The short-term and long-term effect of a pregnancy on
582 breast cancer risk: a prospective study of 802,457 parous Norwegian women. *Br J Cancer*.
583 1995;72(2):480-484. doi:10.1038/bjc.1995.359

584 12. Lambe M, Hsieh C cheng, Chan H wei, Ekbom A, Trichopoulos D, Adami HO. Parity, age at
585 first and last birth, and risk of breast cancer: A population-based study in Sweden. *Breast*
586 *Cancer Res Tr*. 1996;38(3):305-311. doi:10.1007/BF01806150

587 13. Colditz GA, Rosner B. Cumulative Risk of Breast Cancer to Age 70 Years According to Risk
588 Factor Status: Data from the Nurses' Health Study. *American Journal of Epidemiology*.
589 2000;152(10):950-964. doi:10.1093/aje/152.10.950

590 14. Ritte R, Tikk K, Lukanova A, et al. Reproductive factors and risk of hormone receptor
591 positive and negative breast cancer: a cohort study. *BMC Cancer*. 2013;13(1):584.
592 doi:10.1186/1471-2407-13-584

593 15. Lund E. Number of children and death from hormone-dependent cancers. *International*
594 *Journal of Cancer*. 1990;46(6):998-1000. doi:10.1002/ijc.2910460608

595 16. Hinkula M, Pukkala E, Kyyrönen P, Kauppila A. Grand multiparity and the risk of breast
596 cancer: population-based study in Finland. *Cancer Causes Control*. 2001;12(6):491-500.
597 doi:10.1023/A:1011253527605

598 17. Högnäs E, Kauppila A, Pukkala E, Tapanainen JS. Cancer risk in women with 10 or more
599 deliveries. *Obstet Gynecol*. 2014;123(4):811-816. doi:10.1097/AOG.0000000000000182

600 18. Paltiel O, Tajuddin SM, Polanker Y, et al. Grand multiparity and reproductive cancer in the
601 Jerusalem Perinatal Study Cohort. *Cancer Causes Control*. 2016;27(2):237-247.
602 doi:10.1007/s10552-015-0701-6

603 19. Katuwal S, Tapanainen JS, Pukkala E, Kauppila A. The effect of length of birth interval on
604 the risk of breast cancer by subtype in grand multiparous women. *BMC Cancer*.
605 2019;19(1):199. doi:10.1186/s12885-019-5404-z

606 20. Takeuchi T, Kitamura Y, Sobue T, et al. Impact of reproductive factors on breast cancer
607 incidence: Pooled analysis of nine cohort studies in Japan. *Cancer Medicine*.
608 2021;10(6):2153-2163. doi:10.1002/cam4.3752

609 21. Russo J, Balogh G, Russo IH. Breast cancer prevention. *Climacteric*. 2007;10(sup2):47-53.
610 doi:10.1080/13697130701606762

611 22. Russo J, Balogh GA, Russo IH, and the Fox Chase Cancer Center Hospital Network
612 Participants. Full-term Pregnancy Induces a Specific Genomic Signature in the Human
613 Breast. *Cancer Epidemiology, Biomarkers & Prevention*. 2008;17(1):51-66.
614 doi:10.1158/1055-9965.EPI-07-0678

615 23. Krum-Hansen S, Standahl Olsen K, Anderssen E, Frantzen JO, Lund E, Paulssen RH.
616 Associations of breast cancer related exposures and gene expression profiles in normal
617 breast tissue—The Norwegian Women and Cancer normal breast tissue study. *Cancer*
618 *Reports*. n/a(n/a):e1777. doi:10.1002/cnr2.1777

619 24. Lund E, Nakamura A, Snapkov I, et al. Each pregnancy linearly changes immune gene
620 expression in the blood of healthy women compared with breast cancer patients. *Clin*
621 *Epidemiol*. 2018;10:931-940. doi:10.2147/CLEP.S163208

622 25. Lund E, Dumeaux V, Braaten T, et al. Cohort profile: The Norwegian Women and Cancer
623 Study--NOWAC--Kvinner og kreft. *Int J Epidemiol*. 2008;37(1):36-41.
624 doi:10.1093/ije/dym137

625 26. Dumeaux V, Fjukstad B, Fjosne HE, et al. Interactions between the tumorand the blood
626 systemic response of breast cancer patients. PLoS Comput Biol 2017; 13: e1005680.
627 Doi:10.1371/journal.pcbi. 1005680.

628 27. Krum-Hansen S, Standahl Olsen K. Practical and Ethical Issues in Establishing a Collection
629 of Normal Breast Tissue Biopsies. Part of the NOWAC Post-Genome Cohort. *Advancing*
630 *Systems Epidemiology in Cancer*. Published online December 8, 2020:63-76.
631 doi:10.18261/9788215041193-2020-04

632 28. Bøvelstad HM, Holsbø E, Bongo LA, Lund E. A standard operating procedure for outlier
633 removal in large-sample epidemiological transcriptomics datasets. Published online May
634 31, 2017:144519. doi:10.1101/144519

635 29. Lin SM, Du P, Huber W, Kibbe WA. Model-based variance-stabilizing transformation for
636 Illumina microarray data. *Nucleic Acids Research*. 2008;36(2):e11.
637 doi:10.1093/nar/gkm1075

638 30. lumiHumanIDMapping. Bioconductor. Accessed February 28, 2023.
639 http://bioconductor.org/packages/lumiHumanIDMapping/

640 31. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*.
641 2000;406(6797):747-752. doi:10.1038/35021093

642 32. Parker JS, Mullins M, Cheang MCU, et al. Supervised Risk Predictor of Breast Cancer Based
643 on Intrinsic Subtypes. *JCO*. 2009;27(8):1160-1167. doi:10.1200/JCO.2008.18.1370

644 33. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential
645 Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular*
646 *Biology*. 2004;3(1). doi:10.2202/1544-6115.1027

647 34. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and
648 RNA-Seq data. *BMC Bioinformatics*. 2013;14(1):7. doi:10.1186/1471-2105-14-7

649 35. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-
650 based approach for interpreting genome-wide expression profiles. *Proceedings of the*
651 *National Academy of Sciences*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102

652 36. Rotunno M, Sun X, Fgueroa J, et al. Parity-related molecular signatures and breast cancer
653 subtypes by estrogen receptor status. Breast Cancer Res 2914; 16: R74. Doi:10.1186/bcr3689

654 37. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and
655 breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies
656 in 30 countries, including 50302 women with breast cancer and 96973 women without the
657 disease. *Lancet*. 2002;360(9328):187-195. doi:10.1016/S0140-6736(02)09454-0

658 38. Fortner RT, Sisti J, Chai B, et al. Parity, breastfeeding, and breast cancer risk by hormone
659 receptor status and molecular phenotype: results from the Nurses' Health Studies. *Breast*
660 *Cancer Research*. 2019;21(1):40. doi:10.1186/s13058-019-1119-y

661 39. Palmer JR, Viscidi E, Troester MA, et al. Parity, Lactation, and Breast Cancer Subtypes in
662 African American Women: Results from the AMBER Consortium. *JNCI: Journal of the*
663 *National Cancer Institute*. 2014;106(10):dju237. doi:10.1093/jnci/dju237

664 40. Islami F, Liu Y, Jemal A, et al. Breastfeeding and breast cancer risk by receptor status—a
665 systematic review and meta-analysis. *Ann Oncol*. 2015;26(12):2398-2407.
666 doi:10.1093/annonc/mdv379

667 41. Brown SB, Hankinson SE. Endogenous estrogens and the risk of breast, endometrial, and
668 ovarian cancers. *Steroids*. 2015;99:8-10. doi:10.1016/j.steroids.2014.12.013

669 42. The Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous Sex
670 Hormones and Breast Cancer in Postmenopausal Women: Reanalysis of Nine Prospective
671 Studies. *JNCI: Journal of the National Cancer Institute*. 2002;94(8):606-616.
672 doi:10.1093/jnci/94.8.606

673 43. Circulating sex hormones and breast cancer risk factors in postmenopausal women:
674 reanalysis of 13 studies. *Br J Cancer*. 2011;105(5):709-722. doi:10.1038/bjc.2011.254

675 44. International Agency for Research on Cancer, ed. *Hormonal Contraception and Post-*
676 *Menopausal Hormonal Therapy: This Publication Represents the Views and Expert Opinions*
677 *of an IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, Which Met*
678 *in Lyon, 2 - 9 June 1998*. IARC; 1999.

679 45. Dumeaux V, Fjukstad B, Fjosne HE, et al. Interactions between the tumor and the blood
680 systemic response of breast cancer patients. *PLOS Computational Biology*.
681 2017;13(9):e1005680. doi:10.1371/journal.pcbi.1005680

682 46. Lund E, Rasmussen Busund LT, Thalabard JC. Rethinking the carcinogenesis of breast
683 cancer: The theory of breast cancer as a child deficiency disease or a pseudo semi-
684 allograft. *Medical Hypotheses*. 2018;120:76-80. doi:10.1016/j.mehy.2018.08.015

685 47. Li X, Zhou J, Fang M, Yu B. Pregnancy immune tolerance at the maternal-fetal interface.
686 *International Reviews of Immunology*. 2020;39(6):247-263.
687 doi:10.1080/08830185.2020.1777292

688 48. Parasar P, Guru N, Nayak NR. Contribution of macrophages to fetomaternal immunological
689 tolerance. *Human Immunology*. 2021;82(5):325-331. doi:10.1016/j.humimm.2021.02.013

690 49. Orefice R. Immunology and the immunological response in pregnancy. *Best Practice &*
691 *Research Clinical Obstetrics & Gynaecology*. 2021;76:3-12.
692 doi:10.1016/j.bpobgyn.2020.07.013

693 50. Lund E, Kumle M, Braaten T, et al. External validity in a population-based national
694 prospective study--the Norwegian Women and Cancer Study (NOWAC). *Cancer Causes*
695 *Control*. 2003;14(10):1001-1008. doi:10.1023/b:caco.0000007982.18311.2e

696 51. Pu M, Messer K, Davies SR, et al. Research-based PAM50 signature and long-term breast
697 cancer survival. *Breast Cancer Res Treat*. 2020;179(1):197-206. doi:10.1007/s10549-019-
698 05446-y

699 52. Lien TG, Ohnstad HO, Lingjærde OC, et al. Sample Preparation Approach Influences PAM50
700 Risk of Recurrence Score in Early Breast Cancer. *Cancers*. 2021;13(23):6118.
701 doi:10.3390/cancers13236118

702

703

704