Doctoral theses at NTNU, 2023:30

Elise Klæstad

Proliferation in molecular subtypes of breast cancer

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

NTNU
Norwegian University of
Science and Technology

Elise Klæstad

Proliferation in molecular subtypes of breast cancer

Thesis for the Degree of Philosophiae Doctor

Trondheim, January 2023

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

ISBN 978-82-326-6026-1 (printed ver.)
ISBN 978-82-326-6155-8 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:30

Printed by NTNU Grafisk senter

Proliferasjon i molekylære subtyper av brystkreft

Brystkreft er en heterogen sykdom og kan ved hjelp av molekylære markører bli delt inn i forskjellige undergrupper med ulik prognose og respons på behandling. Økt proliferasjon (cellevekst) har blitt identifisert som et av kjennetegnene på kreft, og er assosiert med dårligere prognose i brystkreft. Et av målene til Breast Cancer Subtypes-prosjektet er å studere faktorer som kan ha sammenheng med de spesifikke subtypene av brystkreft og prognose hos brystkreftpasienter. Brystkreftforskning avdekker stadig nye gener assosiert med proliferasjon, noe som kan bidra til en mer presis inndeling av pasienter i prognostiske grupper og på sikt mer målrettet behandling av brystkreftpasienter. To gener som er vist å være assosiert med proliferasjon i brystkreft er *MRPS23* og *ZNF703.*

Denne oppgaven bygger på fire kohorter. Kohort 1, 3 og 4 består av kvinner fra Trøndelag som ble fulgt opp for påvisning av brystkreft fra 1961 og frem til 2012. Kohort 1 består av 25 727 kvinner fra Nord-Trøndelag hvorav 1379 ble diagnostisert med brystkreft i løpet av oppfølgingstiden. Kohort 3 består av 34 221 kvinner fra Nord-Trøndelag hvor av 731 ble diagnostisert med brystkreft, og kohort 4 består av 23 350 kvinner født ved E.C. Dahls stiftelse, hvorav 885 ble diagnostisert med brystkreft. Tumorvev er tilgjengelig for de fleste av brystkreftpasientene som ble diagnostisert ved St. Olavs Hospital. I tillegg til tumorvev har vi tilgang på ytterligere informasjon om kvinnene og deres svulster, samt noe informasjon om de friske kvinnene som utgjør bakgrunnsbefolkningen. Brystkrefttumorene i kohort 1, 3 og 4 har tidligere blitt inndelt i molekylære subtyper ved hjelp av immunhistokjemi og *in situ* hybridisering. Kohort 2 består av genekspresjonsdata fra ca. 2000 brystkreftsvulster fra datasettet METABRIC.

I den første studien undersøkte vi *MRPS23* kopitall ved hjelp av fluorescens *in situ* hybridisering (FISH), og sammenhengen med molekylær subtype, proliferasjon og prognose. I tillegg undersøkte vi genuttrykk av *MRPS23* og sammenheng med molekylær subtype og prognose. Vi fant at amplifikasjon av *MRPS23* var assosiert med økt proliferasjon. Vi identifiserte kopitallsøkning i alle subtyper utenom 5-negativ subtype. Vi fant ingen sikker sammenheng mellom *MRPS23* kopitall og prognose. Høyt genuttrykk av *MRPS23* var assosiert med Luminal B subtype. Det var ingen assosiasjon mellom genuttrykk og prognose.

I den andre studien brukte vi FISH og immunhistokjemi til å studere *ZNF703* kopitall og ZNF703 proteinuttrykk i brystkreft, og undersøke sammenhengen med molekylær subtype, proliferasjon og prognose. Vi fant at høyt kopitall var assosiert med Luminal B subtypene. I tillegg fant vi en sammenheng mellom høyt kopitall og høy proliferasjon, høy histologisk grad og dårlig prognose. Ved å se på Luminal A tumorer separat fant vi at tumorer med høyt

kopitall hadde høyere histologisk grad og dårligere prognose sammenlignet med tumorer uten kopitallsøkning. Vi fant en sammenheng mellom kopitallsøkning og høyt proteinuttrykk av ZNF703, men ingen sammenheng mellom proteinuttrykk og prognose.

I den tredje studien brukte vi Ki-67 og mitosetall som markører for proliferasjon for å studere endringer i insidensrate av høy- og lavproliferative brystkreftsvulster over tid. Vi gjorde analyser på alle brystkreftsvulstene samlet og for noen molekylære subtyper separat. Vi manglet informasjon om proliferasjonsstatus for noen av svulstene, og brukte multippel imputasjon for å estimere manglende verdier for Ki-67 og mitosetall. Vi sammenlignet kvinner født i 1929 eller senere med kvinner født før 1929. Ved å bruke Ki-67 som proliferasjonsmarkør fant vi at det hadde vært en økning av både høy- og lavproliferative svulster blant kvinnene som var født etter 1929. Ved å se på mitosetall fant vi kun økning i lavproliferative svulster. Vi gjorde separate insidensanalyser for svulster av $HER2^+$ og trippel negativ subtype. For $HER2^+$ svulster fant vi at det hadde vært en økning av høyproliferative svulster i henhold til Ki-67 status, mens det i henhold til mitosetall hadde vært en økning av lavproliferative svulster blant kvinnene født i 1929 eller senere. For trippel negative svulster fant vi ingen endring i insidensrate av verken høy- eller lavproliferative brystkreftsvulster.

Funnene fra disse studiene bidrar til ny informasjon om proliferasjonsassosierte gener i molekylære subtyper av brystkreft, og utforsker tidstrender i insidens av brystkreft relatert til proliferasjon.

Navn kandidat: Elise Klæstad
Institutt: Institutt for klinisk og molekylær medisin
Veiledere: Marit Valla og Anna Bofin
Finansieringskilde: Norges teknisk-naturvitenskapelige universitet (NTNU)

Ovennevnte avhandling er funnet verdig til å forsvares offentlig
for graden ph.d. i medisin og helsevitenskap.
Disputas finner sted i auditorium LA21, Laboratoriesenteret.
Fredag 27. januar 2022, kl. 12.15

Table of Contents

Acknowledgements

This thesis was carried out at the Department of Clinical and Molecular Medicine (IKOM) at the Norwegian University of Science and Technology (NTNU) from 2018 to 2022. I am grateful to the Medical Research Program at NTNU for providing the opportunity to attend the Medical Student Research Program (Forskerlinjen), and subsequently the PhD programme at the Faculty of Medicine and Health Sciences. Financial support was provided from NTNU.

First and foremost, I would like to give my sincere thanks to my main supervisor, associate professor of pathology and consultant pathologist, Marit Valla. She has provided excellent guidance, support, and shared her knowledge about breast cancer and the scientific process. She has made me appreciate the writing process, taught me how to make my texts more comprehensible to the reader and to appreciate constructive feedback. In addition, she continuously inspires me with her accomplishments and work ethics.

Furthermore, I would like to thank my co-supervisor, professor of pathology and PI of the Breast Cancer Subtypes Research Group, Anna Mary Bofin, for sharing her knowledge in the field of breast cancer. She has made me question the common truths and see how our research fits into the bigger picture.

I wish to thank associate professor at the Department of Public Health and Nursing, Signe Opdahl, for her assistance with the statistical analyses of this project. Statistical analyses can be difficult to interpret and perform, and she has a brilliant way of explaining complex analyses.

Thank you to associate professor of breast- and endocrine surgery, Monica Engstrøm, for providing valuable feedback in the writing process and assisting during the many hours of FISH analyses. I would also like to thank associate professor Elisabeth Wik at the Department of Clinical Medicine, University of Bergen, for her cooperation, and for sharing her knowledge about the METABRIC data set. Thank you to specialist in pathology and pharmacology, Joanna Ewa Sawicka, for valuable contributions to paper II, and to Sunil Xavier Raj, chief physician at the Cancer Clinic, St Olav's Hospital, for clinical input on paper III.

In addition, I would like to thank senior biomedical engineer Borgny Ytterhus for showing me the laboratory methods of TMA construction, immunostaining and fluorescence *in situ* hybridization and for always providing laboratory work of the highest quality.

This work has been a part of the Breast Cancer Subtypes Research group, which provides extensive and important research on breast cancer. I am thankful that I have had the

opportunity to be a part of such an inspiring environment.

Last but not least, I would like to thank my family and friends for endless support during this project. A huge thank you to my friends for motivation and fun distractions, and for making my years in Trondheim wonderful. I am ever grateful to my parents for their endless love and support. Thank you to my brother, Ole Petter, for reminding me to always look on the bright side. In addition, my appreciation goes to my boyfriend, Bård, for his everlasting patience and for always believing in me.

Thank you all.

Abbreviations

5NP	5 Negative phenotype
AJCC	American Joint Commission of Cancer
BC	Breast cancer
BCI	Breast Cancer Index
BCS	Breast Cancer Subtypes
BL	Basal like
BP	Basal phenotype
CEP	Centromere enumeration probe
CI	Confidence interval
CISH	Chromogenic *in situ* hybridization
CK5	Cytokeratin 5
EBC	Early breast cancer
EGFR	Epidermal growth factor 1
EP	EndoPredict
ER	Oestrogen receptor
FDA	Food and Drug administration
FFPE	Formalin-fixed, paraffin-embedded
FGD5	Faciogenital dysplasia 5
FISH	Fluorescence *in situ* hybridization
HE	Haematoxylin eosin
HER2	Human epidermal growth factor receptor 2
HES	Haematoxylin-erythrosine-saffron
HPF	High power field
HR	Hazard ratio
IHC	Immunohistochemistry
IHC4	Immunohistochemical 4
ISH	*In situ* hybridization
KM	Kaplan Meier
LAR	Luminal androgen receptor
MHT	Menopausal hormone therapy
MRPS23	Mitochondrial Ribosomal Protein S23
PCNA	Proliferating Cell Nuclear Antigen

PD-L1	Programmed death-ligand 1
PhH3	Phosphorylated histone H3
PR	Progesterone receptor
SLN	Sentinel lymph node
TCGA	The Cancer Genome Atlas
TMA	Tissue microarray
TNBC	Triple negative breast cancer
WHO	World Health Organization
ZNF703	Zink finger protein 703

List of publications

This thesis is based on the following publications

Paper I:

Klæstad E, Opdahl S, Engstrøm MJ, Ytterhus B, Wik E, Bofin AM, Valla M. ***MRPS23*** **amplification and gene expression in breast cancer; association with proliferation and the non-basal subtypes.** Breast Cancer Res Treat. 2020 Feb;180(1):73-86.

Paper II:

Klæstad E, Sawicka JE, Engstrøm MJ, Ytterhus B, Valla M, Bofin AM. ***ZNF703*** **gene copy number and protein expression in breast cancer; associations with proliferation, prognosis and luminal subtypes.** Breast Cancer Res Treat. 2021 Feb;186(1):65-77.

Paper III:

Klæstad E, Opdahl S, Raj SX, Bofin AM, Valla M. **Long term trends of breast cancer incidence according to proliferation status.** Included in the thesis as a manuscript. Accepted for publication in BMC Cancer, Dec 2022.

Summary

Breast cancer is a heterogeneous disease. Using immunohistochemistry (IHC) and *in situ* hybridization (ISH) breast cancers can be divided into molecular subtypes with different prognosis and response to treatment. Increased proliferation has been identified as one of the hallmarks of cancer and it is associated with poor prognosis in breast cancer. One of the goals of the Breast Cancer Subtypes project is to identify new prognostic factors and their relation to specific molecular subtypes of breast cancer. Breast cancer research has revealed new genes associated with proliferation. These can contribute to improved prognostication and potentially more targeted treatment of breast cancer patients. Two genes associated with proliferation in breast cancer are *MRPS23* and *ZNF703*.

This study thesis is based on four cohorts. Cohort 1, 3 and 4 comprise women from the county Trøndelag that were followed for breast cancer occurrence from 1961 to 2012. Cohort 1 comprises 25 727 women from Nord Trøndelag, of whom 1379 were diagnosed with breast cancer during follow up. Cohort 3 comprises 34 221 women from Nord Trøndelag of whom 731 were diagnosed with breast cancer, and cohort 4 comprises 23 350 women born at E.C Dahls foundation, of whom 885 were diagnosed with breast cancer. Tumour tissue is available for most of the breast cancer patients that were diagnosed at St. Olav's hospital. In addition to tumour tissue, we have access to information about the patients and their tumours, and some information about the healthy background population. The breast cancer tumours included in cohorts 1, 3 and 4 have previously been reclassified into molecular subtypes using IHC and *in situ* hybridization (ISH). Cohort 2 comprises gene expression data from approximately 2000 breast cancer tumours from the METABRIC dataset.

In the first study, we examined *MRPS23* copy number using fluorescence *in situ* hybridization (FISH), and studied associations with molecular subtype, proliferation, and prognosis. In addition, we examined gene expression of *MRPS23* and associations with molecular subtype and prognosis. We found that amplification of *MRPS23* was associated with increased proliferation. We identified copy number increase in all subtypes except 5-negative phenotype. We found no clear association between *MRPS23* copy number and prognosis. High gene expression of *MRPS23* was associated with the luminal B subtype. There was no association between gene expression and prognosis.

In the second study, we used FISH and IHC to study copy number and protein expression of *ZNF703* in breast cancer, and associations with molecular subtype, proliferation and prognosis. We found that high copy number was associated with the luminal B subtypes. In addition, we found an association between high copy number and increased proliferation,

higher histological grade and poorer prognosis. Furthermore, on looking at luminal A tumours separately, we found that tumours with high copy number had higher grade and poorer prognosis compared to tumours without copy number increase. We found a correlation between high copy number and high protein expression, but no correlation between protein expression and prognosis.

In the third study we used Ki-67 and mitotic count as markers of proliferation to study incidence trends of high- and low-proliferative breast cancer tumours. We performed analyses on all breast cancer tumours combined, and separately for subgroup of molecular subtypes. Multiple imputation was used to estimate missing values of Ki-67 and mitotic count in cases with missing values. We compared incidence in women born in 1929 or later to women born before 1929. According to Ki-67, we found an increase in both high- and low-proliferative tumours among women born in 1929 or later. According to mitotic count, we only found an increase in low-proliferative tumours. We did separate analyses of $HER2^+$ and triple negative tumours. In $HER2^+$ tumours we found that there had been an increase of high-proliferative tumours according to Ki-67, while it according to mitotic count had been an increase of low-proliferative tumours among women born in 1929 or later. There was no change in incidence trends according to proliferation status for triple negative tumours.

Introduction

1.1 Breast cancer epidemiology

Breast cancer is the most common cancer among women worldwide, with an estimated 2.26 million new cases in 2020 (Figure 1) (1). It is also the most common cancer among Norwegian women, comprising 23% of all incident cancers in 2021 (2). It is the most common cause of cancer related death among women in developing countries, and the second most common cause of cancer death in developed countries (3). During the last five decades there has been an increase in breast cancer incidence in most high income countries (3-5), and, more recently, also in low income countries (3).

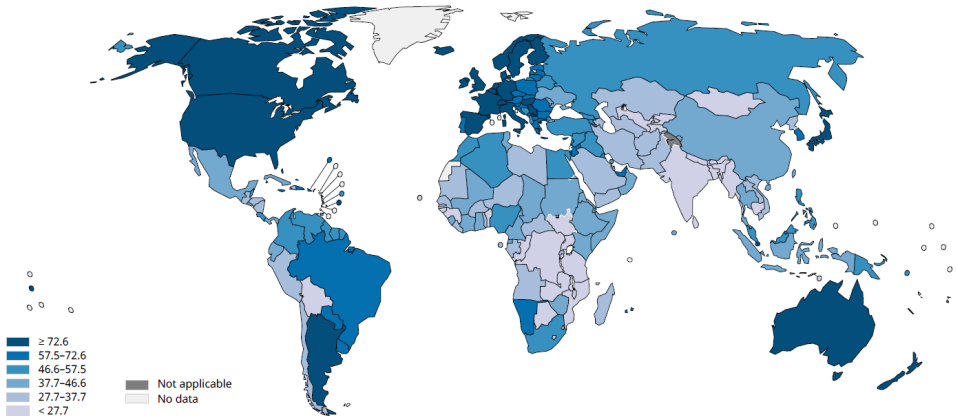


Figure 1: Global breast cancer incidence rates per 100 000 person years in women, 2020. Reprint from Global Cancer Observatory: Cancer Today, *Fact Sheet - Breast*, Available from: https://gco.iarc.fr/today, Copyright (2022) (1).

Since the 1990s there has been a decrease in mortality rates in developed countries (4, 6), whilst rates are still increasing in developing countries (3). Countries in central Africa have the highest mortality rate (Figure 2).

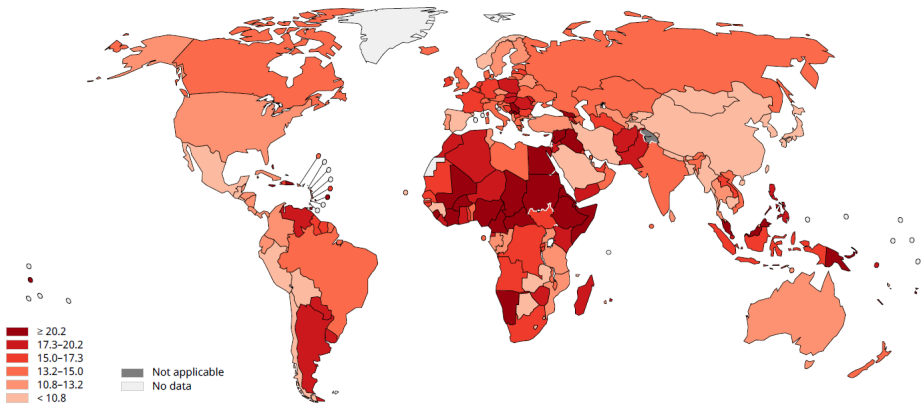


Figure 2: Global mortality rates per 100 000 person years for breast cancer in women, 2020. Reprint from Global Cancer Observatory: Cancer Today, *Fact Sheet - Breast*, Available from: https://gco.iarc.fr/today, Copyright (2022) (1).

In 2021, 3991 new breast cancer cases were registered among Norwegian women, and the cumulative risk of developing breast cancer by the age of 80 was 10.5% (2). The median age at diagnosis is 62 years old (2).

Breast cancer survival among Norwegian women has increased steadily since 1965 (Figure 3), and between 2017-2021 five-year relative survival for breast cancer patients was 92.3% (2). In general, survival decreases with age (Figure 4). Relative survival compares the observed survival in a group of patients with a specific disease to the expected survival of a comparable group in the general population during a given period of time (2). Thus, relative survival shows whether the disease shortens life. As a result of increased incidence and improved survival there is a growing pool of breast cancer survivors in Norway. In December 2021, there were 54 827 prevalent cases of breast cancer in Norway, and 25 409 of these were diagnosed >10 years previously (2). Some of these women may experience complications following breast cancer treatment (7).

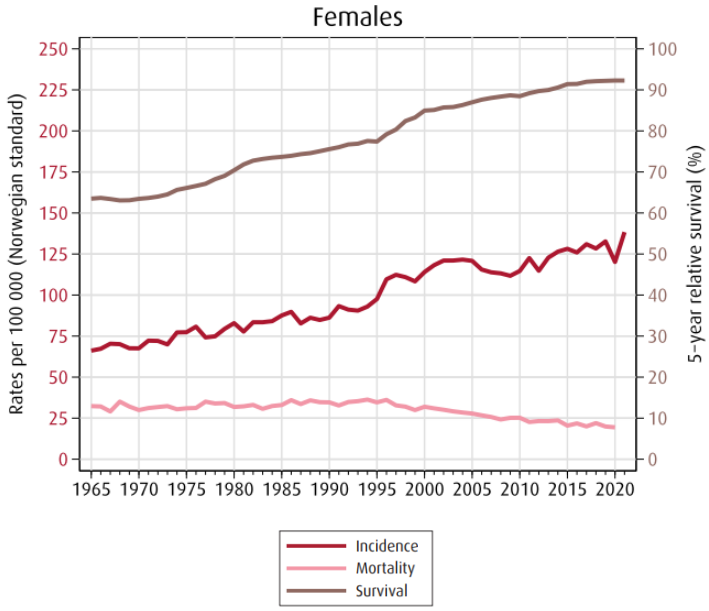


Figure 3: Trends in breast cancer incidence (red) and mortality (pink) rates, and five-year relative survival proportions (brown) among Norwegian women. Reprint from the Cancer Registry of Norway, *Cancer in Norway 2021 - Cancer incidence, mortality, survival and prevalence in Norway (2)*. Page no. 108, Figure 9.1-M, with permission from the Cancer registry of Norway.

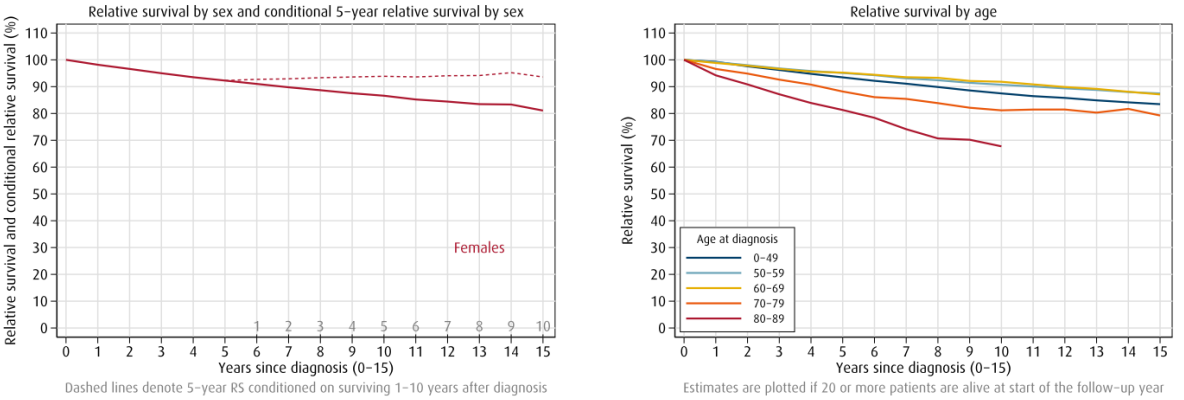


Figure 4: Relative survival from breast cancer up to 15 years after diagnosis, overall (to the left) and by age categories (to the right). Reprint from the Cancer Registry of Norway, *Cancer in Norway 2021 - Cancer incidence, mortality, survival and prevalence in Norway* (2). Page no. 96, Figure 8.1-L, with permission from the Cancer registry of Norway.

1.1.1 Time trends in breast cancer incidence

Globally, breast cancer incidence rates have been rising in many high-income countries the last five decades, and more recently also in lower-income countries (3). In Norway, there has been a considerable increase in breast cancer incidence since the establishment of the

Norwegian Cancer Registry in the early 1950s (Figure 3) (8). The increase in the mid-90s and early 2000s was primarily limited to women aged 50-69 years (Figure 5) (8). Previous studies have found an association between the mid-1990s increase in breast cancer incidence and mammography screening and menopausal hormone therapy (MHT) in Norway (9, 10), and internationally (11, 12)

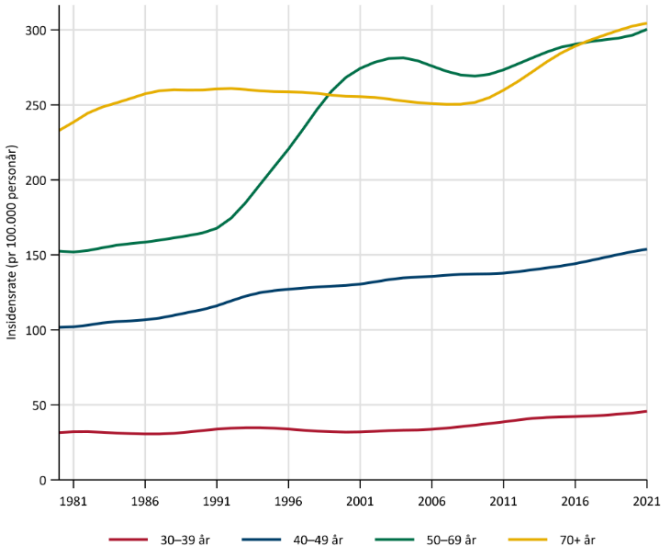


Figure 5: Incidence rates for breast cancer by age at, and year of diagnosis among Norwegian women. Reprint from the National Quality Register for Breast Cancer, *Annual Report 2021 with results and improvent measures from the National Quailty Register for Breast Cancer* (13), page 18, Figure 3.1, with permission from the Cancer Registry of Norway.

Several studies have revealed that breast cancer incidence trends vary according to molecular markers. The recent increase in breast cancer incidence is largely restricted to hormone receptor positive tumours (14, 15). Studies of breast cancer incidence trends through recent decades according to hormone receptor status in the US and several European countries have found increasing incidence rates of oestrogen receptor $(ER)^+$ cancers and decreasing rates of ER^- cancers (14-17). This could partly be explained by the implementation of mammography screening. It has been shown that breast cancers detected through mammography screening have favourable characteristics such as lower stage, lower grade and smaller size compared to interval cancers and cancers diagnosed outside of screening (18-20). Additionally, screen-detected cancers are more often ER^+ (19, 21) and luminal A subtype (18, 19, 21) compared to tumours detected outside of screening. Other than mammography screening, the change in

incidence trends may partly be attributed to changing prevalence and distribution of breast cancer risk factors.

1.1.2 Risk factors associated with breast cancer

Several factors associated with increased breast cancer risk have been identified, such as family history, lifestyle related factors and reproductive factors (22, 23). Female gender, high age, early menarche, late menopause, high age at first pregnancy and low parity are associated with increased risk of breast cancer (23, 24). Steroids are a known risk factor for breast cancer, and use of menopausal hormone therapy (MHT) has been identified as a major cause of the increase in breast cancer incidence in the 90s and early 2000s (9, 10, 12, 25). In addition, it has been estimated that approximately 20% of breast cancer related deaths worldwide are caused by modifiable factors such as alcohol consumption, overweight and obesity, and physical inactivity (26).

Furthermore, family history of breast cancer is associated with increased risk (23, 27), and it has been estimated that nearly a quarter of all breast cancers are related to heritable traits (23). Hereditary breast cancers can be due to both high-penetrance and low-penetrance genes, of which mutations in the *BRCA1* and *BRCA2* genes are among the high-penetrance genes, contributing to a significant increase in the risk of developing breast cancer (23, 28).

Several studies have investigated the association between specific risk factors and molecular subtypes of breast cancer. The results have been inconsistent (29, 30) and different definitions of molecular subtypes further complicate comparison and interpretation of results. In general, there is a stronger association between reproductive risk factors and hormone receptor positive (luminal) tumours than triple negative breast cancers (TNBC; negative for ER, progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2)) (27, 29-31). A pooled study of 38 articles investigating associations between risk factors and breast cancer subtypes found that most established risk factors were associated with the luminal A subtype. For instance, young age at menarche, low parity, old age at menopause, and MHT use were all associated with increased risk of luminal A breast cancer. Associations with risk factors for the other subtypes were less consistent. However, there was a possible association between age at menarche and the luminal B subtype. Young age at menarche, use of oral contraceptives and high parity was associated with TNBC. Family history of breast cancer was a common risk factor for all molecular subtypes (30). Studies have shown that the risk effect of overweight varies according to menopausal status (30, 32, 33). In pre-menopausal women there appears to be an association between obesity and increased risk of

TNBC (30, 32, 33), while in post-menopausal women obesity is associated with increased risk of HR^+ tumours (33). In addition, the time of weight gain may affect the risk of breast cancer (34).

1.2 Breast cancer diagnostics

Patients who present with symptoms or findings suspicious of breast cancer qualify for further examination. These women are referred to a Breast Diagnostic Centre for "triple diagnostics" comprising a clinical examination, imaging and a preoperative biopsy (35). Once a patient is diagnosed with breast cancer, further examinations include clinical/radiological staging, and histopathological diagnostics including assessment of biomarkers in the preoperative biopsy and surgical specimen.

1.2.1 The Norwegian mammography screening programme

The Norwegian mammography screening program started as a pilot project in four counties in 1995/96. Consecutively, more counties were included, and the counties North and South Trøndelag were included in 2001 (36). Since 2004, all Norwegian women aged 50-69 years have been invited to attend a mammography screening every second year (37). The aim of the program is to detect breast cancers at an earlier stage to improve prognosis (36). Yearly attendance is approximately 75% (36). Among women attending the program most tumours are screen-detected, and approximately 25% are interval cancers (36). A recent study of the effect of population-based breast cancer screening in Norway found that there has been a 20% reduction of breast cancer mortality due to screening and a 23% reduction of breast cancer mortality due to improved treatment after implementation of the program (38). These numbers correspond to a previous evaluation of the national screening program estimating that it attributes to a reduction in breast cancer mortality of 20-30% (36). However, the screening program raises issues of overdiagnosis and overtreatment (37, 39) and length time and lead time biases may complicate the estimation of screening effectiveness (40).

1.2.2 Breast cancer diagnostics after malignancy has been established

1.2.2.1 Staging

The TNM-classification system is the system used for breast cancer staging. It takes into consideration the extent of the tumour (T) at the primary site, affected lymph nodes (N) and distant metastases (M) (35, 41-43). Staging is assessed both clinically and by

histopathological examination. Clinical staging is done before treatment is given, and it is based on findings in the clinical examination, imaging, and preoperative biopsy. Pathological staging is based on microscopic examination of the surgical specimen and lymph node metastases. Based on the findings each case is staged from I to IV. Tumour staging is important for the further management of the patient (42, 43). In the most recent edition of the TNM classification of the American Joint Commission of Cancer (AJCC), prognostic stage groups were introduced in addition to the traditional anatomic stage groups (43). The prognostic stage groups incorporate biological markers (ER, PR and HER2), histological grade and gene expression assays into the current staging system in order to give improved prognostic information (43, 44).

1.2.3 Histopathological diagnostics

The final diagnosis from the pathologist includes the type of specimen, histological subtype, histological grade, mitotic count, tumour size including the extent of invasive growth, invasion into surrounding structures (vessels, skin/papillae, muscle, nerves), resection margins, lymph node status and biomarker status (35).

1.2.3.1 Histological subtype

Breast cancer tumours are classified into histological subtypes according to WHO guidelines based on their microscopic appearance (35, 42). Most tumours (70-80%) are classified as "invasive carcinoma of no special type". This is a heterogenous group of tumours with characteristics that fail to demonstrate features compatible with one of the specific histological subtypes. The second most common type is invasive lobular carcinoma, which accounts for approximately 5-15% of breast cancers. The WHO classification system also includes other, less common subtypes such as tubular carcinoma, mucinous carcinoma, and adenoid cystic carcinoma (Figure 6) (42).

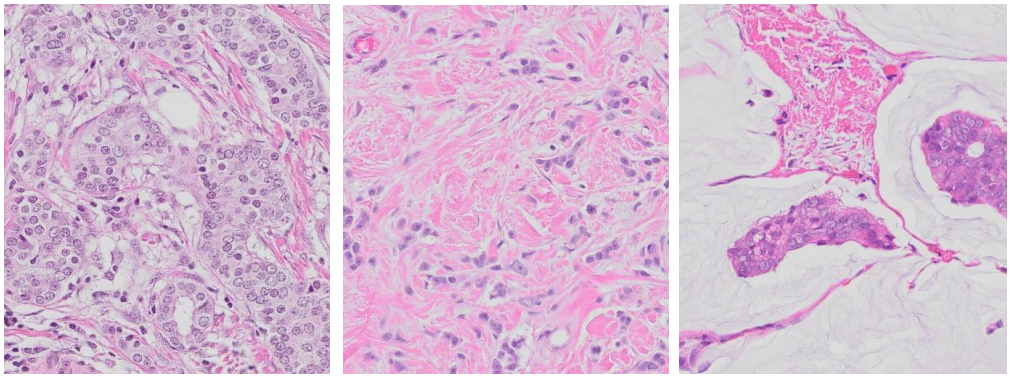


Figure 6: Histological subtypes of breast cancer. Left: Invasive carcinoma of no special type. Middle: Lobular carcinoma. Right: Mucinous carcinoma. HE-staining. Slides were scanned at 40x magnification. Photo: Marit Valla, NTNU.

1.2.3.2 Histological grade

All invasive carcinomas are histologically graded according to the WHO guidelines, which are based on the Nottingham criteria defined by Elston and Ellis in 1991 (42, 45). Grading is usually done on haematoxylin-eosin (HE)-stained whole-sections. Three morphological features are assessed: the proportion of tubule formation, the degree of nuclear pleomorphism and mitotic count (Table 1) (42). Tubule formation is determined by the proportion of the epithelial component of the tumour displaying tubular structures. Nuclear pleomorphism is assessed by evaluating the discrepancy between normal breast epithelial cells and tumour cells in addition to assessment of the variation between the tumour cells. Mitotic count is assessed in the area with the highest number of mitoses (the hotspot method). The mitotic count is the total number of mitoses recorded in 10 high-power fields (HPF) (42). To account for varying sizes of the high-power fields in different microscopes, mitoses/10 HPF is converted to mitoses/mm^2 (Table 2) (42, 46).

Table 1: Method for assessing histological grade in breast tumours

Tubule formation	**Score**
Majority of tumour (>75%)	1
Moderate degree (10-75%)	2
Little or none (<10%)	3
Nuclear pleomorphism	
Small, regular, uniform cells	1
Moderate increase in size and variability	2
Marked variation	3

Mitotic count	
Dependent on microscope field area	1-3

Reprinted from WHO Classification of Tumours Editorial Board. Breast Tumours, 5th edition/vol. 2, Introduction to tumours of the breast, Page No. 87, Copyright (2022).

Table 2: Score thresholds for mitotic counts based on the diameter of the high-power field and its corresponding area

Field diameter	**Field area**	**Mitotic count (score)**		
(mm)	**(mm^2)**	**1**	**2**	**3**
0.50	0.196	≤7	8-14	≥15
0.51	0.204	≤7	8-14	≥15
0.52	0.212	≤7	8-15	≥16
0.53	0.221	≤8	9-16	≥17

Reprinted from WHO Classification of Tumours Editorial Board. Breast Tumours, 5th edition/vol. 2, Introduction to tumours of the breast, Page No. 6, Copyright (2022).

Each characteristic is graded from 1-3 and the total score is thus 3-9. Total score 3 - 5 corresponds to grade 1 (highly differentiated tumours), 6-7 points to grade 2 (moderately differentiated tumours), and 8-9 points to grade 3 (poorly differentiated tumours) (42). Histological grade is strongly associated with prognosis in breast cancer (45, 47).

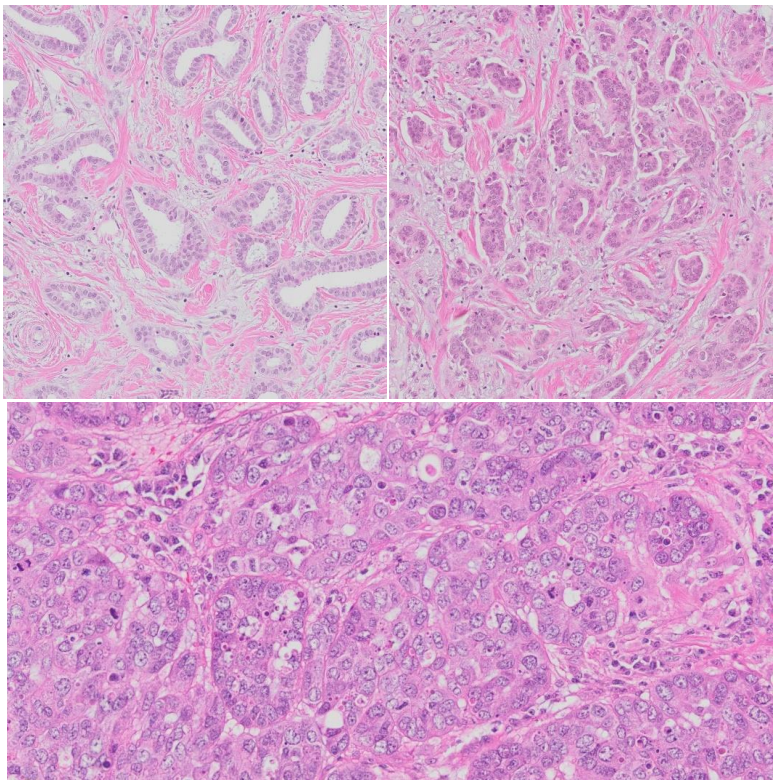


Figure 7: Histopathological grade in breast cancer. Top left: Grade 1. Top right: Grade 2. Bottom: Grade 3. Slides were HE-stained and scanned at 40x magnification. Photo: Marit Valla, NTNU

1.2.3.3 Prognostic and predictive markers

A prognostic marker is a clinical or biological characteristic that gives information about the most likely outcome in an untreated individual (48). Prognostic markers can be assessed by examining clinical outcomes, such as overall, breast cancer specific or recurrence-free/disease-free survival. A predictive marker is a clinical or biological characteristic that gives information about the expected benefit of treatment (48). In breast cancer, tumours are routinely analysed for biomarkers such as ER, PR, HER2 and the proliferation marker Ki-67 (35, 42, 49). In addition, *programmed death-ligand 1* (PD-L1) and gene expression assays can provide useful information in certain cases.

1.2.3.3.1 Hormone receptor status

ER and PR are hormone receptors present in normal breast tissue. ER is expressed in 7% of epithelial cells, in ducts and lobules of the normal breast (50). PR is expressed in 10-20% of luminal epithelial cells. Both ER and PR are prognostic and predictive markers, and can be

used to predict which patients will respond to endocrine therapy, such as Tamoxifen which inhibits oestrogen binding to ER, or aromatase inhibitors which decrease the level of circulating oestrogens (51). Hormone receptor status is evaluated using immunohistochemistry (IHC) in all invasive carcinomas and recurrent breast cancers. Cut-off for ER positivity is ≥1% positive nuclei (35, 49, 52). According to updated ASCO/CAP 2020 guidelines, cases with positive ER staining in 1-10% of tumour cell nuclei should be reported as ER low positive and supplemented with a comment on the percentage of cell staining, intensity, and status on IHC controls (52). Recent Norwegian guidelines state that staining for ER/PR should be re-assessed on the surgical specimen if the preoperative biopsy shows ER from 1-9% (35). American guidelines have a cut-off for PR positivity at 1% (52). According to Norwegian guidelines, a tumour is considered PR positive if >10% of tumour nuclei have positive staining (35). Staining intensity is not considered in either of the hormone receptor markers.

1.2.3.3.2 HER2 status

The *HER2* gene is located on chromosome 17q12 (53). It encodes for a transmembrane growth factor receptor which controls normal cell growth and differentiation (54). The *HER2* gene is overexpressed in 13-15% of breast cancers (22), and it is associated with a poor clinical outcome (54, 55). HER2 is a target for several treatments, for instance the recombinant human anti-HER2 monoclonal antibody (rhuMAb-HER2, trastuzumab) (51). HER2-status is assessed in all invasive breast cancers and metastatic lesions, using IHC and *in situ* hybridization (ISH) according to the algorithm presented in Figure 8.

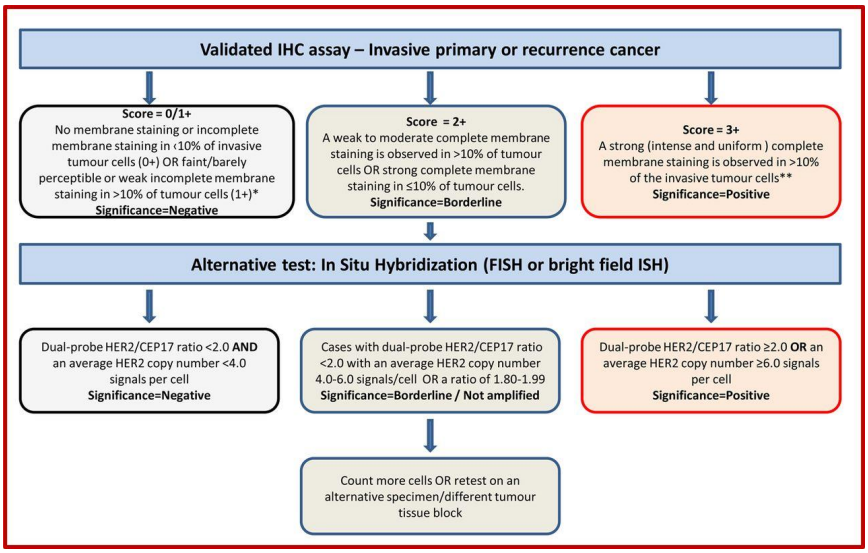


Figure 8: Scoring algorithm for HER2 using IHC and ISH. Reprinted from *Journal of Clinical Pathology* 2015; Rakha EA *et al.* On behalf of the National Coordinating Committee for Breast Pathology, Updated UK Recommendations for HER2 assessment in breast cancer (56), Page No. 93-99. Figure 1. Paper distributed under a Creative Commons license (CC BY-NC 4.0) and can be used for non-commercial purposes.

A tumour is HER2 positive if more than 10% of tumour cells show intense, complete circumferential membrane staining with IHC (35, 56, 57). Using dual-probe ISH, *HER2* status is defined by *HER2*/CEP17-ratio and average *HER2* gene copies per tumour cell. According to Norwegian and British guidelines tumours with *HER2*/CEP17 ratio ≥ 2 and/or average *HER2* copy number ≥6 are classified as HER2 positive (35, 56). In 2018, the American Guidelines were updated with a more complex algorithm for the scoring of HER2 ISH, dividing cases into 5 groups based on *HER2* copy number and *HER2*/CEP17 ratio when dual probe assays are used. The main difference from previous guidelines, is the additional work-up required for cases defined as HER2 ISH group 2-4 (57). The algorithm for HER2 scoring according to American guidelines is presented in Figure 9.

Summary of HER2 ISH Diagnostic Criteria	
HER2 Positive	**HER2 Negative**
Dual Probe Assay	
Group 1	Group 2 AND concurrent IHC 0-1+ or 2+
Group 2 AND concurrent IHC 3+	Group 3 AND concurrent IHC 0-1+
Group 3 AND concurrent IHC 2+ or 3+	Group 4 AND concurrent IHC 0-1+ or 2+
Group 4 AND concurrent IHC 3+	Group 5
Single Probe Assay	
HER2 copy number ≥6.0 signals/cell	HER2 copy number <4.0 signals/cell
HER2 copy number ≥4.0 and <6.0 signals/cell AND concurrent IHC 3+	HER2 copy number ≥4.0 and <6.0 signals/cell AND concurrent IHC 0 or 1+
HER2 copy number ≥4.0 and <6.0 signals/cell AND concurrent dual probe Group 1	HER2 copy number ≥4.0 and <6.0 signals/cell AND concurrent dual probe Group 5

Group 1	Group 2	Group 3	Group 4	Group 5
Ratio ≥2.0 ≥4.0 signals/cell	Ratio ≥2.0 <4.0 signals/cell	Ratio <2.0 ≥6.0 signals/cell	Ratio <2.0 ≥4.0 and <6.0 signals/cell	Ratio <2.0 <4.0 signals/cell

Figure 9: Summary of HER2 ISH diagnostic criteria according to ASCO/CAP Guidelines 2018. Reprinted from College of American Pathologists' webpage (cap.org), HER2 Testing in Breast Cancer – Summary Of Changes And Definitions (58). Permission for reuse granted.

1.2.3.3.3 Ki-67

In breast cancer, an antibody targeting the Ki-67 antigen is used as a proliferation marker. Ki-67 is present in all phases of the cell cycle, except G0 (59). High levels of Ki-67 are associated with poor prognosis in breast cancer patients (60, 61), thus Ki-67 provides prognostic information. In addition, it can be used to separate luminal tumours into luminal A (low proliferation) and luminal B (high proliferation) (62), and to guide selection of patients for chemotherapy (35, 63, 64).

According to Norwegian guidelines, Ki-67 status is reported as a percentage of positively stained nuclei after counting 500 invasive tumour cells within the hot spot area. A hot spot is defined as the area of the tumour with the highest proportion of positively stained nuclei (35). Staining intensity is not considered. All laboratories are advised to participate in external quality control of Ki-67 (35).

1.2.3.3.3.1 Ki-67 controversies:

The clinical use of Ki-67 for predictive purposes is controversial due to inter- (65) and intra-observer and laboratory variations, and a lack of consensus regarding cut-off values, scoring and reporting (49, 63, 66). Preanalytical and analytical variation, in addition to ways of scoring and reporting may affect Ki-67 IHC results (63). In 2009 Cheang *et al.* found that a cut-off value of 13.25% was optimal for separating luminal A from luminal B ($HER2^-$) tumours. They used Ki-67 IHC on TMAs (tissue micro arrays) (62). On the basis of these results, Ki-67 </≥14% was chosen as the cut-off value to separate luminal A from luminal B

in the 2011 St Gallen guidelines (67). Due to variations in Ki-67 assessment, the International Ki-67 in Breast Cancer Working group was convened in 2010 (68). In 2011, they published their first set of recommendations on analytical assessments, interpretation and scoring of Ki-67 (68). However, they were not able to reach consensus on the ideal cut-off for Ki-67 (68). A later study initiated by the working group found substantial interobserver and interlaboratory variations at Ki-67 IHC levels >5% and <30% (63). Subsequently, the Working Group's updated recommendations from 2020 states that Ki-67 levels <5% and ≥30% could be used to reject or recommend chemotherapy to patients with $ER^+/HER2^-$ tumours (63). According to the most recent St Gallen guidelines there are no clear cut-offs for Ki-67 (63). However, the guidelines support the recommendations from the Ki-67 Working Group that tumours with Ki-67 <5% should not receive chemotherapy, whereas tumours with Ki-67 ≥30% should be treated with chemotherapy (64). According to current Norwegian guidelines, Ki-67 score should be interpreted in light of local laboratory values, using median Ki-67 score to define clearly high and clearly low proliferation status (35). The 2021 annual report from the Cancer Registry of Norway shows that the national median value of Ki-67 is 17%, with values ranging from 12 to 22 between the hospitals with the lowest and highest reported median values (13).

In spite of cut-off controversies, the FDA (Food and Drug Administration) recently approved abemaciclib in combination with endocrine therapy as a treatment option for patients with HR^+, $HER2^-$, lymph node positive, high risk, early breast cancers with Ki-67 ≥20% (69), based on results from the MonarchE study (70, 71).

Another aspect of Ki-67 scoring, is the type of specimen the scoring is performed on. By comparing levels of Ki-67 IHC scored on whole sections compared to TMAs, Knutsvik *et al.* found significantly higher levels of Ki-67 when scoring was done on whole sections compared to TMAs. A Ki-67 cut-off at 20% on whole sections and 8% on TMAs resulted in similar distribution of luminal subtypes (72).

1.2.3.3.4 PD-L1

The gene *PD-L1*, also known as *CD274*, is located on chromosome 9p24.1 (73). It encodes an inhibitory receptor ligand, which functions through binding to the PD1 receptor at the cell surface of T-cells, inhibiting T-cell activation and cytokine production (74). Expression of PD-L1 has been identified in non-small cell lung cancer, melanoma, epithelial ovarian cancer, breast cancer and gastrointestinal malignancies, and is generally associated with poor prognosis (74). PD-L1 is a target for immunotherapy, and treatment has already been

implemented for several cancers (74). PD-L1 status can be assessed using IHC (74, 75), and since April 2020, PD-L1 assessment has been included in Norwegian breast cancer guidelines for metastatic TNBCs. A tumour is considered PD-L1 positive if ≥ 1% of immune cells in the invasive tumour show positive staining (35). PD-L1 is expressed in 20% of TNBCs (76). At the 2021 St. Gallen consensus meeting, the panel was against routine reporting of PD-L1 and PD-L1 expression by immune cells (PD-L1ic) in early stage TNBC (64).

1.2.3.3.5 Multiparameter assays

Based on information about the genes and genetic pathways associated with breast cancer development and progression, multigene assays have been developed to provide prognostic and predictive information that can be used in breast cancer diagnostics and treatment. The main commercially available multigene assays and their characteristics are listed in Table 3.

Table 3 – Overview of the main commercially available multigene assays used for chemotherapy decision-making in ER-positive, HER2-negative breast cancer

	MammaPrint	**Oncotype DX**	**Prosigna (PAM50)**	**EndoPredict**	**Breast Cancer Index**	**Genomic Grade Index**
Number of genes	70	21	50	11	7	97
Method	DNA microarray	RT-PCR	NanoString	RT-PCR	RT-PCR	RT-PCR
Tissue sample type	Frozen/FFPE	FFPE	FFPE	FFPE	FFPE	FFPE
Location	Central	Central	Local	Local	Central	Central
Test results	High or low risk + subtype	High, intermediate, or low risk	High, intermediate, or low risk + subtype	High or low risk	High or low risk, high or low benefit	High or low risk
Clinical indication (according to EGMT)	Predicting prognosis and guiding decision-making regarding chemotherapy for women with **ER+/HER2-EBC, LN- or LN+ (1-3)**	Predicting prognosis and guiding decision-making regarding chemotherapy for women with **ER+/HER2-EBC, LN- or LN+ (1-3)**	Predicting prognosis and guiding decision-making regarding chemotherapy for women with **ER+/HER2-EBC, LN- or LN+ (1-3)**	Predicting prognosis and guiding decision-making regarding chemotherapy for women with **ER+/HER2-EBC, LN- or LN+ (1-3)**	Predicting prognosis and guiding decision-making regarding chemotherapy for women with **ER+/HER2-EBC, LN-**	No recommendation
Prospective validation trial(s)	MINDACT (positive)	TAILORx (positive) and RxPONDER (ongoing)	OPTIMA (ongoing)	None	None	ASTER70 (ongoing)
Regulatory approval	EMA, FDA	EMA, FDA	EMA, FDA	EMA, FDA	Not approved	Not approved
Original validation set	Developed in young patients (ages <55 years) who had not received systemic therapy after surgery	Developed in patients who had received tamoxifen only in the NSABP B-20 and B-14 trials	Postmenopausal patients in the training and development sets received heterogenous treatment	Developed in postmenopausal patients who had received endocrine therapy only in the ABCSG-6 and -8 trials	Initially developed in patients treated with tamoxifen only and then further refined in heterogeneously treated patients	Developed in a heterogeneous patient population (62% aged ≤50 years), in order to permit the measurement of histological grade via gene expression profile

Abbreviations: EBC, early breast cancer; EGMT, European Group on Tumor Markers; EMA, European Medicines Agency; FDA, United States Food and Drug Administration; FFPE, formalin-fixed, paraffin-embedded; LN, lymph node; RT-PCR, reverse transcriptase polymerase chain reaction.
Reprint from WHO Classification of Tumours Editorial Board. Breast Tumours (2019) (42). With permission from original publishing by Brandao M, Pondé N, Piccart-Gebhart M. Mammaprint ™: a comprehensive review (License ID: 1263305-1).

Several of the multigene assays include overlapping genes (Figure 10).

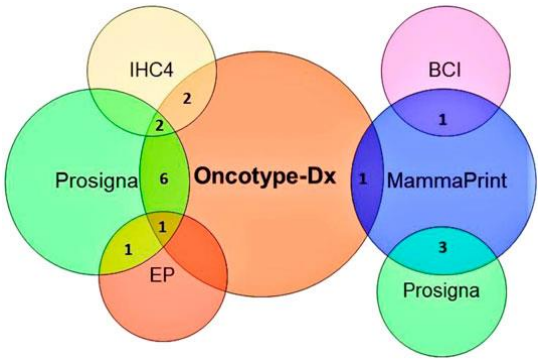


Figure 10: Venn diagram showing the number of overlapping genes/proteins between different gene expression assays. Abbreviations: IHC4: immunohistochemical 4; EP: EndoPredict; BCI: Breast Cancer Index. Reprinted from Front. Pharmacol. (2021), Zubair *M et al.* Advanced Approaches to Breast Cancer Classification and Diagnosis, Figure 5 (77). Paper distributed under a Creative Commons license (CC BY-NC 4.0) and can be used for non-commercial purposes.

In the clinical setting, gene expression tests can be used to make decisions on chemotherapy for patients with ER^+, $HER2^-$ breast cancer (42, 78). The tests provide a risk score ranging from low to high, where patients defined as low risk can be spared chemotherapy, whereas high-risk patients could benefit from chemotherapy (42). A test based on four commonly reported IHC markers has also been developed (IHC4; ER, PR, HER2 and Ki-67), which can provide prognostic information in early ER-positive breast cancer (79). In a comparative study including Oncotype Dx, PAM50, BCI, Endopredict, Clinical Treatment Score and IHC-4, the three tests PAM50, Breast Cancer Index and EndoPredict were found to be most prognostic for overall and late distant recurrence (80).

According to Norwegian guidelines of 2022, the PAM50 (Prosigna) test is the only recommended multiparameter assay (35). In addition to risk score, PAM50 classify the tumours into one of the four intrinsic subtypes (luminal A, luminal B, HER2-enriched and basal-like).

1.3 Treatment options according to Norwegian guidelines

Factors such as tumour size, histological grade, stage, lymph node involvement, biomarkers (ER, PR, HER2, Ki67 and PD-L1) and the patient´s age, comorbidities and own preferences are considered when deciding on treatment. Most patients are metastasis-free at the time of diagnosis. The primary treatment for breast cancers stage T1-2 is surgical resection of the

tumour with examination of sentinel lymph node(s) (SLN), and consideration for radiation therapy postoperatively. Systemic therapy can be given neoadjuvant and/or adjuvant (35).

1.3.1 Surgery

The aim of surgical treatment is to increase survival through local control of the tumour. Surgical treatment options are either mastectomy or breast conserving surgery. In addition, the sentinel lymph node is examined. In case of sentinel node positivity, axillary dissection can be performed. Patients with locally advanced breast cancer receive neoadjuvant treatment (35).

1.3.2 Radiation

Radiation therapy can be directed towards the breast, chest wall and regional lymph nodes. Radiation therapy is performed postoperatively after breast conserving surgery and some mastectomies. The aim is to reduce risk of recurrence and improve survival (35).

1.3.3 Adjuvant and neoadjuvant systemic therapy

Neoadjuvant therapy can be used in cases where breast conserving therapy is difficult due to tumour size or localization, and in cases with locally advanced cancer. It can also be considered for $HER2^+$ and TN tumours when tumour size is >2 cm. In general, adjuvant chemotherapy is offered to breast cancer patients with the following characteristics:

- established lymph node metastases
- patients with non-metastatic disease with:
    - primary tumour size >2 cm
    - primary tumours size >1-10 mm and $HER2^+$ and/or HR^- and/or high proliferative status (Ki-67)/increased risk profile/luminal B-subtype (gene expression analyses).

For $HR^+/HER2^-$ tumours, chemotherapy is considered based on tumour stage, grade, expression of HR, proliferation status (Ki-67), and Prosigna results when available. Targeted therapies are available for certain subgroups of breast cancers, such as endocrine therapy for ER^+ tumours, anti-HER2 treatment such as trastuzumab for $HER2^+$ tumours, and atezolizumab (immunotherapy) for metastatic, TNBCs with positive PD-L1 (35).

1.4 Molecular subtypes of breast cancer

In 2000, Perou *et al.* published "Molecular portraits of human breast tumours", which was a paradigm shift in our understanding of intertumoral heterogeneity in breast cancer. Performing gene expression analyses on 8102 human genes on tissue samples from 42 different patients (36 infiltrating ductal carcinomas, two lobular carcinomas, one ductal carcinoma *in situ*, one fibroadenoma and three normal breast samples) and 17 cultured cell lines, four different subtypes were identified: luminal epithelial/ER^+, *Erb-B2*$^+$, basal-like and normal-breast-like. The subtypes were defined by different gene expression patterns that were suggested to be associated with different aspects of mammary epithelial biology. Luminal epithelial/ER+ tumours were ER^+, *Erb-B2*$^-$, and showed high expression of genes expressed by luminal epithelial cells, the *Erb-B2*$^+$ tumours showed low expression of ER and overexpression of the *Erb-B2* (*HER2)* oncogene and associated genes, the basal-like tumours expressed genes typical for basal epithelial cells and the normal-breast-like tumours expressed genes characteristic of normal breast tissue such as basal epithelial cells and adipose cells (81).

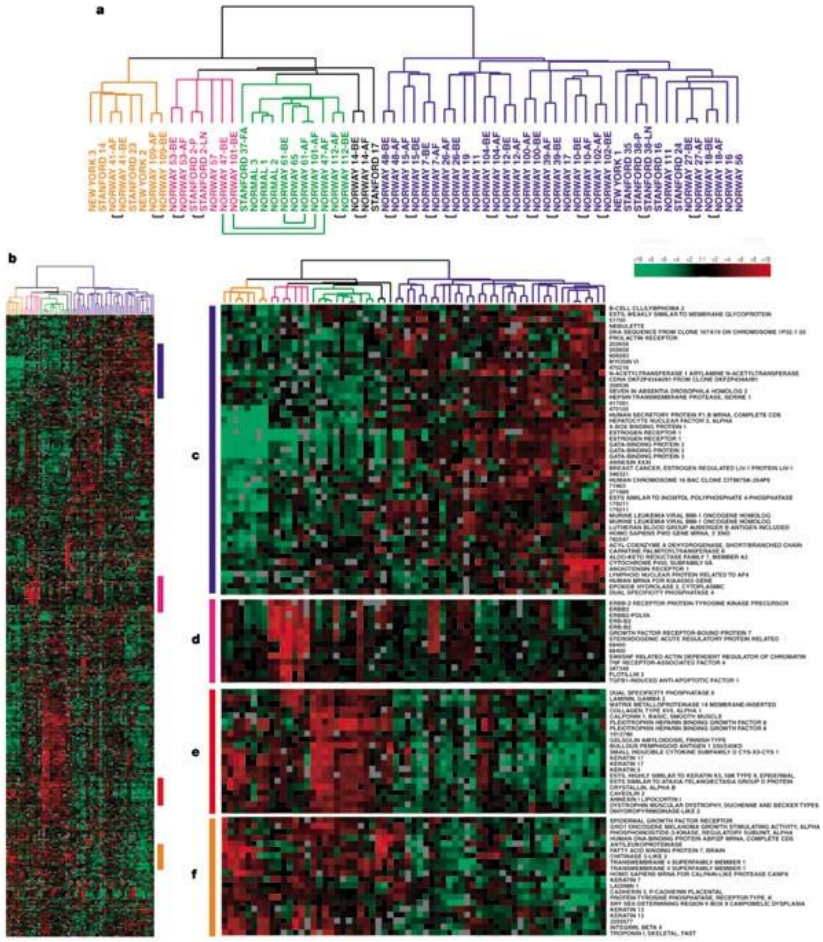


Figure 11: Cluster analyses based on the "intrinsic" subset of 496 genes. Each row represents a single gene and each column represents a sample. A green square indicated that the gene is down-regulated, a red square indicates that the gene is up-regulated, and a black square indicates that the gene expression is unchanged. Each subtype is identified through coloured branches as follows: orange – basal like; pink - *Erb-B2*$^{+}$; green - normal-breast-like; purple - luminal epithelial/ER+. Permission to reuse granted by the Macmillan Magazines Ltd. (License number 5261450385401): Perou, *et al.* Molecular portraits of human breast tumours. Nature vol. 406 (2000) (81), page no 750, Figure 3.

Luminal tumours have been further subdivided into luminal A and luminal B tumours (82). It has been questioned whether the normal-like group is in fact a separate breast cancer subgroup, or rather a result of analysing tissue samples with abundant normal breast tissue (83). In addition, a subtype called claudin-low tumours has been identified (84). Claudin-low tumours have poor prognosis, are mostly triple negative (ER^{-}, PR^{-} and HER2^{-}), and have a poorer response to chemotherapy than basal-like tumours (85). Several studies have validated

the classification of the intrinsic subtypes luminal A, luminal B, $HER2^+$ and basal, and identified them as subgroups of breast cancer with distinct biological features (83, 86) and different prognosis (82, 83, 86, 87).

Luminal A tumours are characterized by high expression of genes characteristic of luminal epithelial cells. Few luminal A tumours have *TP53* mutations (82). Additionally, they are associated with low histological grade, low proliferation, and good prognosis (42, 43, 82, 83). Luminal B tumours are characterized by a lower expression of luminal specific genes (82), a higher expression of proliferation associated genes and poorer survival compared to luminal A tumours (88). A proportion of luminal B tumours are HER2 positive (62).

HER2 tumours are characterized by high expression of *ERBB2* and other genes situated at the same amplicon on chromosome 17 (82, 88), and high frequency of *TP53* mutations (82). Some $HER2^+$ tumours are ER^+ (83). HER2 tumours are associated with poor prognosis (82, 83, 86) and high histological grade (42, 43, 88).

Basal-like tumours are characterized by high expression of genes characteristic of basal epithelial cells such as *keratin 5* and *keratin 17*, and low expression of *ER* and other genes found in the ER/luminal cluster (81, 82, 86). In addition, they are associated with high frequencies of *TP53* (82) and *BRCA1* mutations (86), high histologic grade (42, 83, 88) and poor prognosis (82).

In the mid-2000s patterns of copy number alterations were identified, which were characterized by distinct patterns of deletions, duplications and amplifications of genes and chromosome regions (89). In 2012 Curtis *et al.* identified 10 integrative clusters by combining copy number and gene expression analyses on a collection of approximately 2000 breast cancer samples. The clusters were characterized by different patterns of copy number aberrations and different prognosis. Comparison of the integrative clusters and PAM50 defined intrinsic subtypes showed that six of the clusters were predominantly ER-positive (IntClust 1, 2, 3, 6, 7 and 8) while cluster 10 predominantly comprised basal-like tumours (90). IntClust 4 and 9 comprised both ER-positive and ER-negative tumours of varying intrinsic subtypes, IntClust 5 comprised a combination of HER2-positive and ER-positive tumours (90).

While genomic studies on breast cancer have revealed distinct subtypes, a study performed by Hoadley *et al.* found that certain genomic signatures are shared across tumours from different organs. They examined genomic signatures within and between 12 different cancer types and found that breast cancers could be divided into two distinctly identifiable groups: one group comprising nearly all luminal and HER2-positive tumours, and another

group comprising basal-like tumours. Interestingly, they found that the basal-like group was as different from the luminal group, as it was from lung cancer. Furthermore, the basal-like group showed pathway similarities with squamous cell carcinomas and serous ovarian cancers (91).

Through extensive genome sequencing of breast tumours certain genes have been identified as potential drivers in the evolution of the different subtypes of breast cancer. Among the most frequently mutated genes in breast cancer are *TP53, PIK3CA, MYC, CCND1, PTEN, ZNF703* and *GATA3* (92). Mutations of some of these genes are associated with certain clinical and pathological characteristics, for instance it has been found that *PIK3CA* mutations are most frequent in luminal A tumours, while *TP53* mutations occur more frequently in basal-like tumours (42, 89).

1.4.1 Molecular subtypes defined by surrogate markers

Gene expression analyses used for molecular subtyping are costly and still unavailable for many institutions, both nationally and internationally. Studies have shown that IHC and ISH can be used as surrogates for gene expression analyses to classify tumours into molecular subtypes (62, 93). Different algorithms have been used to classify breast cancer tumours into subtypes, defined by different molecular markers (94, 95).

IHC for ER and PR can be used to separate tumours into luminal and non-luminal subtypes (62, 96). Using IHC for ER, PR, HER2 and Ki-67, Cheang *et al.* were able to classify tumours into luminal A, luminal B ($HER2^-$) and luminal/$HER2^+$ subtypes. Ki-67 cut-off 13.25% was considered optimal to distinguish between luminal A and B tumours. Tumours positive for ER and/or PR and HER2 were classified as luminal-HER2-positive, representing a distinct group of luminal B tumours requiring a HER2-targeted treatment approach. Luminal B and luminal-HER2-positive tumours are associated with younger age, higher grade, larger tumour size and poorer prognosis compared to luminal A tumours (62).

Among the non-luminal tumours, HER2+ tumours are classified by overexpression of several genes in the *HER2* amplicon (81, 82). Tumours that are negative for ER and/or PR and positive for HER2 detected by IHC or FISH can be named HER2 type (96). HER2 type is associated with high grade and poor survival (96, 97).

TNBC are tumours that are ER, PR, and HER2 negative (42, 98-100). Even though there is great overlap between TNBC and the basal-like subtype, not all TNBC are basal-like (98, 99, 101). TNBCs generally have high grade and poor prognosis and occur more frequently among African American (95) and African women (102).

TNBC is a heterogenous group of tumours, and through mRNA analyses Lehman *et al.* identified six TNBC subtypes; luminal androgen receptor type (LAR), basal-like 1 (BL1) and 2 (BL2), immunomodulatory (IM), mesenchymal stem-like (MSL) and mesenchymal-like (M) (101). The six subtypes were later redefined into four subtypes; LAR, BL1, BL2 and M (100), with different characteristics, prognosis, and treatment response (100, 101). This heterogeneity within TNBCs is not captured when subtype classification is solely based on the three biomarkers ER, PR and HER2. Different surrogate markers have been suggested to stratify TNBCs into subgroups. In 2004, Nielsen *et al.* found that basal-like tumours could be identified using a panel of four surrogate markers (ER, HER2, (cytokeratin 5/6) CK5/6 and human epidermal growth factors receptor 1 ((HER1), also known as epidermal growth factor receptor (EGFR)) (93). Subsequently, in 2006 Carey *et al.* added PR to the algorithm and defined basal-like tumours as ER^-, PR^-, $HER2^-$, $CK5/6^+$ and/or HER1 $(EFGR)^+$ (95). In 2008 Cheang *et al.* used five biomarkers to separate TNBCs into two groups; the core basal phenotype (CBP) (ER^-, PR^-, $HER2^-$, $EGFR^+$, $CK5/6^+$), and the 5 negative phenotype (negative for all five markers). They found that the CBP identified basal-like tumours more correctly, and predicted prognosis more precisely compared to the TNBC algorithm based on three markers. (96).

A recently published study has developed a TNBC subtyping algorithm using p16, androgen receptor and tumour infiltrating lymphocytes to classify tumours into four subtypes (basal-like, mesenchymal-like, immunomodulatory, and luminal androgen receptor) that correlates with the TNBC subtypes described by Lehmann *et al.* (103).

TNBCs have generally been treated with chemotherapy, however there is ongoing research to find potential biomarkers and targeted therapies to improve the prognosis for these patients (51, 98). For instance, PD-L1 positive TNBCs qualify for immunotherapy (35).

Our group, the Breast Cancer Subtypes (BCS) Research Group at NTNU, previously reclassified three cohorts of breast cancer tumours into molecular subtypes (Figure 12) (97). The tumours were classified using IHC and ISH as surrogates for gene expression. IHC was done with antibodies for ER, PR, HER2, CK5, Ki-67 and EGFR. HER2 status was determined using both IHC and chromogenic *in situ* hybridization (CISH).

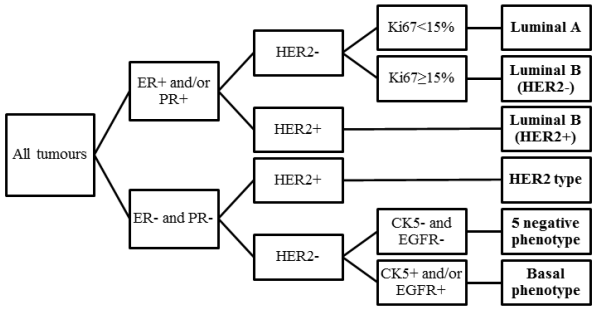


Figure 12: Molecular subtyping of breast cancer tumours using six molecular markers. Modified after Engstrøm MJ, *et al.* Molecular subtypes, histopathological grade, and survival in a historic cohort of breast cancer patients. Breast Cancer Res Treat. 2013 Aug;140(3):463-73 (97).

The St. Gallen Consensus conference in 2011 introduced molecular subtyping of breast cancer using surrogate makers ER, PR, HER2 and Ki-67/grade (67). In 2013 St. Gallen adopted the following algorithm for classification of subtypes (99):

Table 4:

Intrinsic subtype	**Clinicopathological subtypes**	**Surrogate molecular markers**
Luminal A	Luminal A-like	All of: ER and PR positive HER2 negative Ki-67 low recurrence risk "low" based on multi-gene-expression assay (if available)
Luminal B	Luminal B (HER2 negative)	ER positive, HER2 negative and at least one of the following: Ki-67 high PR low/negative recurrence risk "high" based on gene expression assays (if available)
	Luminal B (HER2 positive)	ER and HER2 positive Ki-67 - any expression PR – any expression
ERB-B2 overexpression	HER2 positive	HER2 positive, ER and PR absent
Basal like	Triple negative	ER and PR absent, HER2 negative

This algorithm is still used according to current national and international guidelines (35, 42).

There are discrepancies between subtypes determined by surrogate methods (IHC and ISH), and multiparameter gene expression assays (83, 104, 105). When PAM50 was used to identify intrinsic subtypes in a subset of ER^+ tumours, 11% were classified as HER2-enriched and 5% were basal-like. Furthermore, among ER^- tumours 11% were classified as luminal (83). Nevertheless, using IHC/ISH as surrogate markers for classification of breast cancer subtypes has proven to give important prognostic information (62, 93, 94, 106, 107).

Figure 13 gives an overview of the histological and molecular characteristics of breast cancer.

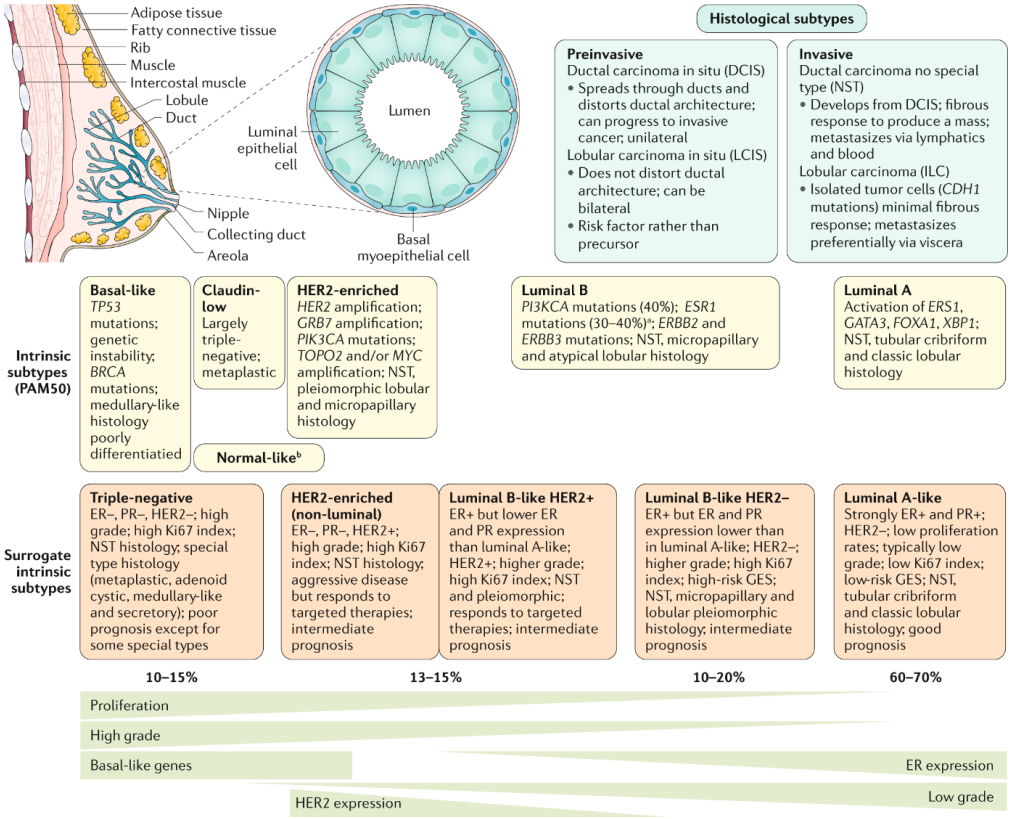


Figure 13: Histological and molecular characteristics of breast cancers. Reprint from Harbeck, N. *et al.* Breast cancer. Nat Rev Dis Primers, Fig 1. Permission for reuse granted. Licence number 5181290578675 (22).

1.5 Proliferation in breast cancer

In 2000 the Hallmarks of Cancer were first introduced with the intent to describe certain functional capabilities that are common steps in the pathogenesis of all cancers. The following

six Hallmarks were introduced: self-sufficiency in growth signals, insensitivity to anti-growth signals, evading apoptosis, limitless replicative potential, sustained replicative potential, and tissue invasion and metastasis (108). In 2011, the concept “enabling characteristics” was introduced, referring to the molecular and cellular mechanisms by which the hallmarks are acquired rather than the capabilities themselves. Furthermore, two functional characteristics; deregulating cellular energetics and avoiding immune destruction, and two enabling characteristics; genome instability and tumour-promoting inflammation were added (109). In 2022, Hanahan proposed adding four new characteristics to the hallmarks, three enabling traits and one functional characteristic (110) (Figure 14).

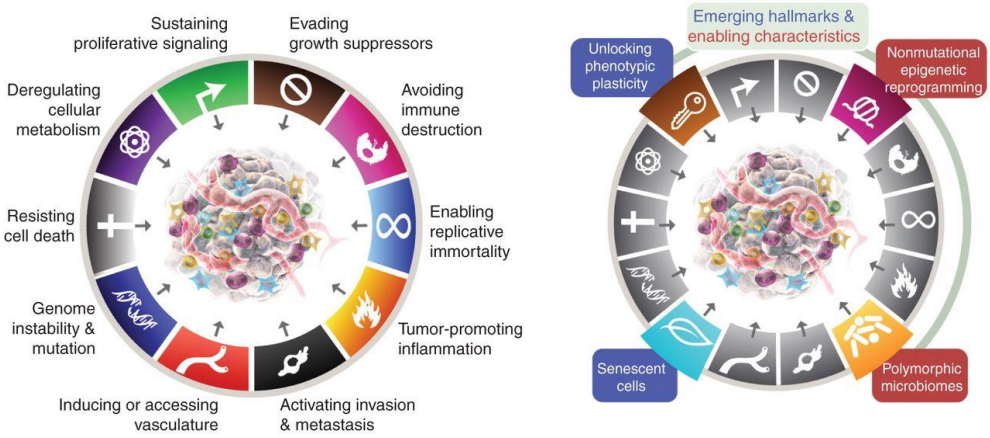


Figure 14: The Hallmarks of Cancer. To the left: The current hallmarks with eight functional characteristics and two enabling characteristics. To the right: The four new proposed additions to the hallmarks. Reprinted from the American Association for Cancer Research (License number 5261860500719); Hanahan, *Hallmarks of Cancer: New Dimension*, Cancer Discovery (2022) (110), Figure 1, Page 32.

Increased proliferative activity was one of the identified hallmarks of cancer (108, 109) (Figure 14). The tumour cells evade the normal cell growth-and-division cycle through several mechanics, for instance by producing their own growth factors, increasing the levels of receptors on the cell surface or through downregulation of negative feedback pathways (109).

High proliferation is strongly associated with poor prognosis in breast cancer (111, 112). Proliferation can be assessed in breast cancer by mitotic count, estimation of Ki-67 levels, or gene expression assays (22). When Perou *et al.* first identified the intrinsic clusters, a cluster of proliferation associated genes was discovered (81). The different breast cancer subtypes are characterised by different gene signatures, and different levels of proliferation (112). A meta-analysis of several gene signatures identified through analyses of breast cancer

subtypes found that their prognostic abilities were mostly due to proliferation associated genes (112). Proliferation associated genes are included in gene expression assays used for subtyping and estimation of risk of recurrence. For instance, a proliferation signature comprising 11 genes is included in the PAM50 gene expression assay (113). Gene expression tests are described in greater detail in section 1.2.3.3.5.

Mitotic count is a well-established part of histological grading, and may also have independent prognostic value (111). A large review found that increased mitotic index and Ki-67 both were associated with reduced overall and disease free survival in early breast cancer (61). Ki-67 is the most commonly used marker of proliferation in breast cancer (68, 111, 114, 115), however, its value as a predictive marker is questioned due to variability in assessment, reporting and lack of clear guidelines for cut-off (63, 65, 66, 68, 115). Mitotic count is also included as a proliferation marker according to Norwegian guidelines (116). Mitoses can be counted during the M-phase of the cell cycle due to the characteristic appearance and alignment of the chromosomes. HE staining is used for visualization (117). Ki-67 IHC is available in most pathology laboratories, and it is an inexpensive measure of proliferation. Using multigene panels, several proliferation markers can be included in the analysis, thus providing more robust data, however such tests are more costly and therefore not available for all.

Several other proliferation markers have been studied in breast cancer, such as proliferating cell nuclear antigen (PCNA) and phosphorylated histone H3 (PhH3) (118). PCNA is a cell cycle marker, and it is significantly elevated during the S and G2 phase (115). However, studies have found variable associations between increased PCNA expression and proliferation and prognosis, and the marker is not included in routine breast cancer diagnostics (115, 118). PhH3 is a nuclear core histone protein which is expressed in the cells in the mitotic phase, thus PhH3 IHC has been proposed as a marker of mitotic activity and consequently of proliferation (118).

In 2014, Gatza *et al.* identified four amplified genes (*Mitochondrial Ribosomal Protein S23* (*MRPS23), Faciogenital dysplasia 5* (*FGD5), Methyltransferase Like 6* (*METTL6)* and *Deltex E3 Ubiquitin Ligase 3* (*DTX3))* that were essential for proliferation and associated with poor prognosis in luminal breast cancers (113). *MRPS23* is described in more detail in section 1.5.2.

1.5.1 Gene amplification in breast cancer

Amplification is defined as a copy number increase of a defined section of the chromosome arm (119). Gene amplification is a common event in breast cancer (120), and among the chromogenic regions containing known driver oncogenes are 8p11 (*Zink finger protein 703* (*ZNF703*)), 8q24, 11q13 and 17q11-22 (*HER2, MRPS23*) (121). Amplification of the chromosomal region 8p11-12 is one of the most frequent in breast cancer, occurring in 10-22% of cases (122). Amplification of the region is associated with poor prognosis in breast cancer (122, 123). Co-amplification with other regions, such as 11q13 can occur (122).

1.5.2 MRPS23

The gene *MRPS23* is a part of the mitochondrial ribosomal protein (MRP) gene family, and it is located on the long arm of chromosome 17 (17q22) (124). *HER2* is also situated on the long arm of chromosome 17 (17q12) (53). The protein MRPS23 is involved in protein synthesis within the mitochondrion (125). To identify amplified genes that were associated with proliferation, Gatza *et al.* used a gene expression proliferation signature from the PAM50 signature to analyze copy number data from the TCGA dataset. To identify genes that were uniquely amplified in highly proliferative luminal tumours, they compared copy number alterations in luminal tumours to other tumours. “Luminal” tumours were defined as non-basal, in other words also including $HER2^+$ tumours. They then used information from a genome-wide RNA-mediated interference screen on breast cancer cell lines, to distinguish essential from non-essential genes. In doing so, they identified 21 genes that were amplified exclusively in highly proliferative luminal tumours, and that were essential for cell proliferation in vitro. Amplified genes that had an association between copy number and gene expression were then selected, and after validation in a second dataset, eight amplified genes remained. Amplification of four of these genes was associated with a poor prognosis (*FGD5, METTL6, DTX3* and *MRPS23)* (113). *MRPS23* was chosen for assessment in paper I. The strategy used by Gatza *et al.* to identify the proliferation associated genes is presented in Figure 15.

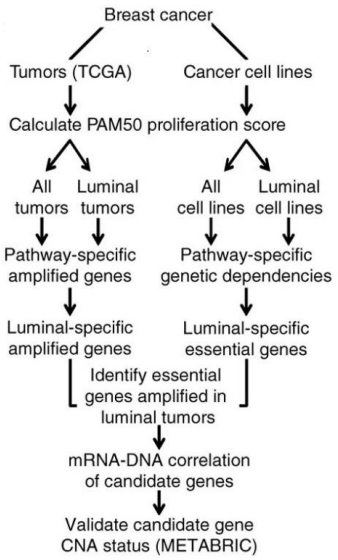


Figure 15: Outlining of the strategy used to identify essential genes amplified in highly proliferative luminal breast tumours. Reprinted from Gatza *et al*. An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer; *Nat Genet* **46** (2014) (113), Figure 5, Page 1056. Permission to reuse granted (License number 5356390668677).

Knock down of *MRPS23* both *in vitro* and in a breast cancer rat model is shown to reduce proliferation and metastasis of cancer cells (126). Increased expression of *MRPS23* has been found in cervical (127, 128), colon (129) and hepatocellular cancer (125, 130). An interaction between p53 and MRPS23 has been described (131), and it is shown that regulation of MRPS23 protein through methylation may play a role in the development of breast cancer metastasis (132).

1.5.3 ZNF703

Amplification of 8p11-12 has been identified in 10% -15% of breast cancers (133). It is associated with a poor prognosis in breast cancer. Four distinct amplicons (A1-A4) have been identified in the 8p12 region (134). The gene *ZNF703* is a part of the A1 amplicon (134-136). *ZNF703* is amplified in 8% of breast tumours (133), and in 9-19% of ER+ tumours (135, 137). The gene is more frequently amplified in ER+, PR-, HER2- tumours than in ER+, PR+, HER2- tumours (138). *ZNF703* is frequently co-amplified with other genes situated in the same chromosomal region, such as *FGFR1* (139), and with genes situated on other chromogenic regions such as *CCND1* on 11q13 (122, 140). *ZNF703* is preferentially

expressed in luminal B tumours (133, 135, 141, 142). *ERBB2* (*HER2*) and *ZNF703* amplifications have been found to be almost mutually exclusive in luminal tumours (135). Recently, it has also been suggested that *ZNF703* plays a potential role in TNBCs as well (143). High expression of *ZNF703* is associated with high proliferation in breast cancer (135, 141, 144), and poor prognosis in luminal tumours (135). High expression of *ZNF703* has also been identified in head and neck cancer (145-147), non-small cell lung cancer (148, 149), gastric cancer (150) and cholangiocarcinoma (151). Studies have found that there is a positive association between *ZNF703* gene copy number and nuclear *ZNF703* mRNA expression in breast cancer cells (135, 141).

Several mechanisms explaining the role of *ZNF703* in breast cancer progression have been suggested. Studies have indicated that overexpression of ZNF703 affects cell cycle regulation (135, 141), through regulation of the ER signalling pathway (135, 141, 152). Zhang *et al*. found that overexpression of ZNF703 in breast cancer cells reduced the effect of tamoxifen-treatment, while knock down of ZNF703 increased the antitumor effect of tamoxifen (152). It has been suggested that *ZNF703* could be a novel human breast cancer oncogene (133, 135).

Aims

The aims of the three papers included in the research thesis were:

I. To describe *MRPS23* copy number change in breast cancer, and to assess associations between copy number alterations, and molecular subtype, proliferation, and prognosis. A second aim was to study associations between *MRPS23* gene expression and molecular subtype and prognosis.

II. To describe *ZNF703* gene copy number and protein expression in breast cancer, and to assess associations between gene copy number, molecular subtype, proliferation, and prognosis; between gene copy number and protein expression; and between protein expression and prognosis.

III. To study long term trends in breast cancer incidence according to proliferation status in three large cohorts of Norwegian women, using Ki-67 protein expression and mitotic count as markers of proliferation.

Materials and methods

Patients in cohort 1, 3 and 4 were followed for breast cancer occurrence through linkage to the Cancer Registry of Norway, the Cause of Death Registry, and the National Registry, using the Norwegian 11-digit identity number.

1.6 Study population

1.6.1 Cohort 1

From 1956 to 1959, all women aged 20-69 years in the Norwegian county of Nord-Trøndelag were invited to attend a clinical screening program for early detection of breast cancer. Birth year ranged from 1886 to 1928 (Figure 16). Participants were interviewed using a standard questionnaire about reproductive variables (menopause, menarche, parity, lactation) and underwent a clinical breast examination. Individual identity numbers given to all Norwegian citizens (since 1961) were used for follow-up. A total of 25 897 women participated in the screening program. Of these, 170 were excluded because they were diagnosed with breast cancer prior to inclusion, restricting our analyses to the remaining 25 727 women. The patients were followed for breast cancer occurrence from January 1st, 1961, to December 31st, 2008, and a total of 1379 new cases were registered. Of these, formalin-fixed, paraffin-embedded (FFPE) tissue was available for 945 cases. Using IHC and ISH, 909 of these tumours were reclassified into molecular subtypes by the BCS Group at NTNU, Norway (97). After diagnosis, patients were followed until time of death from breast cancer or other causes, or until December 31st, 2015, for paper I and II. For paper III, end of follow up was defined as December 31st, 2010, due to lack of more recent follow up data for the healthy background population.

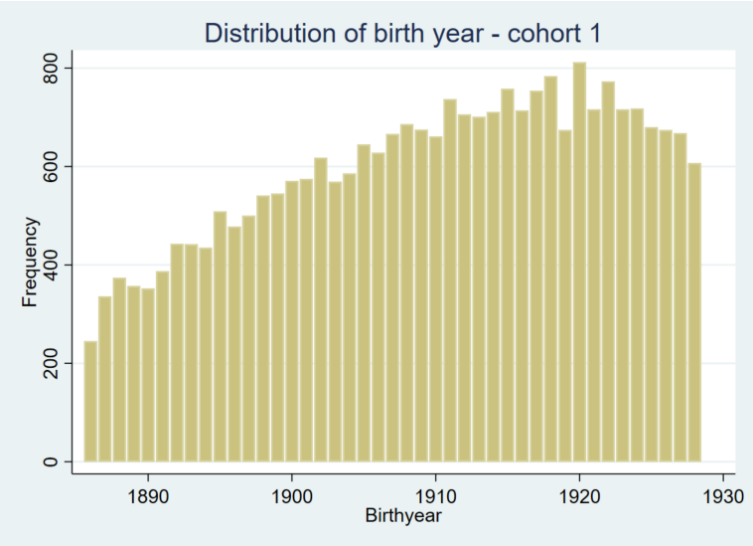


Figure 16: Distribution of birth year in cohort 1.

1.6.2 Cohort 2 - METABRIC

The METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) dataset includes specimens from approximately 2000 breast cancers collected from different tumour banks in the UK and Canada (90, 153). An initial discovery dataset of 997 breast cancers was analysed, and a validation dataset of 995 breast cancers was used to test for reproducibility. The tumours were analysed for copy number variants, copy number aberrations, and single-nucleotide variants. By combining gene expression profiles and clustering of copy number analyses the tumours were divided into ten subgroups with different copy number profiles and prognosis (90).

1.6.3 Cohort 3 – HUNT2

From August 1995 to June 1997, all residents aged 20 years and older in the county of Nord-Trøndelag were invited to participate in the second large health study of its kind in the region, HUNT2 (154). Participants were asked to fill out three questionnaires on a wide range of topics, such as health (including female reproductive data), personal environment, and habits. They also underwent a screening session (examination of blood pressure and heart rate, height and weight, and waist and hip circumference) and blood sampling. Participation was voluntary, and participants signed a written consent allowing for data and blood samples to be used in future research. They agreed to linkage of personal data to other registries (154). A

total of 34 641 women born from 1897 to 1977 participated in the survey (107). Figure 17 shows the distribution and frequency of birth year in cohort 3.

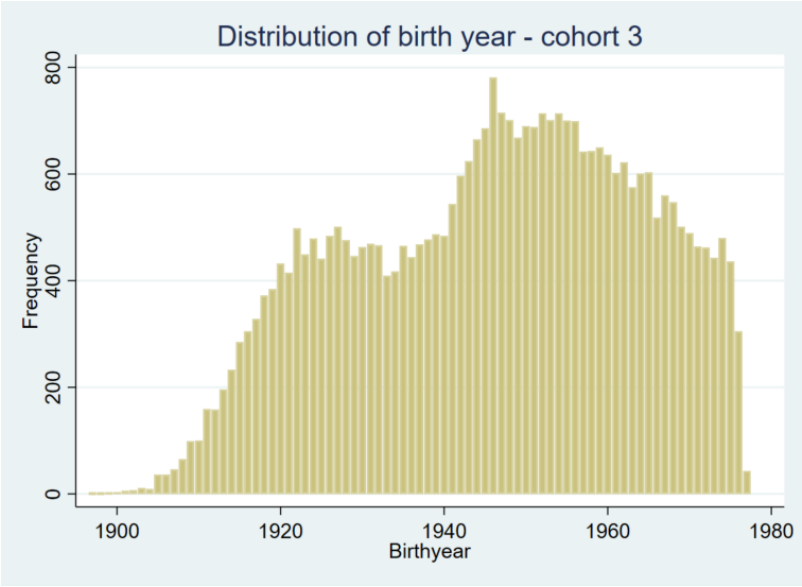


Figure 17: Distribution and frequency of birth year in cohort 3.

Of these, 420 were excluded because they were diagnosed with breast cancer prior to inclusion, restricting our analyses to the remaining 34 221. The participants were followed for breast cancer occurrence from the time of attendance to December 31st, 2009, and 731 incident breast cancer cases were registered (107). Of the 731 tumours, 653 were previously reclassified into molecular subtypes (97, 107). After diagnosis, the patients were followed until death from breast cancer or other causes, or until December 31st, 2015.

1.6.4 Cohort 4 – ECD

All women born at E. C. Dahl's foundation, located in the city of Trondheim, in the county of Sør-Trøndelag, from 1920 to 1966 were followed for breast cancer occurrence (155). Figure 18 shows the distribution and frequency of birth year in the cohort. The study has previously been described in detail (155, 156). The participants were followed for breast cancer occurrence from 1961 to December 31st, 2012. Of the 23 510 participants, 32 were excluded from our analyses due to missing information, 12 were excluded because they were diagnosed with breast cancer prior to inclusion, and 116 were excluded because they were lost to follow up before 20 years of age. Of the 23 350 participants included in our analyses, 885 incident

breast cancers were registered. Of these, 545 tumours were previously reclassified into molecular subtypes (157). After diagnosis, the patients were followed until death from breast cancer or other causes, or until December 31st, 2015.

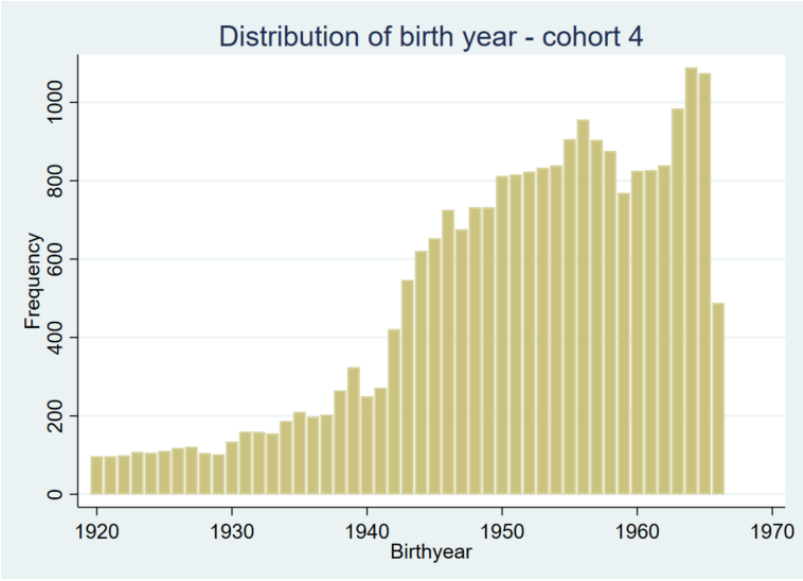


Figure 18: Distribution and frequency of birth year in cohort 4.

1.7 Norwegian public and health registries

1.7.1 The National Registry

The National Registry contains information of everyone that resides or have resided in Norway. The registry contains information such as name and birth, address, death and citizenship, and the information is used by several other institutions, such as the tax authorities, the electoral authority, bank and insurance companies and researchers (158).

1.7.2 The Cancer Registry of Norway

The Cancer Registry of Norway was established in 1951. It combines data from clinical registries, the Norwegian Patient Registry (NPR) and the Cause of Death Registry, in order to give an overview of cancer demographics in Norway (8, 159). Data from the different registries are connected through the personal identification number system. The main aims of the registry are to register cancer incidence, provide data for cancer research and give information about preventive measures and treatment against cancer (8). Data from all malignant neoplasms (except basal cell carcinoma in adults) and precancerous disorders, and

all benign tumours of the central nervous system and meninges are included in the registry (8, 159). Reporting of cancer is regulated by Norwegian law, and the registry provides a close-to-complete, reliable overview of cancer incidence in Norway (160).

1.7.3 The Norwegian Cause of Death Registry

The Norwegian Cause of Death Registry contains information about all deaths occurring in Norway, and all deaths that occur abroad to a person with registered residential address in Norway. The purpose of the registry is to monitor causes of death over time (161). Medical doctors are obliged to write death certificates according to WHO International Classification of Diseases (162).

1.8 Ethical considerations and approvals

The Helsinki declaration from 1964 contains ethical principles for research regarding humans (163). Since 2009, regulations concerning medical and health research have been gathered in the Norwegian Health Research Act (164). Both the Helsinki declaration and the Health Research Act states the need for approval from a research ethics committee, the importance of privacy and confidentiality regarding patient data and the need for informed consent. According to the Health Research Act exemptions from the requirement of consent can be made "if the research is of significant interest to society and the participants' welfare and integrity are ensured" (164). The first regional committee for medical and health research ethics (REK) was established in Norway in 1985. Today, all medical and health research in Norway needs approval from REK (164).

Cohorts 1, 3 and 4 used in this thesis comprise 83 298 women, of whom 2995 were diagnosed with breast cancer. Of the included breast cancer patients, 1829 (61%) had died from breast cancer or other causes during follow-up. The fact that many of the patients were no longer alive or were of a very high age made it difficult to obtain informed consent from all participants. Furthermore, the completeness of the cohorts was of great importance for the strength of the statistical analyses.

Anonymity was obtained by replacing patient identity numbers with case numbers in the research data file. There was no contact between the participants and the researchers conducting the study. The research included in this thesis caused neither harm nor benefit to the included patients. The results of the research could benefit future generations.

The Breast Cancer subtypes project has been approved by REK, Midt Norge (REK 836-09). Permission for linkage between data from the Cancer Registry of Norway and the Cause of Death Registry has been granted by the Norwegian Data Inspectorate. Studies on cohort 1 and 4 were granted dispensation from the usual requirement of informed consent. The participants in cohort 3 signed a consent allowing for the use of patient data and blood samples in further research (154).

In cohort 2, tumours tissue and linked anonymized clinical data was obtained with appropriate consent from the relevant institutional review boards (90).

1.9 Guidelines for tumour marker studies

The REMARK (REporting recommendations for tumour MARKer prognostic studies) guidelines were published by The National Cancer Institute and the European Organisation for Research and Treatment of Cancer in 2006. The aims of the guidelines are to improve the reporting of tumour marker studies and to make it easier to judge the usefulness of the data and to interpret the results in the appropriate context (165).

1.10 Tissue samples

Breast cancer cases were identified through the Cancer Registry of Norway. FFPE tissue samples and corresponding pathology reports were obtained from the department of Pathology at St. Olav's Hospital, Trondheim University Hospital, Norway.

1.11 Laboratory methods

1.11.1 Tissue microarray

TMA is a method that assembles tissue from multiple paraffin blocks into one common recipient block, allowing for cost-effective, rapid analyses of biomarkers on a large number of tumours at a time (166, 167). Another advantage of this method is that it allows for standardized analytical conditions for multiple cases. A disadvantage is that it is not possible to compensate for different preanalytical conditions. Studies have shown that TMAs may be representative of the corresponding whole sections (167-169), however tumour heterogeneity may be a problem due to the small size of the TMA tissue samples (167, 169). Therefore, specific guidelines regarding patient cohort and tissue samples, manufacturing of the TMAs and reporting of results have been suggested for TMA studies (169). The TMA blocks used in our studies were made using a Tissue Arrayer Mini-Core with TMA Designer2 software

(Alphelys). Three 1mm-in-diameter tissue cores were drawn from peripheral regions of FFPE primary tumours, and from lymph node metastases and transferred to TMA recipient blocks (Figure 19). The periphery of the primary tumour was selected because it was expected to be the most proliferative part of the tumour. Slides of 4 µm thickness were cut from the TMA blocks and used for IHC and ISH.

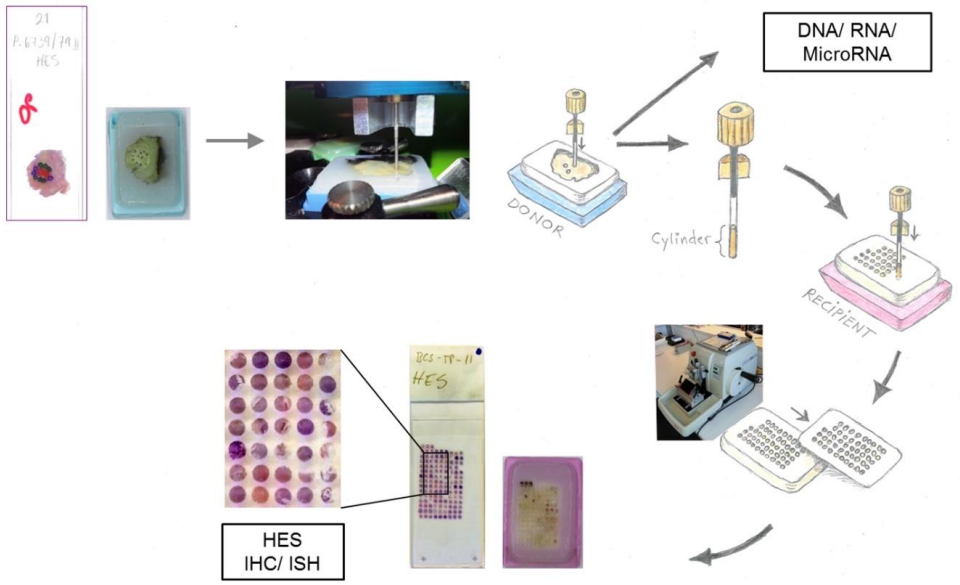


Figure 19: Illustration of TMA construction, by Linda A. Dyrnes, Breast Cancer Subtypes Research Group, NTNU. Selected areas are marked by a pathologist and cylinders from the marked areas are drawn/punched from the donor block and transferred to the recipient block. New slides are then made from the recipient block and used for HE-staining, IHC and ISH.

1.11.2 Immunohistochemistry

Immunohistochemistry is a method that uses antibodies to detect antigens in tissue (170, 171). In breast cancer, IHC is used for diagnostic, predictive and prognostic purposes (116). An antibody is an immunoglobulin molecule with complementary 3D formation to a second molecule, the antigen (171). The location where the antibody is complementary to the antigen is called the epitope. Both polyclonal and monoclonal antibodies can be used for immunostaining. Monoclonal antibodies recognize only one epitope, while polyclonal antibodies contain different antibodies with varying specificity to different epitopes of the antigen (171). In general, monoclonal antibodies have lower sensitivity, but higher specificity than polyclonal antibodies. Today, most commercially produced monoclonal antibodies are highly specific (171). Before treatment with the antibody, the tumour section is preincubated

with a compound that reduces non-specific binding of the antibody, such as normal serum. The antibody can be visualized by attaching a label. There are different types of labels, for instance fluorescent labels that can be visualized using a fluorescence microscope, labels that can be directly visualised using a light microscope such as gold and ferritin, or enzymes that produce colour through reactions with other compounds (chromogens) (170, 171). Figure 20 shows different IHC detections methods.

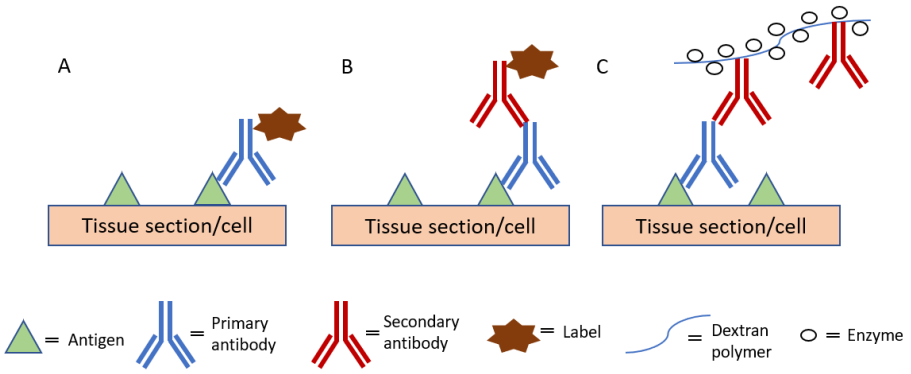


Figure 20: Illustration of IHC detection systems, by Elise Klæstad. A) Direct-conjugate method: The label is attached directly to the primary antibody, B) Indirect-conjugate method: a secondary labelled antibody detects the unlabelled primary antibody, C) secondary antibodies and enzyme molecules are attached to a polymer "backbone". This method allows for attachment of several enzyme molecules to the antigen.

The tumours in cohort 1, 2 and 4 have previously been reclassified into molecular subtypes using IHC and ISH. For paper II, protein expression of ZNF703 was assessed on tumour tissue in TMAs using IHC. For ZNF703, the tissue was stained using a monoclonal mouse ZNF703 antibody.

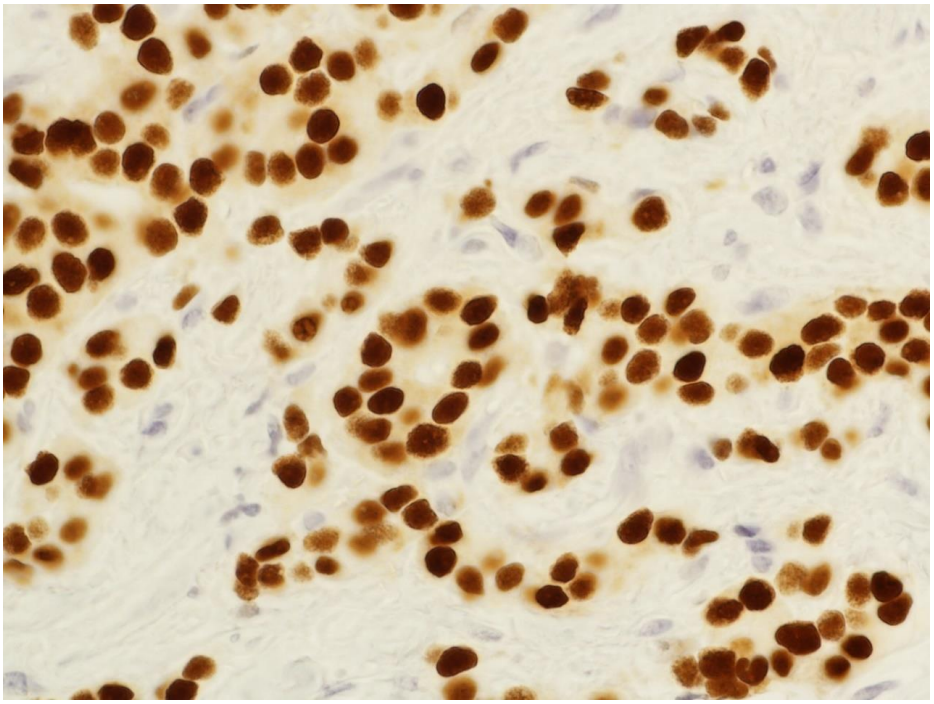


Figure 21: Tissue section form an invasive breast cancer tumour with immunostaining for oestrogen receptor (ER) (600 x magnification). Photo: AM Bofin, BCS, NTNU.

1.11.3 *In situ* hybridization

ISH can be used for copy number assessment of a segment of DNA or RNA, using probes targeting specific sequences (172). A probe is a labelled fragment complementary to the selected target sequence. The method can be used to examine copy number and chromosome rearrangements such as translocations, deletions, and chromosomal breaks. The most frequently used methods for ISH are FISH and chromogenic *in situ* hybridization (CISH). The two methods differ in the method employed for visualisation of the probes. For CISH, the probes are labelled with a chromogenic tag that can be visualised in a bright-field microscope, while for FISH the probe is labelled with a fluorescent tag or fluorochrome and detected using a fluorescence microscope. According to current breast cancer guidelines, ISH is routinely used to determine *HER2* status (56, 173). FISH was used to examine copy numbers of *MRPS23* and chromosome 17 centromere in paper I, and to examine copy numbers of *ZNF703* and chromosome 8 centromere in paper II.

1.11.4 Gene expression

Gene expression analyses have given us more insight into the heterogeneity of breast cancer (81, 82), and gene expression profiling has become a more commonly available laboratory technique. Furthermore, gene expression data from large datasets such as the TCGA (174) and METABRIC (90) datasets are now available for researchers.

1.12 Scoring and reporting

In paper I, *MRPS23* copy number was examined with FISH using probes targeting *MRPS23* and centromere 17. There are no current guidelines defining a cut-off for *MRPS23* amplification. We chose to base our scoring and reporting on current *HER2* guidelines. Thus, cases were divided into three categories based on *MRPS23* copy number: mean <4; mean ≥4 <6; and mean ≥6 (Figure 22).

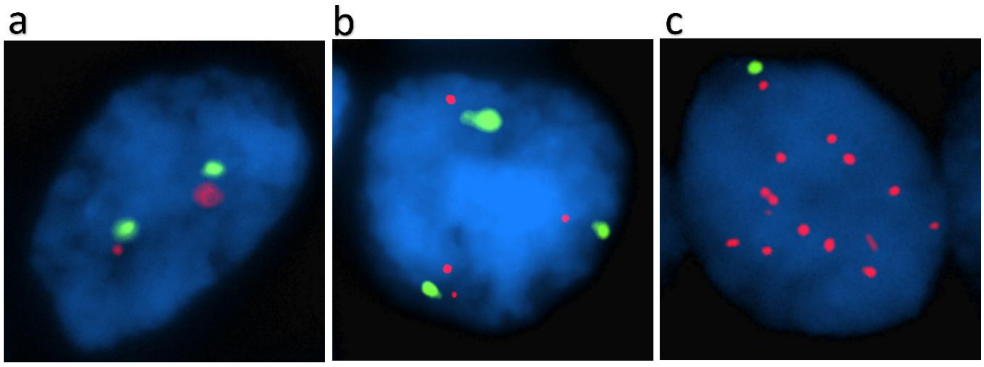


Figure 22: Breast cancer cell nucleus with a) two copies of *MRPS23* (red) and centromere enumeration probe 17 (CEP17 (green)), b) four copies of *MRPS23* and three copies of CEP17 and c) copy number increase of *MRPS23* without corresponding increase in CEP17. Klæstad E, Opdahl S, Engstrøm MJ, *et al.* MRPS23 amplification and gene expression in breast cancer; association with proliferation and the non-basal subtypes. *Breast Cancer Res Treat*. 2020;180(1):73-86. Licensed under a Creative Commons Attribution 4.0 International License.

To estimate the impact of gene/centromere ratio, cases were also divided into two categories: *MRPS23*/CEP17 <2 and *MRPS23*/CEP17 ≥2. In addition, we defined *MRPS23* amplification (*MRPS23*$^+$) as mean *MRPS23* copy number ≥6 and/or *MRPS23*/CEP17 ratio ≥2. Cases with *MRPS23* copy number <6 and *MRPS23*/CEP17 ratio <2 were defined as non-amplified (*MRPS23*$^-$). To study the prognostic value of *MRPS23* and *HER2* status combined, cases were also divided into four groups: *MRPS23*$^-$/HER2$^-$; *MRPS23*$^-$/HER2$^+$; *MRPS23*$^+$/HER2$^-$; and *MRPS23*$^+$/HER2$^+$. The prognostic influence of *MRPS23* gene expression was examined by

dividing the cases into subgroups based on gene expression quartiles. Cut-off for high expression was set as the upper quartile.

In paper II, cases were divided into three categories based on copy number status of *ZNF703*: mean <4; mean ≥4 <6; and mean ≥6 (Figure 23). Cases were divided into two categories based on the proportion of positive nuclear staining of ZNF703: <50%; and ≥50% (Figure 23). In addition, cases were divided into four categories based on copy number status and nuclear staining; *ZNF703* copy number <6 and ZNF703$^-$; *ZNF703* copy number <6 and ZNF703$^+$; *ZNF703* copy number ≥6 and ZNF703$^-$; and *ZNF703* copy number ≥6 and ZNF703$^+$.

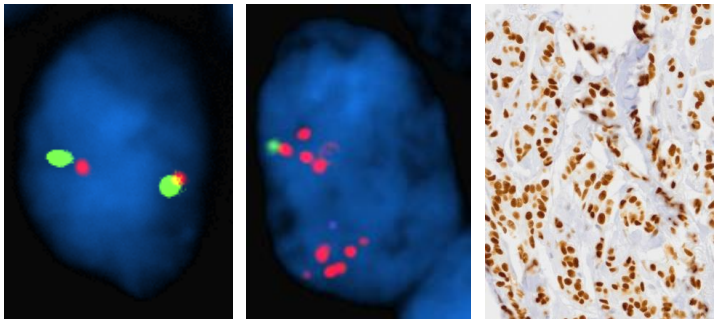


Figure 23: To the left: breast cancer cell nucleus with two copies of *ZNF703* (red) and CEP8 (green). In the middle: Breast cancer cell nucleus with increased copy number of *ZNF703* without corresponding increase in CEP8. To the right: Positive IHC staining of ZNF703 (≥50% nuclei with positive staining).

In paper III, proliferation was estimated using Ki-67 and mitoses as markers of proliferation. For Ki-67, we used two different cut-off levels: mean </≥15% and mean </≥30%. Cut-off at 15% has previously been used for subtyping in our group (97). This choice of cut-off was based on a previous study by Cheang *et al* (62), and the St. Gallen 2011 guidelines (67). Furthermore, 30% was chosen as the second cut-off based on the latest recommendations from the Ki-67 working group which states that only Ki-67 values <5% and ≥30% should be used for prognostics purposes and treatment decisions (63).

For mitotic count, we used the established thresholds for mitotic count in histological grading (42). We calculated the number of mitoses/mm^2 for each case based on the pathologists' mitotic count in 10 HPFs and the microscope's field area. The following two cut-offs were used: ≤/>3.6 mitoses/mm^2 (mitotic score 1 versus mitotic score 2 and 3), and </≥7.7 mitoses/mm^2 (mitotic score 1 and 2 versus mitotic score 3).

1.13 Statistical analyses

STATA 16.1 and 17 were used for all statistical analyses.

1.13.1 Point estimates, p-values and confidence intervals

Three important terms in statistical analyses are point estimates, p-values, and confidence intervals. The point estimate is a measure of the association between the outcome variable and the exposure variable. Point estimates can be in the form of a difference or a ratio, for instance mean difference or incidence rate ratios (175, 176). The p-value is the probability of observing this point estimate or a more extreme value, given that the null hypothesis is correct. The null hypothesis generally assumes that there is no association between variables being examined, or no difference between groups being compared (177). The null hypothesis is tested against the alternative hypothesis which assumes that there is a difference. The confidence interval (CI) describes the precision of the point estimate (176). By convention, the CI is set to 95%, which means that if a study was repeated several times, the confidence interval would contain the true value in 95% of the cases. A wide CI indicates low precision (176, 178). In our analyses statistical significance was set at the 5% level. Values between 5-10% were considered borderline significant.

1.13.2 Sources of error in epidemiological studies

There are three main sources of error one has to consider in the interpretation of results in epidemiological studies; information bias, selection bias and confounding (177).

As a common term, bias is a deviation from the truth that leads to an incorrect measure of an association (179). Information bias and selection bias are the two main types of bias. Information bias is also referred to as misclassification and can occur as a consequence of measurement errors. Selection bias occurs when the association between exposure and outcome in the sample selection differ from the population intended to study.

In this thesis we used data from four different cohorts, with different inclusion methods. Cohort 1, 3 and 4 include women followed for breast cancer occurrence, while cohort 2 includes tumour tissue from breast cancer patients collected in the UK and Canada. In cohort 1 all invited women were included in the study and followed up for breast cancer, irrespective of participation in the baseline survey. Whereas cohort 3 was restricted to women who participated in the baseline survey. In cohort 4 women were included on the basis that they were born at E. C. Dahl's foundation during a specific time-period. Inclusion was

therefore not based on active participation. One could argue that people who actively participate in a study differ from the general population in terms of health awareness, socioeconomic status and so on. Thus, cohort 3 may be subject to selection bias.

In our studies FISH and IHC analyses were performed on TMAs. Even though TMAs are considered representative for whole sections, this method may have led to misclassification of tumours, for example if a hotspot of Ki-67 was missed in the TMA, and consequently a high proliferative tumour was classified as low proliferative.

1.13.3 Pearson Chi-Squared test

Pearson chi-squared test was performed to assess the association between our selected biomarkers and different patient and tumour characteristics.

The Pearson chi-squared (χ^2) test compares the expected and the observed values presented in a contingency table, usually 2x2 tables, and obtains a p-value. The null hypothesis assumes that there is no association. The larger the difference between the observed values and the expected values, the lower the p-value. The test is preferably used on larger sample sizes, and should not be used if any of the expected values in the contingency table are below 5 (176).

1.13.4 McNemar and marginal homogeneity test

Marginal homogeneity test (paper I and II) and McNemar test (paper I) was used to compare copy number status of *MRPS23* and *ZNF703* in the primary tumours and in the corresponding lymph node metastases.

The two tests can be used to investigate association between matched pairs, such as primary tumours and their corresponding lymph nodes. McNemar test is used on variables with two categories, while marginal homogeneity test is preferred on ordered variables with more than 2 categories, such as our categories of mean *ZNF703* (<4, $\geq 4 < 6$, ≥ 6). The two tests examine whether the distribution of observations to the rows and columns in the contingency table are proportionally equal, for example in our study we examined whether the proportion of primary tumours with *MRPS23* amplification was the same as the proportion of lymph nodes with *MRSPS23* amplification. In light of our example, a p-value below the defined significance level means that there is a difference between the proportion of primary tumours and the proportion of lymph node metastases with *MRPS23* amplification (176).

1.13.5 Survival analysis

The aim of survival analysis is to estimate how much time passes before a given event occurs. Survival analyses can for example be used to study time to disease, effect of treatment or time until death. In survival analysis, information about time is used to estimate and compare survival and/or hazard functions, and to assess the relationship between explanatory variables and survival time. In principle there are two outcomes in survival analyses, occurrence of the event of interest or censoring. Censoring occurs when participants do not experience the defined event during their time of follow-up. The most common reasons for censoring are loss to follow up, withdrawal from the study, or the event not occurring during the study period (176, 177). We performed survival analyses in paper I and paper II, where death from breast cancer was defined as our occurring event. Participants were censored due to emigration or end of follow up. Death from other causes was considered a competing risk in our analysis of cumulative incidence of BC death. Competing risk is an event whose occurrence prohibit the occurrence of the event of interest (176). For instance a patient who has already died from a heart attack cannot die from breast cancer (176).

1.13.5.1 Kaplan Meier method

By registering time to the event of interest and the number of participants at risk of an event, the Kaplan Meier (KM) method can be used to estimate the probability of surviving up until a certain time. Censoring of a participant leads to a reduction of the number of participants at risk (176). The KM method gives a graphical display of the survival function. The KM plot is presented as a step function, starting at 1 before any event has occurred, followed by decreasing survival probability as time passes. The method assumes that competing risks are independent and handles competing events by censoring them. It is impossible to determine whether a person who has died from a heart attack would have died from breast cancer if she had not died from the heart attack. Therefore, the KM method could lead to biased estimates if the censored subjects are more, or less, likely to experience the defined event than uncensored subjects. Hence, the KM method is preferably used for analysis with only one possible event (175).

1.13.5.2 Cumulative incidence of breast cancer death

In paper I and II, we calculated cumulative incidence of breast cancer death for subgroups of patients defined by the biomarkers MRPS23 and ZNF703. We used cumulative incidence of

breast cancer death instead of KM because we wanted to include death from other causes as a competing risk.

Cumulative incidence curves are an alternative to the KM method when competing risks are present. In contrast to KM there is no assumption of independence between competing events in cumulative incidence curves. The cumulative incidence of death from breast cancer F(t) can be interpreted as the risk of dying from breast cancer at a certain point, given that you have not died from other causes until that time (175). In cumulative incidence curves, the time from 0 to t is divided into small intervals, and the cumulative incidence equals the sum of the probability of the event occurring within a specific time interval (175, 180). Gray's test was used to compare equality between cumulative incidence curves (175).

1.13.5.3 Cox Proportional Hazard Model

Cox proportional hazard models were used in papers I and II to estimate hazard ratios of breast cancer death in subgroups of patients according to *MRPS23*-status (paper I), *MRPS23*/HER2-status (paper I) and *ZNF703* copy number status and protein expression (paper II). Where applicable, adjustments were made for the following factors: age (≤49, 50-59, 60-64, 65-69, 70-74, ≥75), stage (I-IV), histological grade (I-III), Ki-67 status (</≥15%), and HER2 status. In addition, we calculated hazard rates for subdivisions of molecular subtypes according to proliferation status and birth year (paper III).

Cox regression analyses can be used to estimate the effect of defined risk factors on survival time (177). The hazard rate can be interpreted as the risk of the occurrence of an event at a time *t*, for instance breast cancer death, given that the subject has survived until *t* (177). Using Cox proportional hazard models, one can compare the hazard rate in one group to the hazard rate in a designated reference group (176). While cumulative incidence functions provide an absolute measure of the risk of death, the Cox proportional hazard model provides a relative risk of death associated with one factor compared to another.

In Cox regression analyses there is an assumption that the effect of the risk factors is constant over time, i.e., the proportionality assumption (176, 177). We performed log-minus-log plots to test for proportionality, and the proportionality assumptions were met in all studies.

1.13.6 Incidence rates

In paper III, breast cancer incidence rates were used to examine time trends according to proliferation markers Ki-67 and mitotic count, age, and birth year.

An incidence rate is a measure of new incidents of a defined event in a population during a given period of time (177). Incidence rates are calculated as the number of new incident cancers divided by the combined observation time for all participants at risk in the same time period (177).

We estimated age-specific incidence rates for women born before 1929 and for women born in 1929 or later, to examine the change in incidence of breast cancer with high and low proliferation status over time. Poisson regression was used to estimate incidence rate ratios with 95% CI, to compare incidence rates between women born in 1929 or later to women born before 1929. Similar analyses of incidence rates and incidence rate ratios were done for $HER2^+$ and TN tumours to examine if there had been a change in incidence according to proliferation status within the molecular subtypes.

Some of the incident breast cancers had missing values of the proliferation markers. Ki-67 status was missing in 920 (31%) of the tumours and mitotic count was missing in 886 (30%) of the tumours. The majority of the cases with missing values were diagnosed at other hospitals and were therefore unavailable for analysis. Others had missing values because tumour blocks were not found, IHC staining was unsuccessful, or there was limited amount of tumour tissue available. To prevent underestimation of incidence rates due to missingness, multiple imputations were performed.

1.13.7 Missing data

Missing data can occur in research and must be accounted for to prevent biased estimates and loss of information. There are three types of missing data:

1. Missing completely at random – when the missing values are missing completely at random. In other words, the probability that an observation is missing is not related to any other patient characteristic. An example could be if a TMA-slide containing random tumour tissue samples was lost.
2. Missing at random – the probability that a value is missing depends on other observed patient characteristics. In other words, any systematic difference between the missing values and the observed values can be explained by differences in the observed dataset. In paper III, the majority of cases with missing values were missing because they were diagnosed at other hospitals. Patients diagnosed at other hospitals were on

average a little younger compared to patients diagnosed at St Olav's hospital. To compensate for this, age was included in the imputation model.

3. Missing not at random – the probability that a value is missing depends on unobservable data, for instance the missing value itself. Values missing not at random cannot be accounted for and could lead to loss of valuable information (181, 182). An example of missing not at random could be if high-proliferative tumours were more frequently unsuccessfully stained, and therefore had missing values.

1.13.8 Multiple imputations

In paper III we used multiple imputations to predict mitotic count and Ki-67 status for tumours with missing values for these variables.

Multiple imputation is a multistep method for replacing the missing data values of a variable, using estimates based on the observed data. Missing values are replaced with values based on findings in patients with similar characteristics. For instance, in our study factors such as patient age, tumour size and survival status could help predict a missing value. Several imputed data sets are created, each with different estimates of the missing values to reflect the possible distribution of the missing variable. The imputed datasets are then analysed and averaged together to give overall estimated associations (181, 182).

We included the following variables in our imputation model: age (5-year categories) and calendar year at diagnosis (continuous), stage (I-IV, unknown), extent of disease (localized to the breast, local invasion, regional lymph nodes, distant lymph nodes or organ metastases, metastases detected, unknown) as reported by the Cancer Registry of Norway, year of birth (5-year categories), follow-up time after diagnosis (log transformed, continuous) and survival status (alive, death from breast cancer, death from other causes). Based on 50 imputed datasets, we calculated imputed incidence rates and incidence rate ratios according to birth year and age.

Imputation was performed assuming that values were missing at random (181, 182). The major reason for missing values of Ki-67 and mitotic count was that cases were diagnosed at other hospitals and therefore unavailable for analysis. A comparison of cases with missing values, and cases with known proliferation status showed that cases with missing values were on average a few years younger at diagnosis compared to cases without missing values, and there was a higher frequency of cases diagnosed in the 60's and 70's. Inclusion of age and year of diagnosis in the imputation model should have compensated for

this to some extent. Nevertheless, imputed results are only estimates of associations and we cannot know how well they describe the true associations.

Summary of results

1.14 Paper I

In paper I, the aims were to examine *MRPS23* copy number status in breast cancer, and to assess possible associations between *MRPS23* copy number status and molecular subtype, proliferation, and prognosis, and between *MRPS23* gene expression and molecular subtypes and prognosis.

We found *MRPS23* amplification (mean *MRPS23* copy number ≥6 and/or *MRPS23*/CEP17 ratio ≥2) in 8% of primary tumours. *MRPS23* amplification was found in all subtypes, except 5NP and BP, and amplification of the gene was associated with the luminal B (HER2-) subtype. We found that the proportion of HER2+ cases was larger in the *MRPS23* amplified group compared to the non-amplified. There was an association between *MRPS23* amplification and high Ki-67, high mitotic count, and high histologic grade. However, we found no significant association between *MRPS23* copy number and prognosis. We also divided the cases into four groups based on *MRPS23* and *HER2* copy number status. We found that *MRPS23* amplified cases with positive HER2 status had the poorest prognosis.

Using the METABRIC dataset we found that high gene expression of *MRPS23* was associated with the luminal B subtype. We found no association between *MPRS23* gene expression and prognosis.

1.15 Paper II

In paper II, the aims were to assess *ZNF703* copy number and ZNF703 protein expression in primary breast cancer tumours. In addition, we aimed to investigate associations between *ZNF703* copy number and molecular subtypes, proliferation, and prognosis, between *ZNF703* copy number and ZNF703 protein expression, and between ZNF703 protein expression and prognosis.

We identified high *ZNF703* copy number (copy number ≥6) in 7% of primary tumours. We found an association between high *ZNF703* copy number, and high proliferation, luminal B subtypes and high histological grade. When divided into groups according to mean copy number, we found that patients with mean copy number ≥6 had a cumulative risk of death from breast cancer of 48% (95% CI 35-63%) after ten years, compared to 32% (95% CI 28-36%) in patients with copy number <4. Cox regression analyses also showed a higher rate of death in patients with high copy number (≥6) compared to patients with copy number <4 (HR 1.6 (95% CI 1.1-2.5)). Within the luminal A subtype,

we found that high *ZNF703* copy number was associated with higher histological grade and poorer prognosis compared to cases without copy number increase (borderline significant). We found no association between *ZNF703* copy number and prognosis among luminal B tumours. We found positive nuclear staining (≥50%) in 76% of the primary tumours. Positive IHC staining was found across all subtypes and categories of *ZNF703* copy number. We found a positive correlation between copy number status and protein expression, but no association between protein expression and prognosis.

1.16 Paper III

In paper III, the aim was to examine long-term trends of breast cancer incidence according to proliferation, using Ki-67 and mitotic count as markers of proliferation. We performed incidence analyses for all breast cancers cases combined and did separate incidence analyses for $HER2^+$ and triple negative breast cancers. Multiple imputation was used to replace missing values of Ki-67 and mitoses.

In the incidence analyses, we compared women born in 1929 or later, to those born before 1929. We found that there had been an increase in overall breast cancer incidence in the age groups 40 to 69 years among women born in 1929 or later. Rates for imputed and observed incidence rates followed the same patterns, with imputed rates being higher than the observed rates, as expected. Using Ki-67 as a measure of proliferation, we found and increase in both low- and high-proliferative tumours at 15% and 30% cut-off, in women aged 40-69 years, who were born in 1929 or later. Using mitotic count, we found an increase in low-proliferative tumours at cut-offs ≤3.6 mitoses/mm^2 and <7.7 mitoses/mm^2 in the age groups 40-69, among women born in 1929 or later. There was no increase in incidence of high-proliferative tumours neither at >3.6 mitoses/mm^2 nor at ≥7.7 mitoses/mm^2 cut-off. We found that mean values of Ki-67 were higher among women aged <50 and >75, and lower among women aged 50-74 when we compared women born in 1929 or later to those born before 1929. Mean values of mitoses/mm^2 were higher in the age groups <50 years and lower in the ages >50 for women born in 1929 or later. According to Ki-67 we found that there had been an increase in high-proliferative $HER2^+$ tumours among women born in 1929 or later. While according to mitotic count there had been an increase in low-proliferative $HER2^+$ tumours. There was no change in incidence rate according to proliferation markers in triple negative breast cancers.

Discussion

1.17 Discussion of main findings:

1.17.1 Paper I and II

In diagnostic pathology, proliferation in breast cancer tumours is currently evaluated using Ki-67, mitotic count, and the PAM50 test (116). In Paper I and II, we studied copy number of two genes suggested to play a role in breast cancer proliferation: *MRPS23* and *ZNF703*. Furthermore, we studied gene expression of *MRPS23* (Paper I), and protein expression of ZNF703 (Paper II), and associations between these markers, and molecular subtypes, proliferation, and prognosis in breast cancer.

Gatza *et al*. identified *MRPS23* as a driver of proliferation in breast cancer, and when amplified, it was associated with a poor prognosis. The gene was found to be uniquely amplified in highly proliferative, luminal (non-basal) subtypes (113).

Defined by copy number analyses, Gatza *et al.* found amplification of *MRPS23* in 20% and 33% of patients in the two datasets included, while we found amplification in 8% of cases using FISH. Like Gatza *et al.*, we found an association between *MRPS23* amplification and high proliferation. Gatza *et al.* assessed proliferation using the PAM50 gene expression test, while we used Ki-67 and mitotic count as measures for proliferative activity. Furthermore, we confirmed their finding that *MRPS23* was uniquely amplified in non-basal tumours. Gatza *et al*. classified the tumours into molecular subtypes by gene expression analyses. We used FISH and IHC, reclassifying our tumours into six molecular subtypes. In contrast to Gatza *et al*, we did not find an association between copy number increase and a poor prognosis. Different methods for molecular subtyping and copy number assessment, different definitions of amplification and assessment of proliferation can possibly partly explain the diverging results.

In our study, *MRPS23* gene expression in luminal tumours (luminal A, luminal B and HER2-type tumours defined by PAM50) was analysed using data from the METABRIC dataset. We divided cases into quartiles of gene expression to investigate associations with prognosis. We found no difference in prognosis between tumours in the upper gene expression quartile compared to the rest. Gatza *et al.* made no assessment of the association between *MRPS23* gene expression and prognosis, however they found an association between gene copy number and gene expression, and between copy number and prognosis (113). We found that increased gene expression was associated with the luminal B subtype. However, we found overlapping levels of gene expression between the different subtypes, even between

the luminal B and basal-like subtypes, the latter known to be non-amplified. Transcription is regulated by several mechanism, and there is often a discrepancy between gene copy number and gene expression (183). The lack of correlation between gene expression levels and prognosis could possibly be explained by other transcriptional regulatory mechanisms.

In our analyses we found that 42% of the *MRPS23* amplified tumours were $HER2^+$ compared to 11% of the non-amplified tumours. Since $HER2^+$ is an important prognostic marker in breast cancer we wanted to examine prognosis in subgroups of patients according to *MRPS23* amplification and HER2-status. We found that the patients with $HER2^+$, *MRPS23* amplified tumours had the poorest prognosis, however, the results were not significant.

In breast cancer, *ZNF703* have been suggested as a driver gene in the A1 amplicon in the chromogenic region 8p11-12. In paper II, we found high copy number (≥6) of *ZNF703* in 7% of primary tumours. This is in accordance with another study which found amplification in 8% of breast tumours (133). We found *ZNF703* copy number ≥6 in 7.5% of ER^+ tumours, and in 3% of ER^- tumours, while others have found *ZNF703* amplification in 9-19% of ER^+ tumours (135, 137) and in 2% of ER^- tumours (135). We also found that *ZNF703* copy number increase was most frequent in luminal B tumours, in accordance with other studies (135, 137).

Gelsi-Boyer *et al.* identified four amplicons in the 8p12 amplicon (A1 – A4), and 14 genes which showed a correlation between amplification and overexpression, among them *ZNF703* (134). The A1 amplicon was located as the most telomeric of the four amplicons, and analyses showed that it did not include the centromere. Holland *et al.* later confirmed that the A1 amplicon does not include the centromere (135). Studying CEP8 using FISH, we found that of the cases with *ZNF703* copy number ≥6, only 2% had CEP8 ≥6. Thus, our results indicate that an increase in *ZNF703* copy number is not accompanied by an increase in centromere 8 copy number. This is in accordance with previous findings.

Holland *et al.* found that *ZNF703* regulates proliferation in human luminal breast cancer cell lines (135), and Sircoulomb *et al.* found a positive association between ZNF703 protein overexpression and increased proliferation in human luminal breast cancer cell lines (141). We found that high copy number of *ZNF703* was associated with increased proliferation in breast cancer, measured by Ki-67-level and mitotic count.

Using copy number analyses, Holland *et al.* found that amplification of *ZNF703* was associated with a poor prognosis in ER positive tumours (135). Sircoulomb *et al.* found that high level of *ZNF703* mRNA was associated with poor survival in luminal tumours (A and B) (141). Using FISH, we found that high copy number of *ZNF703* was associated with poor

prognosis. Separate analyses for luminal A and luminal B showed poor prognosis among luminal A cases with high copy number (borderline significant), but no association between *ZNF703* copy number and prognosis among luminal B cases. We also found that luminal A cases with high copy number of *ZNF703* had higher grade than those without copy number increase (Table 5).

Table 5: Histologic grade and mean *ZNF703* copy number in Luminal A tumours

	Mean copies of *ZNF703* (%)				
	<4	**≥4 <6**	**≥6**	**Total**	χ^2
Grade (%)					
I	67 (21)	1 (6)	0	68	p <0.001
II	228 (70)	9 (56)	4 (44)	241	
III	29 (9)	6 (38)	5 (56)	40	
Total	324	16	9	349	

It can be debated whether luminal A tumours with grade 3 in fact should be classified as luminal A. Most luminal A tumours are grade I or II (97). According to the 2015 St. Gallen Consensus, luminal A are characterized by low proliferation and good prognosis, and they are defined by high ER/PR, low Ki-67, small tumour size and rarely involvement of lymph nodes (49). In contrast, luminal B tumour are characterized by high proliferation and poor prognosis, and they are defined by low ER/PR and high Ki-67, histological grade 3 and more often affected lymph nodes. In addition, a third group with intermediate scores of molecular markers, with uncertain prognosis and response to endocrine therapy was also mentioned (49). The classification of all grade 3 luminal tumours as luminal B subtype has also been used in molecular subtyping algorithm by others (94). Thus, luminal A cases with high *ZNF703* copy number could represent a more aggressive group of cancers than luminal A tumours in general, or a group of misclassified luminal B tumours.

Furthermore, Holland *et al.* and Sircoulomb *et al.* used gene expression analyses to classify molecular subtype, and copy number analyses to determine *ZNF703* status. In our analyses, subtypes were classified using IHC and ISH, while *ZNF703* copy number was assessed using FISH. Thus, different methodologies could also explain the diverging results.

1.17.2 Paper III

Studies from Europe and the US show that there has been an increasing incidence of ER^+ tumours, and a decrease in ER^- tumours the last decades (15-17, 184). An increase in luminal A and luminal B ($HER2^-$) tumours has previously been described by our group (107). The previous increase in ER^+ tumours has partly been attributed to the use of MHT and mammography screening (9, 10). In addition, changing patterns of risk factors may also affect the observed changes in incidence trends.

In paper III, we studied long-term trends in incidence of high and low-proliferative breast cancers, using Ki-67 and mitotic counts as markers of proliferation. In our analyses, we compared age-specific incidence rates among women born in 1929 or later to women born before 1929. According to Ki-67, we found that there had been an increase in both high and low-proliferative tumours, at both 15% and 30% cut-off, among women in the ages 50-69 years born in 1929 or later. According to mitotic count we found that there had been an increase in low-proliferative breast cancers, but no increase in high-proliferative tumours. We did separate analyses for $HER2^+$ and TNBCs. For $HER2^+$ tumours we found an increase in high-proliferative tumours according to Ki-67 for women born in 1929 or later, while there was an increase in low-proliferative tumours according to mitotic count. There was no change in incidence according to proliferation makers in TNBCs.

The mammography screening program was implemented in the county of Trøndelag in 2001, and since then all women aged 50-69 years have been invited to screening every other year. Due to age at implementation of the program the participants in cohort 1 were not included in the screening program, while most women in cohorts 3 and 4 were. Screening detected cancers are more frequently luminal A subtype compared to interval cancers (19, 20). In addition, they often show other favourable characteristics such as lower grade and less often high expression of Ki-67 compared to cancers detected outside of screening (19, 20). MHT was frequently used in Norway from the late 1980s until 2002, when the sale decreased due to the results from the Women's Health Initiative studies (9). Thus, MHT and mammography screening may have affected the observed increase in incidence of low-proliferative tumours.

Studies on risk factors associated with the different molecular subtypes of breast cancer have found that most established breast cancer risk factors, such as early menarche, low parity, and menarche at old age are associated with the low-proliferative luminal A subtype (30). In Norway there has been a steady decrease in the number of births per woman since the 1960s (185). Due to the association between reproductive patters such as low parity, and the

low-proliferative Luminal A subtype, changes in reproductive patterns could partly explain the increase in low-proliferative tumours found in our study. Lifestyle related factors are also associated with increased risk of breast cancer – for instance obesity and overweight, inactivity and excessive alcohol consumption. Obesity and overweight have become an increasing problem in Norway the last 40-50 years, and among women in the age group 40-49 years approximately 1 in 4 are obese (186). Thus, several risk factors may have affected the change in breast cancer incidence according to proliferation.

In the analyses of Ki-67 we used the two different cut-offs: </≥15% and </≥30%. Mitotic count was divided using the following cut-offs: ≤/>3.6 and </≥7.7 mitoses/mm^2. The distribution of Ki-67 and mitotic count are presented in Table 6 (numbers not presented in paper III).

Table 6: Distribution of Ki-67 and mitotic count

	Mitoses/mm2 (%)				
	≤3.6	**>3.6 <7.7**	**≥7.7**	**Total**	χ^2
Ki-67 (%)					
<15	972 (77)	175 (38)	41 (12)	1188	p<0.001
≥15 <30	241 (19)	180 (39)	100 (30)	521	
≥30	51 (4)	108 (23)	188 (58)	347	
Total	1264	463	329	2056	

We found that there was good correspondence between low Ki-67 and low mitotic count as well as high Ki-67 and high mitotic count. However, in the mid-range the mitotic count and corresponding Ki-67 levels varied. A study on HR$^+$/HER2$^-$ breast cancers comparing Ki-67 and mitotic index on whole sections found discrepant results (high MI/low Ki-67 or low Ki-67/high MI) in 15,9% of the tumours (187). In this study mitoses were counted in 10 HPFs in the peripheral most cellular region of the tumour slide and scored from 1 to 3 points according to Elston and Elli's thresholds for mitotic count used in histologic grading. A score equal to 1 or 2 were considered low mitotic index, while a score ≥3 was considered high mitotic index. Ki-67 was considered high if >20% of nuclei within the area of highest positivity showed positive staining (187). Discrepancy between mitotic count and Ki-67 levels may be affected by several factors. For instance, mitoses can only be assessed during M-phase, while Ki-67 is expressed through all phases of cell cycle, except the G0 (117). This may explain some of the cases with high Ki-67 and low/intermediate mitotic

count. Furthermore, aspects regarding material and methodology may affect the results. For instance, in our study Ki-67 was assessed on TMAs while mitotic count was done on whole sections. Thus, we may have missed the true Ki-67 hotspot in some cases. This may partly explain cases with low Ki-67 and higher mitotic count. Reduced antigenicity of Ki-67 because of prolonged storage of FFPE blocks could have led to underestimation of the true Ki-67 level. Reduction of Ki-67 antigenicity due to prolonged storage would have led to an underestimation of Ki-67 levels in the oldest tumours, and consequently an underestimation of the increase in low proliferative tumours and an overestimation of the increase in high proliferative tumours. A third factor that could have led to discrepant results between mitotic count and Ki-67 levels is interobserver variability. To minimize the risk of misclassification, markers were assessed by two observers. In case of discrepant results, cases were discussed, and consensus was reached.

1.18 Methodological considerations, strengths, and weaknesses

1.18.1 Reliability of results

The studies included in this thesis are based on four large cohorts of breast cancer patients. Cohorts 1, 3 and 4 are all large cohorts of Norwegian women with long term follow-up for breast cancer occurrence, and breast cancer cases with long follow-up after their primary diagnosis. Breast cancer can relapse several years after the primary diagnosis (97, 107), therefore data with long follow-up after diagnosis is of great value in breast cancer research.

Data on the cases have been linked with data from high-quality, national registries providing information about cancer incidence and outcome. According to law, all incident cancers must be reported to the National Cancer Registry, providing a close to complete overview of cancer incidence in Norway (160).

All laboratory work was performed by highly experienced biomedical engineers, at the same laboratory, using the same antibodies, and the same subtyping algorithm for all cases. All IHC markers were assessed by two independent observers. In case of interobserver disagreement, consensus was reached.

Cohort 2, the METABRIC dataset, is a dataset containing copy number and gene expression data from approximately 2000 tissue samples with long-term clinical follow-up data available (90). It is widely used in international breast cancer research (188).

1.18.2 Characteristics of study population

Cohorts 1, 3 and 4 comprise women from the Norwegian county Trøndelag. The women in cohort 1 were born between 1886 and 1928, and they were followed for breast cancer occurrence from 1961 to 2008 (97). Mean age at diagnosis in the cohort was 70.7 years. National identity numbers were introduced in Norway in 1961, making it possible to follow breast cancer patients after diagnosis through linkage to national registries. Therefore, women diagnosed with breast cancer before 1961 were not included, and it is likely that cancers diagnosed at younger ages are underrepresented in this cohort. The women in cohort 1 were not invited to participate in the Norwegian mammography screening program, due to their age at implementation of the program in 2001. Screening programs lead to earlier detection of breast cancer tumours, and tumours diagnosed through screening are more frequently smaller in size, lower grade, and luminal A subtype compared to tumours discovered symptomatically (19, 21). Furthermore, most of the participants in cohort 1 were followed for breast cancer occurrence until time of death. All these factors could explain the high age at diagnosis in cohort 1.

Cohort 3 comprises women born from 1897 to 1977, who were followed for breast cancer occurrence from inclusion in 1995-97 to 2009 (107). Mean age at diagnosis was 63.2 years. The majority of the participants were born after 1940, thus most patients were followed from a relatively young age. Cohort 4 comprises women born between 1920 and 1966. These women were followed for breast cancer occurrence from 1961 until 2012 (155). Mean age at diagnosis was 54 years. The majority of these women were born after 1945. Compared to cohort 1, the other two cohorts represent a far younger group of women. Most of the participants in cohorts 3 and 4 would have been invited to attend the screening program.

Exposure to risk factors and treatment will have varied for the women included in these three cohorts. For instance, due to patient age and time of diagnosis, treatment of patients in cohort 1 was largely restricted to surgery, while breast cancer patients in cohort 3 and 4 may have received additional adjuvant treatment. Since most patients in cohort 1 were diagnosed prior to modern breast cancer treatment, this gives us the opportunity to follow the near natural course of the disease after surgery.

Cohort 2 comprise nearly 2000 breast cancer tumours from the widely used, international dataset METABRIC. Mean age at diagnosis was 63.2 years. The ER^+ and lymph node-positive patients were treated with chemotherapy, while most of the ER^- and/or lymph node-negative patients were not. None of the $HER2^+$ patients received anti-HER2 treatment.

1.18.3 Storage, fixation, and staining

In a breast cancer study, it is nearly impossible to combine long-term follow up and full control over preanalytical conditions. The present studies are based on archival tumour tissue collected over a period of 50 years, thus preanalytical conditions and storage practices for FFPE blocks may have varied. Studies have shown that variations in preanalytical conditions can affect IHC and FISH results (56, 189). Such variations can occur between samples from the same diagnostic period, or they can be due to time-dependent factors. For instance, delay of fixation of a tissue sample could affect immunostaining. Time dependent factors could be due to changes in the buffers and reagents used during the fixation process (189). In general, FFPE tissue blocks are considered to preserve antigen expression through longer periods of storage, and a review published in 2011 stated that the duration of storage of FFPE tumour tissue did not affect IHC. The conclusion was based on four studies that had examined storage periods ranging from 2-25 years (189). Nevertheless, a more recent study on FFPE breast tumours stored for several decades found that immunostaining intensity of ER, HER2 and Ki-67 decreased with tissue age, and that the decrease was most prominent for Ki-67 (190). In our study material the oldest tumours were diagnosed in the early 1960s.

In paper III, we found tumours with high expression of Ki-67 across all storage periods. However, in the incidence analyses we found an increase of high-proliferative tumours according to Ki-67 status, but not according to mitotic count. We cannot exclude that the observed difference in incidence trend could be partly explained by reduced antigenicity of Ki-67.

We found poor FISH signals in some of the TMAs containing tissue from the oldest tumours. The tumours from more recent time periods (≥1980) largely displayed good visualization, hence the variations in probe uptake are most likely due to preanalytical conditions. Therefore, we restricted analyses in paper I and II to TMAs containing tumours diagnosed primarily after 1980.

Studies have shown that storage of tissue slides reduces antigenicity, and preferably new slides should be made for every analysis. However, every time a new slide is cut, tumour material is lost. Thus, there is a trade-off between loss of tumour material and loss of antigenicity due to storage. In our study, serial sections of slides were cut and stored in a light proof freezer at -20 degrees, and assessment of markers used in subtyping were evaluated within days to weeks after staining.

Both IHC and FISH analyses were done on TMA sections. Three tissue cores were drawn from each tumour. Even though TMAs can be representative for the corresponding

whole section, heterogeneity within the tumour could be a challenge and it would be of interest to validate our findings in whole sections. For instance, Ki-67 hotspots that were present on the whole section of the tumour but missing on the TMA could have led to misclassification of luminal B as luminal A tumours, or lack of correspondence between mitotic count and Ki-67, as preciously described. Nevertheless, assessment on whole sections leads to greater loss of tumour tissue and is more costly and time consuming compared to TMA assessment.

1.18.4 Scoring and reporting

A challenge in both paper I and II were deciding on cut-offs for amplification and high copy number, and whether to include ratio in the subclassification or not. In paper I, we chose to include ratio in the definition of subgroups, while we chose to only consider gene copy number for the *ZNF703* subgroup classification in paper II. There are no current guidelines as to how amplification of these genes should be defined. *HER2* status is defined by both mean copy number and ratio. In addition, *HER2* is located on the same arm of chromosome 17 as *MRSP23*, and we therefore chose to use current HER2 guidelines to define subgroups of *MPRS23.* However, studies have shown that the majority of cases with average *HER2* copy number >6 and ratio <2 have *HER2* amplifications that include the centromere, and that true polysomy of chromosome 17 is a rare event (191, 192). Hence, one could argue that ratio has no place in the definition of *HER2* status. We examined the association between *MRPS23* copy number and CEP17 copy number and found that of cases with *MRPS23* mean copies ≥4 82% had CEP17 mean copies <4 (Table 7). Thus, copy number increase of *MRPS23* was rarely accompanied by increase in CEP17 copy number, and one could argue that it may not be necessary to include ratio in the definition of *MRPS23* amplification. If *MPS23* amplification had been defined solely as copy number ≥6, we would have had a total of 21 amplified cases, compared to 45 when ratio ≥2 was included. Our definition may therefore have overestimated the number of amplified cases.

Table 7: Association between *MRPS23* copy number and CEP17 chromosome copies in primary tumours

	Mean *MRPS23*				
	<4	**≥4<6**	**≥6**	**Total**	χ^2
Mean CEP17					

<4	535 (99)	24 (83)	17 (81)	576	p<0.001
≥4 <6	4 (1)	5 (17)	3 (14)	12	
≥6	1 (0)	0 (0)	1 (5)	2	
Total	540	29	21	590	

ZNF703 is located in the A1 amplicon, in the chromogenic region 8p11-12 (134). The amplicon does not include the centromere of chromosome 8 (134, 135, 141). Using FISH, we found that 88% of cases with *ZNF703* mean copies ≥4 had CEP8 mean copies <4 (Table 8). Thus, our results were in accordance with previous findings, indicating that increase of *ZNF703* was not accompanied by an increase of centromere copies. Therefore, for *ZNF703*, we only took mean copy number into consideration in our definition of subgroups.

Table 8: Association between *ZNF703* copy number and CEP8 chromosome copies in primary tumours

	Mean *ZNF703*				
	<4	**≥4<6**	**≥6**	**Total**	χ^2
Mean CEP8					
<4	587 (97)	44 (86)	43 (90)	674	p <0.001
≥4 <6	16 (3)	6 (12)	4 (8)	26	
≥6	0	1 (2)	1 (2)	2	
Total	603	51	48	702	

There are no established guidelines for assessment of ZNF703 IHC. We chose not to include staining intensity in our analyses and defined <50 and ≥50% positive nuclear staining of ZNF703 as cut-off. We found positive nuclear staining in > 75% of primary tumours, and in 70% of tumours with *ZNF703* copy number <4. Due to the high proportion of positive cases, one could argue that the cut-off for positive nuclear staining for ZNF703 should be set higher than 50%.

In paper III, we examined proliferation in breast cancer using mitotic count and Ki-67 as markers of proliferation. Mitotic count was recalculated from mitoses/10 HPF to mitoses/mm2. This is in accordance with current Norwegian guidelines (35), and in guidelines used for histological grading (42). Recalculation into mitoses/mm^2 accounts for differences in visual field diameter between different microscopes (46). In our study, cut-offs for mitotic

count were based on the well-established cut-offs for mitotic scores used for histological grading (42).

As previously explained, there is no international consensus on the optimal Ki-67 cut-off. Ki-67 cut-off controversies are thoroughly explained in section 1.2.3.3.3.1. However, we chose the following cut-offs for Ki-67; </≥15% and </≥30%. Cut-off at 15% was chosen because it is similar to the cut-off previously shown to separate Luminal A from Luminal B tumours (62) and because it has been used for molecular subtyping on the same cohorts previously. In addition, we chose 30% as a second cut-off in accordance with recent recommendations from the International Ki-67 Working Group (63). Another option would have been to use mean values to define high and low status of Ki-67 and mitotic count.

Staining intensity is not considered in assessment of Ki-67 (63). Nevertheless, the different observers' perception of Ki-67 staining could still affect our results. For example, one observer may classify a cell as Ki-67 negative while another observer may classify the same cell as Ki-67 positive because they perceive the staining differently. To minimize the risk of misclassification, all sections were in general assessed by two different observers, and in case of disagreement consensus was reached.

1.18.5 Analytical considerations

The aims of the survival analyses in paper I and II were to study whether *MRPS23* or *ZNF703* provide additional prognostic information beyond what is already known through conventional prognostic factors. We adjusted for established prognostic factors such as age, stage, grade, Ki-67 and HER2 separately. Adjusting for these factors did not alter the results.

In paper III, the main aim was to study whether there had been a change in incidence rate of breast cancer according to the proliferation status. We used Ki-67 and mitotic count as markers of proliferation. Women from cohort 1, 3 and 4 were included in our analyses. The cohorts have been described in detail in section 1.6. Participants in the three cohorts were all included from what is now a united Trøndelag County, and across overlapping range of birth years. Thus, some participants may be included in more than one cohort. The overlap in incident breast cancers could be identified through case specific identity numbers. However, because of anonymization of the study participants it was not possible for us to identify overlap in the background population. Elimination of all participants overlapping based on birth year would eliminate a substantial number of study participants while elimination of only the identified overlapping breast cancer cases would lead to underestimation of incidence rates. We assumed that the overlap in the background population was proportionally similar to

the overlap of incident breast cancer cases. Thus, we chose to not exclude overlapping breast cancer cases to prevent underestimation of breast cancer incidence rates. Not eliminating duplicate incidents may however have led to some overestimation of precision.

In paper III we used multiple imputations to predict mitotic count and Ki-67 status for tumours with missing values of these variables. We imputed missing values for dichotomized values, rather than absolute values. We included all available information in the imputation model to minimize biased results. As expected, observed incidence rates were lower than imputed incidence rates, but the two followed the same pattern.

In all three papers we used 95% confidence intervals as a measure of precision. A 95% confidence interval can be interpreted as follows: If a study was repeated several times, the confidence interval would contain the true value 95% of the times (177). A wide confidence interval indicates low precision while a narrow confidence interval indicates high precision. The width of the confidence interval can be affected by the sample size and the variability in the samples.

Some of the subgroups had few cases and the results must be interpreted accordingly. In general, both for paper I and paper II, the number of amplified cases was small. When we subdivided further, such as the category combining *HER2* and *MRPS23* status or copy number status within the different molecular subtypes, numbers were even smaller. This is for example reflected in the wide confidence intervals for the cumulative incidence estimates and hazard ratios for MRPS23+/HER2+ tumours in paper I. Similarly for paper III, in the analyses of incidence rates of molecular subtypes according to proliferation status numbers were small. Thus, the results must be interpreted with caution.

In every step of a study there is a risk of random misclassification of information. In all studies presented in this thesis the observers were blinded to patient outcome during annotation to prevent it from affecting the results.

Conclusion and future perspectives

In paper I, we found that *MRPS23* amplification was associated with increased proliferation. Amplification was only found among luminal and HER2 type tumours. There was no association between *MRPS23* copy number and prognosis, nor between *MRPS23* gene expression and prognosis.

In paper II, we found that *ZNF703* copy number increase was associated with luminal B subtypes, high proliferation, and poor prognosis. We found a subgroup of amplified luminal A tumours with poor prognosis. There was no association between ZNF703 protein expression and prognosis.

In paper III, we examined the change in breast cancer incidence according to proliferation markers Ki-67 and mitotic count by comparing women born in 1929 or later to women born before 1929. We found that according to Ki-67, there has been an increase of both high and low-proliferative tumours, while according to mitotic count, there has only been an increase in low-proliferative tumours.

High proliferation is a hallmark of cancer, and it is associated with poor prognosis in breast cancer. Today, proliferation is routinely evaluated using Ki-67, mitotic count and indirectly through PAM50. Given the important role of proliferation in cancer development, controversies regarding Ki-67 assessment, and high costs of gene expression analyses, identification of new proliferation markers could be of great importance for breast cancer patients. We need new markers able to stratify patients into more precise prognostic and treatment groups, and identification of new proliferation markers could lead to novel targeted treatment in the future. In addition, our studies illustrate the heterogeneity within the existing molecular subtypes.

References

1. Ferlay J EM, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. Global Cancer Observatory: Cancer Today https://gco.iarc.fr/today/home: International Agency for Research on Cancer; 2020 [Available from: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf.
2. Cancer in Norway 2021 - Cancer incidence, mortality, survival and prevalence in Norway. 2022.
3. World Cancer Report: Cancer Research for Cancer Prevention. Wild CP WE, Stewart BW, editor. Lyon, France: International Agency for Research on Cancer; 2020.
4. Toriola AT, Colditz GAJBcr, treatment. Trends in breast cancer incidence and mortality in the United States: implications for prevention. 2013;138(3):665-73.
5. Cancer in Norway 2018 - Cancer incidence, mortality, survival and prevalence in Norway. . Oslo, Norway; 2020.
6. Autier P, Boniol M, Gavin A, Vatten LJ. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. BMJ (Clinical research ed). 2011;343:d4411.
7. Soerjomataram I, Lortet-Tieulent J, Ferlay J, Forman D, Mathers C, Parkin DM, et al. Estimating and validating disability-adjusted life years at the global level: a methodological framework for cancer. BMC Medical Research Methodology. 2012;12(1):125.
8. Norway CRo. Cancer in Norway 2020 - Cancer incidence, mortality, survival and prevalence in Norway. Oslo, Norway; 2020.
9. Hofvind S, Sakshaug S, Ursin G, Graff-Iversen S. Breast cancer incidence trends in Norway--explained by hormone therapy or mammographic screening? Int J Cancer. 2012;130(12):2930-8.
10. Weedon-Fekjær H, Bakken K, Vatten LJ, Tretli S. Understanding recent trends in incidence of invasive breast cancer in Norway: age-period-cohort analysis based on registry data on mammography screening and hormone treatment use. BMJ (Clinical research ed). 2012;344:e299.
11. Glass AG, Lacey JV, Jr., Carreon JD, Hoover RN. Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. J Natl Cancer Inst. 2007;99(15):1152-61.
12. Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, et al. The decrease in breast-cancer incidence in 2003 in the United States. New England Journal of Medicine. 2007;356(16):1670-4.
13. Årsrapport 2021 med resultater og forbedringstiltak fra Nasjonalt kvalitetsregister for brystkreft. . Oslo: Kreftregisteret; 2022.
14. Aromaa-Häyhä A, Auvinen P, Sarkeala T, Heinävaara S, Lamminmäki A, Malila N, et al. Improved diagnostics and change of tumour characteristics in breast cancer: a retrospective study over two decades. Acta Oncol. 2018;57(10):1331-8.
15. Anderson WF, Katki HA, Rosenberg PS. Incidence of breast cancer in the United States: current and future trends. Journal of the National Cancer Institute. 2011;103(18):1397-402.
16. Mullooly M, Murphy J, Gierach GL, Walsh PM, Deady S, Barron TI, et al. Divergent oestrogen receptor-specific breast cancer trends in Ireland (2004–2013): Amassing data from independent Western populations provide etiologic clues. European Journal of Cancer. 2017;86:326-33.
17. Mesa-Eguiagaray I, Wild SH, Rosenberg PS, Bird SM, Brewster DH, Hall PS, et al. Distinct temporal trends in breast cancer incidence from 1997 to 2016 by molecular subtypes: a population-based study of Scottish cancer registry data. British Journal of Cancer. 2020;123(5):852-9.
18. Hofvind S, Holen Å, Román M, Sebuødegård S, Puig-Vives M, Akslen L. Mode of detection: an independent prognostic factor for women with breast cancer. J Med Screen. 2016;23(2):89-97.
19. Crispo A, Barba M, D'Aiuto G, De Laurentiis M, Grimaldi M, Rinaldo M, et al. Molecular profiles of screen detected vs. symptomatic breast cancer and their impact on survival: results from a clinical series. BMC Cancer. 2013;13:15.
20. Sihto H, Lundin J, Lehtimäki T, Sarlomo-Rikala M, Bützow R, Holli K, et al. Molecular subtypes of breast cancers detected in mammography screening and outside of screening. Clin Cancer Res. 2008;14(13):4103-10.
21. Farshid G, Walters D. Molecular subtypes of screen-detected breast cancer. Breast cancer research and treatment. 2018;172(1):191-9.
22. 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.
23. Sun Y-S, Zhao Z, Yang Z-N, Xu F, Lu H-J, Zhu Z-Y, et al. Risk factors and preventions of breast cancer. International journal of biological sciences. 2017;13(11):1387.
24. Horn J, Vatten LJ. Reproductive and hormonal risk factors of breast cancer: a historical perspective. Int J Womens Health. 2017;9:265-72.

25. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. Jama. 2002;288(3):321-33.
26. Danaei G, Vander Hoorn S, Lopez AD, Murray CJL, Ezzati M. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. The Lancet. 2005;366(9499):1784-93.
27. Tamimi RM, Colditz GA, Hazra A, Baer HJ, Hankinson SE, Rosner B, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. Breast cancer research and treatment. 2012;131(1):159-67.
28. Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. Annals of Oncology. 2015;26(7):1291-9.
29. Anderson KN, Schwab RB, Martinez ME. Reproductive risk factors and breast cancer subtypes: a review of the literature. Breast cancer research and treatment. 2014;144(1):1-10.
30. Barnard ME, Boeke CE, Tamimi RM. Established breast cancer risk factors and risk of intrinsic tumor subtypes. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2015;1856(1):73-85.
31. Horn J, Opdahl S, Engstrøm MJ, Romundstad PR, Tretli S, Haugen OA, et al. Reproductive history and the risk of molecular breast cancer subtypes in a prospective study of Norwegian women. Cancer Causes Control. 2014;25(7):881-9.
32. Pierobon M, Frankenfeld CL. Obesity as a risk factor for triple-negative breast cancers: a systematic review and meta-analysis. Breast cancer research and treatment. 2013;137(1):307-14.
33. Garcia-Estevez L, Moreno-Bueno G. Updating the role of obesity and cholesterol in breast cancer. Breast Cancer Research. 2019;21(1):35.
34. Alsaker MD, Janszky I, Opdahl S, Vatten LJ, Romundstad PR. Weight change in adulthood and risk of postmenopausal breast cancer: the HUNT study of Norway. Br J Cancer. 2013;109(5):1310-7.
35. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft. helsebiblioteket.no: Helsedirektoratet; 2022.
36. Research-based evaluation of the Norwegian Cancer Screening Program Regjeringen.no: The Research Council of Norway; 2015 [Available from: https://www.regjeringen.no/contentassets/444d08daf15e48aca5321f2cefaac511/mammografirapport-til-web.pdf.
37. Hofvind S, Tsuruda K, Mangerud G, Ertzaas A, Holen Å, Pedersen K, et al. The Norwegian Breast Cancer Screening Program, 1996-2016: celebrating 20 years of organised mammographic screening. Cancer in Norway. 2016:S1-S72.
38. Sebuødegård S, Botteri E, Hofvind S. Breast Cancer Mortality After Implementation of Organized Population-Based Breast Cancer Screening in Norway. JNCI: Journal of the National Cancer Institute. 2020;112(8):839-46.
39. Fakta om overdiagnostikk kreftregisteret.no [Available from: https://www.kreftregisteret.no/screening/Mammografiprogrammet/Evaluering/Fakta-om-overdiagnostikk/.
40. Mucci LA, Wilson KM, Rider JR. Cancer Screening. In: Loda M, Mucci LA, Mittelstadt ML, Van Hemelrijck M, Cotter MB, editors. Pathology and Epidemiology of Cancer. Cham: Springer International Publishing; 2017. p. 73-82.
41. Brierley JD, Union for International Cancer C. TNM classification of malignant tumours. Chichester, England: Wiley Blackwell; 2017.
42. 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.
43. Amin MB, American Joint Committee on C. AJCC cancer staging manual. 8th ed. ed. New York, N.Y: Springer; 2017.
44. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, et al. Breast cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA: a cancer journal for clinicians. 2017;67(4):290-303.
45. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology. 1991;19(5):403-10.
46. Cree IA, Tan PH, Travis WD, Wesseling P, Yagi Y, White VA, et al. Counting mitoses: SI(ze) matters! Mod Pathol. 2021;34(9):1651-7.
47. Bloom H, Richardson W. Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. British journal of cancer. 1957;11(3):359.
48. Italiano A. Prognostic or predictive? It's time to get back to definitions! Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(35):4718; author reply -9.
49. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, et al. Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the

Primary Therapy of Early Breast Cancer 2015. Annals of oncology : official journal of the European Society for Medical Oncology. 2015;26(8):1533-46.
50. Petersen OW, Høyer PE, van Deurs B. Frequency and distribution of estrogen receptor-positive cells in normal, nonlactating human breast tissue. Cancer research. 1987;47(21):5748-51.
51. Waks AG, Winer EP. Breast Cancer Treatment: A Review. Jama. 2019;321(3):288-300.
52. Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, et al. Estrogen and Progesterone Receptor Testing in Breast Cancer: ASCO/CAP Guideline Update. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2020;38(12):1346-66.
53. GeneCards. ERBB2 gene genecards.org [Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=ERBB2&keywords=HER2.
54. Rubin I, Yarden YJAoo. The basic biology of HER2. 2001;12:S3-S8.
55. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science (New York, NY). 1989;244(4905):707-12.
56. Rakha EA, Pinder SE, Bartlett JM, Ibrahim M, Starczynski J, Carder PJ, et al. Updated UK Recommendations for HER2 assessment in breast cancer. Journal of clinical pathology. 2015;68(2):93-9.
57. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2018;36(20):2105-22.
58. Human Epidermal Growth Factor Receptor 2 (HER2) Testing in Breast Cancer Guidelines: 2018 Focused Update. Summary of Changes and Definitions. cap.org: College of American Pathologists; 2018 [Available from: https://www.cap.org/protocols-and-guidelines/cap-guidelines/current-cap-guidelines/recommendations-for-human-epidermal-growth-factor-2-testing-in-breast-cancer.
59. Gerdes J, Lemke H, Baisch H, Wacker HH, Schwab U, Stein H. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. Journal of immunology (Baltimore, Md : 1950). 1984;133(4):1710-5.
60. de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. British Journal of Cancer. 2007;96(10):1504-13.
61. Stuart-Harris R, Caldas C, Pinder SE, Pharoah P. Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. Breast (Edinburgh, Scotland). 2008;17(4):323-34.
62. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009;101(10):736-50.
63. Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, et al. Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group. JNCI: Journal of the National Cancer Institute. 2021;113(7):808-19.
64. Burstein HJ, Curigliano G, Thürlimann B, Weber WP, Poortmans P, Regan MM, et al. Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021. Annals of Oncology. 2021;32(10):1216-35.
65. Focke CM, Bürger H, van Diest PJ, Finsterbusch K, Gläser D, Korsching E, et al. Interlaboratory variability of Ki67 staining in breast cancer. European journal of cancer (Oxford, England : 1990). 2017;84:219-27.
66. Focke CM, van Diest PJ, Decker T. St Gallen 2015 subtyping of luminal breast cancers: impact of different Ki67-based proliferation assessment methods. Breast cancer research and treatment. 2016;159(2):257-63.
67. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Annals of oncology : official journal of the European Society for Medical Oncology. 2011;22(8):1736-47.
68. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in Breast Cancer: Recommendations from the International Ki67 in Breast Cancer Working Group. JNCI: Journal of the National Cancer Institute. 2011;103(22):1656-64.
69. FDA approves abemaciclib with endocrine therapy for early breast cancer fda.gov: U.S Food & Drug Administration; 2021 [Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-abemaciclib-endocrine-therapy-early-breast-cancer.
70. Harbeck N, Rastogi P, Martin M, Tolaney SM, Shao ZM, Fasching PA, et al. Adjuvant abemaciclib combined with endocrine therapy for high-risk early breast cancer: updated efficacy and Ki-67 analysis from the monarchE study. Annals of oncology : official journal of the European Society for Medical Oncology. 2021;32(12):1571-81.

71. Johnston SRD, Harbeck N, Hegg R, Toi M, Martin M, Shao ZM, et al. Abemaciclib Combined With Endocrine Therapy for the Adjuvant Treatment of HR+, HER2-, Node-Positive, High-Risk, Early Breast Cancer (monarchE). Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2020;38(34):3987-98.
72. Knutsvik G, Stefansson IM, Aziz S, Arnes J, Eide J, Collett K, et al. Evaluation of Ki67 expression across distinct categories of breast cancer specimens: a population-based study of matched surgical specimens, core needle biopsies and tissue microarrays. PloS one. 2014;9(11):e112121.
73. CD274 Genecards.org [Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CD274&keywords=PDL1.
74. Kythreotou A, Siddique A, Mauri FA, Bower M, Pinato DJ. PD-L1. Journal of clinical pathology. 2018;71(3):189-94.
75. Bertucci F, Gonçalves A. Immunotherapy in Breast Cancer: the Emerging Role of PD-1 and PD-L1. Current oncology reports. 2017;19(10):64.
76. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer. 2007;109(9):1721-8.
77. Zubair M, Wang S, Ali N. Advanced Approaches to Breast Cancer Classification and Diagnosis. Frontiers in Pharmacology. 2021;11.
78. Kwa M, Makris A, Esteva FJ. Clinical utility of gene-expression signatures in early stage breast cancer. Nature Reviews Clinical Oncology. 2017;14(10):595-610.
79. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(32):4273-8.
80. Sestak I, Buus R, Cuzick J, Dubsky P, Kronenwett R, Denkert C, et al. Comparison of the Performance of 6 Prognostic Signatures for Estrogen Receptor–Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial. JAMA Oncology. 2018;4(4):545-53.
81. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.
82. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(19):10869-74.
83. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(8):1160-7.
84. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. 2007;8(5):R76.
85. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Research. 2010;12(5):R68.
86. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the national academy of sciences. 2003;100(14):8418-23.
87. Prat A, Pineda E, Adamo B, Galván P, Fernández A, Gaba L, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast (Edinburgh, Scotland). 2015;24 Suppl 2:S26-35.
88. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006;7(1):96.
89. Russnes HG, Lingjærde OC, Børresen-Dale AL, Caldas C. Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters. Am J Pathol. 2017;187(10):2152-62.
90. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346-52.
91. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. Cell. 2014;158(4):929-44.
92. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47-54.
93. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res. 2004;10(16):5367-74.

94. Dawood S, Hu R, Homes MD, Collins LC, Schnitt SJ, Connolly J, et al. Defining breast cancer prognosis based on molecular phenotypes: results from a large cohort study. Breast cancer research and treatment. 2011;126(1):185-92.
95. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama. 2006;295(21):2492-502.
96. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res. 2008;14(5):1368-76.
97. Engstrøm MJ, Opdahl S, Hagen AI, Romundstad PR, Akslen LA, Haugen OA, et al. Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. Breast cancer research and treatment. 2013;140(3):463-73.
98. Sakach E, O'Regan R, Meisel J, Li X. Molecular Classification of Triple Negative Breast Cancer and the Emergence of Targeted Therapies. Clinical Breast Cancer. 2021;21(6):509-20.
99. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Annals of oncology : official journal of the European Society for Medical Oncology. 2013;24(9):2206-23.
100. Lehmann BD, Jovanović B, Chen X, Estrada MV, Johnson KN, Shyr Y, et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. PloS one. 2016;11(6):e0157368.
101. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. The Journal of clinical investigation. 2011;121(7):2750-67.
102. Popli P, Gutterman EM, Omene C, Ganesan S, Mills D, Marlink R. Receptor-Defined Breast Cancer in Five East African Countries and Its Implications for Treatment: Systematic Review and Meta-Analysis. JCO Glob Oncol. 2021;7:289-301.
103. Yoo T-K, Kang J, Lee A, Chae BJ. A triple-negative breast cancer surrogate subtype classification that correlates with gene expression subtypes. Breast cancer research and treatment. 2022;191(3):599-610.
104. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor–positive breast cancer. Clinical cancer research. 2010;16(21):5222-32.
105. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol. 2011;5(1):5-23.
106. Blows FM, Driver KE, Schmidt MK, Broeks A, Van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS medicine. 2010;7(5):e1000279.
107. Valla M, Vatten LJ, Engstrom MJ, Haugen OA, Akslen LA, Bjorngaard JH, et al. Molecular Subtypes of Breast Cancer: Long-term Incidence Trends and Prognostic Differences. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2016;25(12):1625-34.
108. Hanahan D, Weinberg RAJc. The hallmarks of cancer. 2000;100(1):57-70.
109. Hanahan D, Weinberg RAJc. Hallmarks of cancer: the next generation. 2011;144(5):646-74.
110. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discovery. 2022;12(1):31-46.
111. van Diest PJ, van der Wall E, Baak JP. Prognostic value of proliferation in invasive breast cancer: a review. Journal of clinical pathology. 2004;57(7):675-81.
112. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res. 2008;10(4):R65.
113. Gatza ML, Silva GO, Parker JS, Fan C, Perou CMJNg. An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. 2014;46(10):1051.
114. Pathmanathan N, Balleine RL. Ki67 and proliferation in breast cancer. Journal of clinical pathology. 2013;66(6):512-6.
115. Juríková M, Danihel Ľ, Polák Š, Varga I. Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer. Acta Histochem. 2016;118(5):544-52.
116. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft, 2019. In: Helsedirektoratet, editor.: Helsedirektoratet; 2019.
117. Beresford MJ, Wilson GD, Makris A. Measuring proliferation in breast cancer: practicalities and applications. Breast Cancer Research. 2006;8(6):216.
118. Penault-Llorca F, Radosevic-Robin N. Proliferation Markers in Breast Cancer. In: Badve S, Gökmen-Polar Y, editors. Molecular Pathology of Breast Cancer. Cham: Springer International Publishing; 2016. p. 81-98.

119. Albertson DG. Gene amplification in cancer. Trends in genetics : TIG. 2006;22(8):447-55.
120. Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, et al. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer research. 2002;62(21):6240-5.
121. Ooi A, Inokuchi M, Horike SI, Kawashima H, Ishikawa S, Ikeda H, et al. Amplicons in breast cancers analyzed by multiplex ligation-dependent probe amplification and fluorescence in situ hybridization. Human pathology. 2019;85:33-43.
122. Letessier A, Sircoulomb F, Ginestier C, Cervera N, Monville F, Gelsi-Boyer V, et al. Frequency, prognostic impact, and subtype association of 8p12, 8q24, 11q13, 12p13, 17q12, and 20q13 amplifications in breast cancers. BMC Cancer. 2006;6(1):245.
123. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer cell. 2006;10(6):529-41.
124. Conticello SG. The AID/APOBEC family of nucleic acid mutators. Genome Biology. 2008;9(6):229.
125. Pu M, Wang J, Huang Q, Zhao G, Xia C, Shang R, et al. High MRPS23 expression contributes to hepatocellular carcinoma proliferation and indicates poor survival outcomes. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2017;39(7):1010428317709127.
126. Gao Y, Li F, Zhou H, Yang Y, Wu R, Chen Y, et al. Down-regulation of MRPS23 inhibits rat breast cancer proliferation and metastasis. Oncotarget. 2017;8(42):71772-81.
127. Lyng H, Brøvig RS, Svendsrud DH, Holm R, Kaalhus O, Knutstad K, et al. Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. BMC genomics. 2006;7(1):268.
128. Li B, Zhang YL. Identification of up-regulated genes in human uterine leiomyoma by suppression subtractive hybridization. Cell Research. 2002;12:215.
129. Staub E, Gröne J, Mennerich D, Röpcke S, Klamann I, Hinzmann B, et al. A genome-wide map of aberrantly expressed chromosomal islands in colorectal cancer. Molecular cancer. 2006;5:37.
130. Kurokawa Y, Matoba R, Nakamori S, Takemasa I, Nagano H, Dono K, et al. PCR-array gene expression profiling of hepatocellular carcinoma. JOURNAL OF EXPERIMENTAL AND CLINICAL CANCER RESEARCH. 2004;23(1):135-42.
131. Oviya RP, Gopal G, Shirley SS, Sridevi V, Jayavelu S, Rajkumar T. Mitochondrial ribosomal small subunit proteins (MRPS) MRPS6 and MRPS23 show dysregulation in breast cancer affecting tumorigenic cellular processes. Gene. 2021;790:145697.
132. Liu L, Zhang X, Ding H, Liu X, Cao D, Liu Y, et al. Arginine and lysine methylation of MRPS23 promotes breast cancer metastasis through regulating OXPHOS. Oncogene. 2021;40(20):3548-63.
133. Reynisdottir I, Arason A, Einarsdottir BO, Gunnarsson H, Staaf J, Vallon-Christersson J, et al. High expression of ZNF703 independent of amplification indicates worse prognosis in patients with luminal B breast cancer. Cancer medicine. 2013;2(4):437-46.
134. Gelsi-Boyer V, Orsetti B, Cervera N, Finetti P, Sircoulomb F, Rougé C, et al. Comprehensive profiling of 8p11-12 amplification in breast cancer. 2005;3(12):655-67.
135. Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, Chin SF, et al. ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. 2011;3(3):167-80.
136. Information) NNCfB. ZNF703 zinc finger protein 703 [Homo sapiens (human)] https://www.ncbi.nlm.nih.gov/2020 [updated 13.03.2020. Available from: https://www.ncbi.nlm.nih.gov/gene/80139?report=full_report.
137. Fumagalli D, Wilson T, Salgado R, Lu X, Yu J, O'Brien C, et al. Somatic mutation, copy number and transcriptomic profiles of primary and matched metastatic estrogen receptor-positive breast cancers. 2016;27(10):1860-6.
138. Liu XY, Ma D, Xu XE, Jin X, Yu KD, Jiang YZ, et al. Genomic Landscape and Endocrine-Resistant Subgroup in Estrogen Receptor-Positive, Progesterone Receptor-Negative, and HER2-Negative Breast Cancer. Theranostics. 2018;8(22):6386-99.
139. Bofin AM, Ytterhus B, Klæstad E, Valla M. FGFR1 copy number in breast cancer: associations with proliferation, histopathological grade and molecular subtypes. Journal of clinical pathology. 2021.
140. Valla M, Klæstad E, Ytterhus B, Bofin AM. CCND1 Amplification in Breast Cancer -associations With Proliferation, Histopathological Grade, Molecular Subtype and Prognosis. J Mammary Gland Biol Neoplasia. 2022;27(1):67-77.
141. Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousselet E, et al. ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. 2011;3(3):153-66.
142. Glodzik D, Purdie C, Rye IH, Simpson PT, Staaf J, Span PN, et al. Mutational mechanisms of amplifications revealed by analysis of clustered rearrangements in breast cancers. Annals of oncology : official journal of the European Society for Medical Oncology. 2018;29(11):2223-31.
143. Zhang X, Mu X, Huang O, Wang Z, Chen J, Chen D, et al. ZNF703 promotes triple-negative breast cancer cells through cell-cycle signaling and associated with poor prognosis. BMC Cancer. 2022;22(1):226.

144. Slorach EM, Chou J, Werb Z. Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. Genes & development. 2011;25(5):471-84.
145. Baltaci E, Karaman E, Dalay N, Buyru N. Analysis of gene copy number changes in head and neck cancer. Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery. 2018;43(4):1004-9.
146. Yang H, Jiang WQ, Cao Y, Sun YA, Wei J, An X, et al. Elevated ZNF703 Protein Expression Is an Independent Unfavorable Prognostic Factor for Survival of the Patients with Head and Neck Squamous Cell Carcinoma. Disease markers. 2015;2015:640263.
147. Wang H, Deng X, Zhang J, Ou Z, Mai J, Ding S, et al. Elevated Expression of Zinc Finger Protein 703 Promotes Cell Proliferation and Metastasis through PI3K/AKT/GSK-3beta Signalling in Oral Squamous Cell Carcinoma. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology. 2017;44(3):920-34.
148. Baykara O, Dalay N, Kaynak K, Buyru NJCm. ZNF 703 Overexpression may act as an oncogene in non-small cell lung cancer. 2016;5(10):2873-8.
149. Baykara O, Bakir B, Buyru N, Kaynak K, Dalay N. Amplification of chromosome 8 genes in lung cancer. Journal of Cancer. 2015;6(3):270-5.
150. Yang G, Ma F, Zhong M, Fang L, Peng Y, Xin X, et al. ZNF703 acts as an oncogene that promotes progression in gastric cancer. Oncology reports. 2014;31(4):1877-82.
151. Li K, Wang J, Han J, Lan Y, Xie C, Pan S, et al. Overexpression of ZNF703 facilitates tumorigenesis and predicts unfavorable prognosis in patients with cholangiocarcinoma. Oncotarget. 2016;7(46):76108-17.
152. Zhang X, Mu X, Huang O, Xie Z, Jiang M, Geng M, et al. Luminal breast cancer cell lines overexpressing ZNF703 are resistant to tamoxifen through activation of Akt/mTOR signaling. 2013;8(8).
153. Dawson SJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. Embo j. 2013;32(5):617-28.
154. Holmen m.fl J. The Nord-Trøndelag Health Study 1995-97 (HUNT 2). Norsk Epidemiologi. 2011;13(1).
155. Sandvei MS, Lagiou P, Romundstad PR, Trichopoulos D, Vatten LJ. Size at birth and risk of breast cancer: update from a prospective population-based study. Eur J Epidemiol. 2015;30(6):485-92.
156. Vatten LJ, Nilsen TI, Tretli S, Trichopoulos D, Romundstad PR. Size at birth and risk of breast cancer: prospective population-based study. Int J Cancer. 2005;114(3):461-4.
157. Sandvei MS, Opdahl S, Valla M, Lagiou P, Vesterfjell EV, Rise TV, et al. The association of women's birth size with risk of molecular breast cancer subtypes: a cohort study. BMC Cancer. 2021;21(1):299.
158. This is the National Registry skatteetaten.no: The Norwegian Tax Administration; [Available from: https://www.skatteetaten.no/en/person/national-registry/about/this-is-the-national-registry/.
159. Om kreftregisteret Kreftregisteret.no: Kreftregisteret; 2020 [updated 20.03.2020. Available from: https://www.kreftregisteret.no/Generelt/Om-Kreftregisteret/.
160. 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. European journal of cancer (Oxford, England : 1990). 2009;45(7):1218-31.
161. Om dødsårsakregisteret FHI.no: Norwegian Insitute of Public Health; [Available from: https://www.fhi.no/hn/helseregistre-og-registre/dodsarsaksregisteret/dodsarsaksregisteret2/.
162. Slik kodes og kvalitetssikres dødsårsaker i Dødsårsaksregisteret FHI.no: Norwegian Institute of Public Health; [updated 08.06.2021. Available from: https://www.fhi.no/hn/helseregistre-og-registre/dodsarsaksregisteret/dodsarsaken-kodes-med-icd-koder/.
163. Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects etikkom.no: The Norwegian National Committees for Research Ethics; 2014 [Available from: https://www.etikkom.no/en/ethical-guidelines-for-research/medical-and-health-research/declaration-of-helsinki/.
164. Sønderland K. The Health Research Act etikkom.no: The Norwegian National Research Ethics Committees; 2016 [Available from: https://www.etikkom.no/en/library/practical-information/legal-statutes-and-guidelines/the-health-research-act/.
165. Hayes DF, Ethier S, Lippman ME. New guidelines for reporting of tumor marker studies in breast cancer research and treatment: REMARK. Breast cancer research and treatment. 2006;100(2):237-8.
166. Kononen J, Bubendorf L, Kallioniemi A, Bärlund M, Schraml P, Leighton S, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med. 1998;4(7):844-7.
167. Sapino A, Marchiò C, Senetta R, Castellano I, Macrì L, Cassoni P, et al. Routine assessment of prognostic factors in breast cancer using a multicore tissue microarray procedure. Virchows Archiv. 2006;449(3):288-96.
168. Zhang D, Salto-Tellez M, Putti TC, Do E, Koay ES-C. Reliability of Tissue Microarrays in Detecting Protein Expression and Gene Amplification in Breast Cancer. Modern Pathology. 2003;16(1):79-85.

169. Pinder SE, Brown JP, Gillett C, Purdie CA, Speirs V, Thompson AM, et al. The manufacture and assessment of tissue microarrays: suggestions and criteria for analysis, with breast cancer as an example. Journal of clinical pathology. 2013;66(3):169-77.
170. Buchwalow IB, Böcker W. Immunohistochemistry. Basics and Methods. 2010;1:1-149.
171. Dabbs DJ. Diagnostic Immunohistochemistry E-Book: Theranostic and Genomic Applications: Elsevier Health Sciences; 2017.
172. Schwarzacher T, Heslop-Harrison P. Practical in situ hybridization. Oxford: BIOS; 2000.
173. Wang B, Ding W, Sun K, Wang X, Xu L, Teng X. Impact of the 2018 ASCO/CAP guidelines on HER2 fluorescence in situ hybridization interpretation in invasive breast cancers with immunohistochemically equivocal results. Scientific Reports. 2019;9(1):16726.
174. The Cancer Genome Atlas Program cancer.gov: National Cancer Institute; [Available from: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.
175. Kleinbaum DG, Klein M. Survival Analysis : A Self-Learning Text, Third Edition. New York, NY: Springer New York : Imprint: Springer; 2012.
176. Veierød MB, Lydersen S, Laake P. Medical statistics : in clinical and epidemiological research. Oslo: Gyldendal akademisk; 2012.
177. Aalen OO, Frigessi A. Statistiske metoder i medisin og helsefag. 2. utg. ed. Oslo: Gyldendal akademisk; 2018.
178. Rosner B. Fundamentals of biostatistics. 8th ed. ed. Boston, Mass: Cengage Learning; 2016.
179. Carneiro I, Howard N. Introduction To Epidemiology. Maidenhead, UNITED KINGDOM: McGraw-Hill Education; 2011.
180. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol. 2012;41(3):861-70.
181. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ (Clinical research ed). 2009;338:b2393.
182. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59(10):1087-91.
183. Geiger T, Cox J, Mann M. Proteomic changes resulting from gene copy number variations in cancer cells. PLoS genetics. 2010;6(9):e1001090.
184. Anderson WF, Rosenberg PS, Petito L, Katki HA, Ejlertsen B, Ewertz M, et al. Divergent estrogen receptor-positive and-negative breast cancer trends and etiologic heterogeneity in Denmark. International journal of cancer. 2013;133(9):2201-6.
185. Tømmerås AM. Nok en gang rekordlav fruktbarhet ssb.no: Statistics Norway; 2021 [Available from: https://www.ssb.no/befolkning/fodte-og-dode/statistikk/fodte/artikler/nok-en-gang-rekordlav-fruktbarhet.
186. Meyer HE, et al. Overvekt og fedme i Norge fhi.no: Norwegian Insitute of Public Health; [updated 01.07.2022. Available from: https://www.fhi.no/nettpub/hin/ikke-smittsomme/overvekt-og-fedme/?term=&h=1.
187. Rossi L, Laas E, Mallon P, Vincent-Salomon A, Guinebretiere JM, Lerebours F, et al. Prognostic impact of discrepant Ki67 and mitotic index on hormone receptor-positive, HER2-negative breast carcinoma. Br J Cancer. 2015;113(7):996-1002.
188. Clare SE, Shaw PL. “Big Data” for breast cancer: where to look and what you will find. npj Breast Cancer. 2016;2(1):16031.
189. Engel KB, Moore HM. Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue. Arch Pathol Lab Med. 2011;135(5):537-43.
190. Combs SE, Han G, Mani N, Beruti S, Nerenberg M, Rimm DL. Loss of antigenicity with tissue age in breast cancer. Laboratory Investigation. 2016;96(3):264-9.
191. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. Arch Pathol Lab Med. 2018.
192. Hanna WM, Rüschoff J, Bilous M, Coudry RA, Dowsett M, Osamura RY, et al. HER2 in situ hybridization in breast cancer: clinical implications of polysomy 17 and genetic heterogeneity. Mod Pathol. 2014;27(1):4-18.

Errata

Paper I: It is written in the article, in the section on “Statistical analyses”, that cox analyses were adjusted for molecular subtypes where applicable. However, this was no included in the analyses and should not have been written in the text.

Paper II: It is written in the article, in the section on “Statistical analyses”, that cox analyses were adjusted for molecular subtypes when applicable. However, this was no included in the analyses and should not have been written in the text.

Paper I

Breast Cancer Research and Treatment (2020) 180:73–86
https://doi.org/10.1007/s10549-020-05532-6

PRECLINICAL STUDY

MRPS23 amplification and gene expression in breast cancer; association with proliferation and the non-basal subtypes

Elise Klæstad[1] · Signe Opdahl[2] · Monica Jernberg Engstrøm[1,3] · Borgny Ytterhus[1] · Elisabeth Wik[4,5] · Anna Mary Bofin[1] · Marit Valla[1,6]

Received: 27 July 2019 / Accepted: 9 January 2020 / Published online: 16 January 2020

Abstract

Purpose *MRPS23* is recognized as a driver of proliferation in luminal breast cancer. The aims of the present study were to describe *MRPS23* copy number change in breast cancer, and to assess associations between *MRPS23* copy number change and molecular subtype, proliferation and prognosis, and between *MRPS23* gene expression and molecular subtype and prognosis.

Methods Using fluorescence in situ hybridization (FISH), we examined *MRPS23* and centromere 17 copy number in 590 formalin-fixed, paraffin-embedded primary tumours and 144 corresponding lymph node metastases from a cohort of Norwegian breast cancer patients. Furthermore, we analysed *MRPS23* gene expression data in 1971 primary breast cancer tumours from the METABRIC dataset. We used Pearson's χ^2 test to assess associations between *MRPS23* copy number and molecular subtype and proliferation, and between *MRPS23* expression and molecular subtype. We studied prognosis by estimating hazard ratios and cumulative incidence of death from breast cancer according to *MRPS23* copy number and *MRPS23* expression status.

Results We found *MRPS23* amplification (mean *MRPS23* copy number ≥ 6 and/or *MRPS23*/chromosome 17 ratio ≥ 2) in 8% of primary tumours. Copy number increase associated with non-basal subtypes and higher tumour cell proliferation (Ki67). Higher *MRPS23* expression associated with the Luminal B subtype. We found no significant association between *MRPS23* amplification or *MRSP23* gene expression, and prognosis.

Conclusion Amplification of *MRPS23* is associated with higher proliferation and non-basal subtypes in breast cancer. High *MRPS23* expression is associated with the Luminal B subtype.

Keywords *MRPS23* · Breast cancer · Proliferation · Copy number · Amplification · METABRIC

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10549-020-05532-6) contains supplementary material, which is available to authorized users.

✉ Elise Klæstad
elisekl@stud.ntnu.no

[1] Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Erling Skjalgssons gate, 7030 Trondheim, Norway

[2] Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway

[3] Department of Breast and Endocrine Surgery, St. Olav's Hospital, Trondheim University Hospital, 7006 Trondheim, Norway

[4] Department of Clinical Medicine, Section for Pathology, Centre for Cancer Biomarkers CCBIO, University of Bergen, 5021 Bergen, Norway

[5] Department of Pathology, Haukeland University Hospital, 5021 Bergen, Norway

[6] Department of Pathology, St. Olav's Hospital, Trondheim University Hospital, 7006 Trondheim, Norway

Introduction

Increased proliferation is a hallmark of cancer [1, 2], and identification of genetic drivers of proliferation could be important for prognostication and development of new targeted treatment. By high-throughput genomic analyses, Gatza et al. identified proliferation driving genes in non-basal breast cancer [3]. Amplification of four of these genes (*MRPS23*, *FGD5*, *DTX3* and *METTL6*) was associated with a poor prognosis. Mitochondrial ribosomal protein S23 (*MRPS23*) is located on the long arm of chromosome 17 (17q22) and belongs to the mitochondrial ribosomal protein gene family [4, 5]. Mitochondrial ribosomes are composed of a small 28S subunit and a large 39S subunit. *MRPS23* encodes the 28S subunit [4, 5]. High *MRPS23* expression has been found in colon [6], cervical [7, 8] and hepatocellular cancer [9, 10], and associated with poor prognosis in non-basal breast cancer [3], hepatocellular [10] and cervical cancer [7, 8]. In a breast cancer mouse model, *MRPS23* knock-down reduced proliferation, induced apoptosis and limited angiogenesis and lymph node metastasis [11].

Our group has previously reclassified breast cancer tumours from a large cohort of Norwegian women into six molecular subtypes based on immunohistochemistry (IHC) and chromogenic in situ hybridization (CISH) [12]. The aims of the present study were to characterize *MRPS23* copy number alterations by fluorescence in situ hybridization (FISH) on formalin-fixed, paraffin-embedded (FFPE) primary tumour tissue and corresponding lymph node metastases from this cohort, and to assess how these copy number alterations associates with molecular subtypes, proliferation and prognosis. Furthermore, using the METABRIC dataset [13], we assess how *MRPS23* gene expression levels correlate with molecular subtypes and prognosis.

Materials and methods

Study populations and specimen characteristics

Cohort 1

Between 1956 and 1959, 25,727 women born 1886–1928 were invited to attend a clinical examination for early detection of breast cancer in Nord-Trøndelag County, Norway [14]. Through linkage with data from the Cancer Registry of Norway, these women were followed for breast cancer occurrence. Between 1961 and 2008, 1393 new breast cancers were registered. All tumours were reclassified according to histological type and grade [12, 15]. Tissue microarray (TMA) blocks were made using the Tissue Arrayer Mini-Core with TMA Designer2 software (Alphelys). Three 1-mm-in-diameter tissue cores from the periphery of the FFPE primary tumours and lymph node metastases were transferred to TMA recipient blocks. TMA sections (4 µm) were cut and stained, and the tumours were reclassified into molecular subtypes. Human epidermal growth factor receptor 2 (HER2) was assessed using both CISH and IHC. Tumours with *HER2*/chromosome enumeration probe 17 (CEP17) ratio ≥ 2 were defined as HER2 positive. When CISH was unsuccessful, tumours with intense membranous staining (IHC 3+) in $> 10\%$ of tumour cells were considered HER2 positive.

Of the 1393 tumours, 909 were successfully reclassified into molecular subtypes [12]:Luminal A (ER and/or PR+, HER2−, Ki67 < 15%), Luminal B (HER2−) (ER+ and/or PR+, HER2−, Ki67 ≥ 15%), Luminal B (HER2+) (ER+ and/or PR+, HER2+), HER2 type (ER− and PR−, HER2+), 5 negative phenotype (5NP; ER−, PR−, HER2−, CK5− and EGFR−) and Basal phenotype (BP; ER−, PR−, HER2−, CK5+ and/or EGFR+) (Supplementary Fig. 1). From the time of diagnosis (baseline), patients were followed until death from breast cancer, death from other causes or until December 31st, 2015. Individual information about adjuvant treatment is unavailable. However, due to age and/or time of diagnosis, few would have received chemotherapy. Some would have been treated with antihormonal treatment, but none qualified for trastuzumab. In the present study, TMAs containing cores from tumours diagnosed mainly in the 1980s or later ($n = 636$) were included. Of these, 46 were excluded due to unsuccessful FISH ($n = 30$) or insufficient amounts of tumour tissue ($n = 16$). Thus, 590 cases were suitable for *MRPS23* and CEP17 copy number assessment. Of these, 192 had lymph node metastases, and lymph node tissue from 150 was available in TMAs. Due to unsuccessful FISH ($n = 5$) or insufficient amounts of tissue ($n = 1$), six were excluded. Hence, lymph node metastases from 144 cases were included.

Cohort 2 METABRIC

The METABRIC dataset includes a discovery dataset ($n = 997$), and a validation dataset ($n = 995$). The cohorts have previously been described in detail [13]. In the present study, 1971 cases had available follow-up data and *MRPS23* gene expression data from all primary breast tumours. Tumours of basal-like ($n = 329$), normal-like ($n = 202$) and unknown subtype ($n = 6$) were excluded from our analysis [3, 16, 17]. Thus, 1434 tumours were included. Patients with ER positive and/or lymph node negative tumours had not received chemotherapy, whereas patients with ER negative and lymph node positive

tumours did [13]. None of the HER2+ patients were treated with trastuzumab. To assess possible associations between gene expression levels and prognosis, cases were separated into quartiles. Prognosis for each quartile was analysed separately, and for dichotomization of gene expression values, patients with gene expression levels in the upper quartile were compared to all other cases.

MRPS23 FISH cohort 1

FISH was done according to the manufacturer's guidelines using Dako Histology FISH Accessory Kit K 579911. After de-waxing and rehydration, TMA slides were boiled in a microwave oven (10 min) in Pre-Treatment Solution, cooled (15 min) and washed in Wash Buffer (2 × 3 min). Protein digestion was performed with Pepsin Solution at 37 °C (30 min), and then washed in Wash buffer (2 × 3 min). Dehydration was done in ethanol (70, 80 and 95%) for 2 min at each concentration, and the slides were then air dried at room temperature for 15 min. FISH-custom probes for *MRPS23* (3 µL, Empire Genomics) and CEP17 (1 µL, Abbott/VYSIS) were mixed with hybridization buffer (9 µL, Empire Genomics) and applied to TMA slides. Coverslips were applied and sealed with coverslip sealant (Dako). Denaturation was performed at 83 °C (3 min) followed by hybridization at 37 °C overnight in a DAKO Hybridizer. Post hybridization, TMA slides were rinsed in 0.4xSSC/0.3%NP-40 at 72 °C (2 min), and in 2xSSC/0.1%NP-40 at RT (15 s). Slides were air dried at 37 °C (15 min). DAPI (15 µL, VYSIS. Abbott no 06J50-001) was applied and the slides were coverslipped.

MRPS23 and CEP17 copy numbers were counted using a fluorescence microscope (Nikon Eclipse 90i). All available tissue cylinders from each case were examined, and *MRPS23* and CEP17 copy number in 20 well-preserved, non-overlapping tumour cell nuclei were recorded. Mean copy number of *MRPS23*/tumour cell and *MRPS23*/CEP17 ratio were estimated for each case. To assess the impact of mean *MRPS23* copy number, cases were divided into three categories based on recent HER2 guidelines [18]: mean *MRPS23* copy number < 4; mean ≥ 4 < 6 and mean ≥ 6. To estimate the impact of gene/centromere ratio, cases were divided into two categories: *MRPS23*/CEP17 < 2 and *MRPS23*/CEP17 ≥ 2. Finally, *MRPS23* amplification (*MRPS23*+) was defined as mean *MRPS23* copy number ≥ 6 and/or *MRPS23*/CEP17 ratio ≥ 2. Cases with *MRPS23* copy number < 6 and *MRPS23*/CEP17 ratio < 2 were defined as non-amplified (*MRPS23*−). To study the prognostic value of HER2 status and *MRPS23* status combined, cases were divided into four groups: *MRPS23*−/HER2−; *MRPS23*−/HER2+; *MRPS23*+/HER2− and *MRPS23*+/HER2+. The REMARK criteria for tumour marker reporting were followed [19].

Statistical analyses

Pearson chi square tests were used to compare *MRPS23* copy number status and *MRPS23* gene expression levels in the primary tumours across patient and tumour characteristics. To compare copy number status in the primary tumours and their corresponding lymph node metastases, paired analyses were performed using McNemar's and marginal homogeneity test.

For each category of *MRPS23* copy number status and *MRPS23*/HER2 status in primary tumours, and for the *MRPS23* gene expression categories, we estimated cumulative incidence of death from breast cancer, with death from other causes as a competing event. We used Gray's test to compare equality of cumulative incidence curves. Cox proportional hazard models were used to estimate hazard ratios (HR) of breast cancer death with 95% confidence intervals (CI). In the Cox regression analyses, patients were censored at time of death from other causes. Where applicable, adjustments were made for age at baseline (≤ 49, 50–59, 60–64, 65–69, 70–74, ≥ 75), histological grade (I–III), stage (I–IV), Ki67 status (< / ≥ 15%) and molecular subtype. We found no clear violations of proportionality in log-minus-log plots. Linear regression analyses were used for comparison of *MRPS23* expression levels between different molecular subtypes.

All statistical tests were two-sided, and statistical significance was assessed at 5% level. *p*-values between 5 and 10% were regarded borderline significant. All statistical analyses were done using STATA version 15.1 (Stata Corp., College Station, TX, USA).

Results

Cohort 1

Mean age at diagnosis was 75.6 (SD 8.6, range 41–96) years and mean follow-up after diagnosis was 8.9 (SD 7.2) years. By the end of follow-up, 217 (37%) patients had died from breast cancer, and 318 (54%) had died from other causes (Table 1).

MRPS23 copy number status in primary tumours

When *MRPS23* and CEP17 copy number alterations were present, a homogenous pattern was seen, with alterations in the majority of tumour cells. Three different phenotypes were seen, tumours without copy number alterations; tumours with *MRPS23* and CEP17 copy number increase and tumours with *MRPS23* copy number increase only (Fig. 1). In total, 29 of the primary tumours (5%) had mean *MRPS23* copy number ≥ 4 < 6, 21 (4%) had

Table 1 Characteristics of cohort 1

	Total study population	Categories defined by mean *MRPS23*				Categories defined by *MRPS23*/CEP17 ratio			Categories defined by amplification status		
		< 4	≥ 4 to < 6	≥ 6	*p* value (χ^2)	< 2	≥ 2	*p* value (χ^2)	*MRPS23*/CEP17 < 2 and mean *MRPS23* < 6	*MRPS23*/CEP17 ≥ 2 and/or mean *MRPS23* ≥ 6	*p* value (χ^2)
N (%)	590 (100)	540 (92)	29 (5)	21 (4)		549 (93)	41 (7)		545 (92)	45 (8)	
Mean age at diagnosis, years (SD)	75.6 (8.6)	75.6 (8.5)	76.3 (7.7)	74.4 (11.2)		75.6 (8.5)	74.5 (9.5)		75.6 (8.5)	74.4 (9.4)	
Mean follow-up, years (SD)	8.9 (7.2)	8.8 (7.2)	10.7 (8.1)	7.6 (6.6)		8.9 (7.2)	8.7 (7.6)		8.9 (7.2)	8.6 (7.3)	
Deaths from BC (%)	217 (37)	196 (36)	11 (38)	10 (48)		199 (36)	18 (44)		198 (36)	19 (42)	
Deaths from other causes (%)	318 (54)	292 (54)	15 (52)	11 (52)		297 (51)	21 (51)		294 (54)	24 (53)	
Histologic grade (%)											
I	70 (12)	69 (13)	0	1 (5)	< 0.001	69 (13)	1 (2)	0.006	69 (13)	1 (2)	0.001
II	327 (55)	308 (57)	12 (41)	7 (33)		309 (56)	18 (44)		308 (56)	19 (42)	
III	193 (33)	163 (30)	17 (58)	13 (61)		171 (31)	22 (54)		168 (31)	25 (56)	
Unknown											
Lymph node metastasis (%)											
Yes	192 (33)	173 (32)	13 (45)	6 (29)	0.691	176 (32)	16 (40)	0.283	175 (32)	17 (38)	0.458
No	245 (41)	226 (42)	12 (42)	7 (33)		231 (42)	14 (34)		228 (42)	17 (38)	
Unknown histology	153 (26)	141 (26)	4 (14)	8 (38)		142 (26)	11 (27)		142 (26)	11 (24)	
Tumour size (%)											
≤ 2 cm	282 (48)	260 (48)	15 (52)	7 (33)	0.40	262 (48)	20 (49)	0.514	261 (48)	21 (47)	0.264
> 2 cm, ≤ 5 cm	96 (16)	87 (16)	4 (14)	5 (24)		90 (16)	6 (15)		89 (16)	7 (16)	
> 5 cm	10 (2)	10 (2)	0	0		10 (2)	0		10 (2)	0	
Uncertain, but > 2 cm	74 (12)	63 (12)	6 (21)	5 (24)		66 (12)	8 (19)		64 (12)	10 (22)	
Uncertain	128 (22)	120 (22)	4 (14)	4 (19)		121 (22)	7 (17)		121 (22)	7 (16)	
Stage (%)											
I	286 (49)	267 (49)	9 (31)	10 (48)	0.024	269 (49)	17 (41)	0.157	267 (49)	19 (42)	0.182
II	239 (41)	214 (40)	17 (59)	8 (38)		220 (40)	19 (46)		219 (40)	20 (44)	
III	35 (6)	33 (6)	1 (3)	1 (5)		33 (6)	2 (5)		33 (6)	2 (4)	
IV	28 (5)	25 (5)	2 (7)	1 (5)		26 (5)	2 (5)		25 (5)	3 (7)	
Unknown	2	1 (0.2)	0	1 (5)		1 (0)	1 (2)		1 (0.2)	1 (2)	
Molecular subtype (%)											
Luminal A	309 (52)	295 (55)	11 (38)	3 (14)	< 0.001	301 (55)	8 (20)	< 0.001	301 (55)	8 (18)	< 0.001
Luminal B (HER2−)	148 (25)	131 (24)	9 (31)	8 (38)		132 (24)	16 (39)		130 (24)	18 (40)	
Luminal B (HER2+)	44 (7)	34 (6)	4 (14)	6 (29)		31 (6)	13 (32)		31 (6)	13 (29)	
HER2 type	33 (6)	26 (5)	3 (10)	4 (19)		29 (5)	4 (10)		27 (5)	6 (13)	
5NP	14 (2)	14 (3)	0	0		14 (3)	0		14 (3)	0	
BP	42 (7)	40 (7)	2 (7)	0		42 (8)	0		42 (8)	0	

Table 1 (continued)

	Total study population	Categories defined by mean *MRPS23*				Categories defined by *MRPS23*/CEP17 ratio			Categories defined by amplification status		
		< 4	≥ 4 to < 6	≥ 6	*p* value (χ^2)	< 2	≥ 2	*p* value (χ^2)	*MRPS23*/CEP17 < 2 and mean *MRPS23* < 6	*MRPS23*/CEP17 ≥ 2 and/or mean *MRPS23* ≥ 6	*p* value (χ^2)
Histological subtype (%)											
Ductal carcinoma	409 (69)	373 (69)	24 (83)	12 (57)	0.695	380 (69)	29 (71)	0.690	377 (69)	32 (71)	0.840
Lobular carcinoma	78 (13)	75 (14)	0	3 (14)		74 (13)	4 (10)		74 (14)	4 (9)	
Tubular carcinoma	3 (1)	3 (1)	0	0		3 (1)	0		3 (1)	0	
Mucinous carcinoma	30 (5)	28 (5)	1 (3)	1 (5)		28 (5)	2 (5)		28 (5)	2 (4)	
Medullary carcinoma	14 (2)	12 (2)	1 (3)	1 (5)		14 (3)	0		13 (2)	1 (2)	
Papillary carcinoma	27 (5)	23 (4)	2 (7)	2 (10)		24 (4)	3 (7)		24 (2)	3 (7)	
Metaplastic	8 (1)	8 (1)	0	0		8 (1)	0		8 (1)	0	
Other	21 (4)	18 (3)	1 (3)	2 (10)		18 (3)	3 (7)		18 (3)	3 (7)	
Ki67 high/low (%)											
Ki67 < 15%	349 (59)	330 (61)	13 (45)	6 (29)	0.003	335 (61)	14 (34)	0.001	334 (61)	15 (33)	< 0.001
Ki67 ≥ 15%	241 (41)	210 (39)	16 (55)	15 (71)		214 (39)	27 (66)		211 (39)	30 (66)	
Mitoses/10 HPF, median (IQR)	5 (1.12)	5 (1.12)	9 (3.17)	10 (5.16)		5 (1.12)	8 (3.16)		5 (1.11)	8 (4.17)	
Mitoses/10 HPF, quartiles (%)											
≤ 1	154 (26)	148 (27)	5 (17)	1 (5)	0.146	148 (27)	6 (15)	0.197	148 (27)	6 (13)	0.058
> 1, ≥ 5	158 (27)	146 (27)	6 (21)	6 (29)		148 (27)	10 (24)		147 (27)	11 (24)	
> 5, ≤ 12	140 (24)	126 (23)	8 (28)	6 (29)		129 (24)	11 (27)		129 (24)	11 (24)	
> 12	138 (23)	120 (22)	10 (34)	8 (38)		124 (23)	14 (34)		121 (22)	17 (38)	

N number of patients, *SD* standard deviation, *BC* breast cancer, *HER2* Human epidermal growth factor receptor 2, *5NP* five negative phenotype, *BP* basal phenotype, *HPF* High- power filed, *IQR* inter quartile range

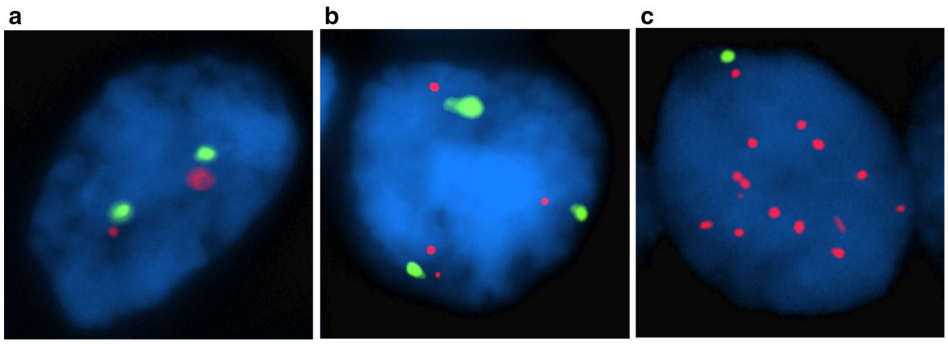


Fig. 1 Breast cancer cell nucleus with **a** two copies of *MRPS23* and CEP17, **b** copy number increase of both *MRPS23* and CEP17 and **c** copy number increase of *MRPS23* without corresponding increase of CEP17

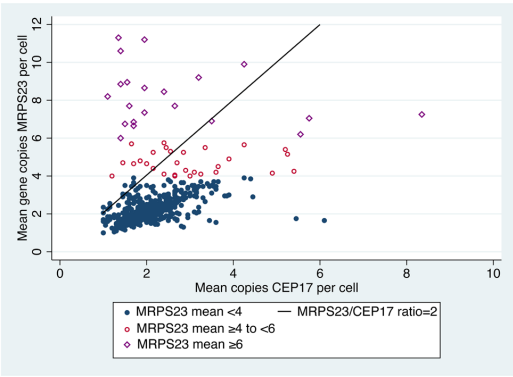


Fig. 2 Scatter plot of *MRPS23* and CEP17 copy number in 590 breast cancer tumours

mean copy number ≥ 6 and 41 (7%) had *MRPS23*/CEP17 ratio ≥ 2 (Table 1, Fig. 2). A total of 45 tumours (8%) were amplified (mean *MRPS23* ≥ 6 and/or *MRPS23*/CEP17 ≥ 2). Among cases with mean *MRPS23* ≥ 6, four cases had *MRPS23*/CEP17 ratio < 2. Of cases with *MRPS23*/CEP17 ratio ≥ 2, 24 had mean *MRPS23* < 6.

MRPS23 copy number increase (mean ≥ 4) was found within all molecular subtypes except the 5NP (Table 1). Amplifications were found in all molecular subtypes except the 5NP and the BP. Amplifications were seen in 30% of Luminal B (HER2+), 18% of HER2 type, 12% of Luminal B (HER2−), and 3% of Luminal A. Of the *MRPS23* amplified tumours, 19 (42%) were HER2+, compared to 58 (11%) of non-amplified tumours.

Copy number status in lymph node metastases

In total, 144 cases were examined for *MRPS23* copy number status in their lymph node metastases (Table 2). There were no significant changes in *MRPS23* copy number status in the lymph node metastases compared to the corresponding primary tumours. Among the pairs of primary tumours and lymph node metastases, 14 (10%) primary tumours were classified as *MRPS23*+, and in 10 of these (71%), the corresponding lymph node metastases were also *MRPS23*+. *MRPS23* amplification was also identified in the lymph node metastases of two *MRPS23*-tumours.

MRPS23 copy number status and proliferation

Of the *MRPS23* amplified tumours, 66% had high Ki67 (≥ 15%), compared to 39% of the non-amplified tumours (Table 1). Mitotic counts were also higher in amplified tumours (borderline significance). Of the *MRPS23* amplified tumours, 56% were grade III, compared to 31% of the non-amplified tumours. An association between *MRPS23* copy number increase and high Ki67, and high histological grade was also found when *MRPS23* copy number status was defined by *MRPS23* mean and *MRPS23*/CEP17 ratio (Table 1). In the lymph node metastases, *MRPS23* amplification was associated with high Ki67 (borderline significance, Table 3).

MRPS23 copy number status and prognosis

Mean *MRPS23*

The cumulative risk of death from breast cancer for cases with no copy number increase was 35% (95% CI 31–39) 10 years after diagnosis (Table 4; Fig. 3a). The corresponding risks for cases with mean copy number ≥ 4 < 6

Table 2 *MRPS23* status in primary tumours and lymph node metastases according to *MRPS23*/CEP17 ratio, mean *MRPS23* and amplification status

	Mean *MRPS23*/tumour cell, primary tumours				Marginal homogeneity test
	< 4	≥ 4, < 6	≥ 6	Total	
Mean *MRPS23*/tumour cell, lymph nodes					
< 4	125 (97)	4 (40)	1 (20)	130	$p = 0.637$
≥ 4, < 6	3 (2)	5 (50)	2 (40)	10	
≥ 6	1 (1)	1 (10)	2 (40)	4	
Total	129	10	5	144	
	MRPS23/CEP17 ratio, primary tumours				
	< 2	≥ 2	Total		McNemar test
MRPS23/CEP17 ratio, lymph nodes					
< 2	130 (99)	3 (23)	133		$p = 0.625$
≥ 2	1 (1)	10 (76)	11		
Total	131	13	144		
	Amplification status, primary tumours				
	MRPS23−	*MRPS23*+	Total		McNemar test
Amplification status, lymph nodes					
MRPS23−[a]	128 (98)	4 (29)	132		$p = 0.688$
MRPS23+[b]	2 (2)	10 (71)	12		
Total	130	14	144		

[a]*MRPS23*/CEP17 < 2 and mean *MRPS23* < 6

[b]*MRPS23*/CEP17 ≥ 2 and/or mean *MRPS23* ≥ 6

Table 3 *MRPS23* amplification status and Ki67 levels in lymph nodes

	MRPS23 amplification status, lymph node			
	MRPS23−[a]	*MRPS23*+[b]	Total	χ^2
Ki67, lymph node (%)				
Ki67 < 15%	69 (53%)	3 (25%)	72	$p = 0.07$
Ki67 > 15%	62 (47%)	9 (75%)	71	
Total	131	12	143	

[a]*MRPS23*/CEP17 < 2 and mean *MRPS23* < 6

[b]*MRPS23*/CEP17 ≥ 2 and/or mean *MRPS23* ≥ 6

and mean ≥ 6 was 34% (95% CI 20–55), and 43% (95% CI 25–66), respectively. In the Cox regression analysis, there were no clear differences in the rates of death between categories. Adjustments for age, stage, grade, Ki67 or HER2 status did not influence the results.

MRPS23/CEP17 ratio

After 10 years of follow-up, cases with *MRPS23*/CEP17 ratio < 2 had a cumulative risk of death from breast cancer of 30% (95% CI 27–34) (Table 4; Fig. 3b), whereas cases with ratio ≥ 2 had a corresponding risk of 41% (95% CI 30–58). When comparing rates of death from breast cancer, there was no clear difference between the two categories (HR 1.2, 95% CI 0.8–2.0). Adjusting for age, stage, grade, Ki67 or HER2 status did not influence the results.

MRPS23 amplification status

After 10 years of follow-up, patients with *MRPS23* amplified tumours had 40% (95% CI 27–56) cumulative risk of death from breast cancer, compared to 30% (95% CI 27–34) for patients without amplification (Table 4; Fig. 3c). The rates of death from breast cancer were similar for cases with and without amplification (HR 1.2, 95% CI 0.7–1.9). Separate adjustments for age, grade, histological grade, stage and HER2 status did not influence the results. Analysis of prognosis was also done for Luminal A cases separately, and for all luminal subtypes combined. In these analyses, no clear differences in prognosis between *MRPS23* amplified and non-amplified cases were seen (data not shown).

MRPS23/HER2− status

When tumours were reclassified into four categories based on *MRPS23* and HER2 status, the highest risk of death was

Table 4 Absolute and relative risk of death from breast cancer according to mean *MRPS23*/tumour cell, *MRPS23*/CEP17 ratio and amplification status

	Mean *MRPS23*/tumour cell, primary tumour			*MRPS23*/CEP17 ratio, primary tumour		Amplification status, primary tumours	
	< 4	≥ 4, < 6	≥ 6	< 2	≥ 2	*MRPS23*−[a]	*MRPS23*+[b]
Cum. risk after 5 years (%) (95% CI)	21 (18–24)	24 (12–44)	38 (21–62)	21 (18–25)	29 (18–46)	21 (18–25)	29 (18–44)
Cum. risk after 10 years (%) (95% CI)	35 (31–39)	34 (20–55)	43 (25–66)	30 (27–34)	41 (30–58)	30 (27–34)	40 (27–56)
HR, unadjusted (95% CI)	1.0	0.9 (0.5–1.7)	1.5 (0.8–2.8)	1.0	1.2 (0.8–2.0)	1.0	1.2 (0.7–1.9)
HR adjusted for age (95% CI)	1.0	0.9 (0.5–1.7)	1.5 (0.8–2.8)	1.0	1.3 (0.8–2.1)	1.0	1.2 (0.7–1.9)
HR adjusted for stage (95% CI)	1.0	0.8 (0.4–1.5)	1.2 (0.6–2.4)	1.0	1.2 (0.7–1.9)	1.0	1.0 (0.6–1.7)
HR adjusted for grade (95% CI)	1.0	0.7 (0.4–1.4)	1.3 (0.7–2.4)	1.0	1.1 (0.7–1.7)	1.0	1.0 (0.6–1.6)
HR adjusted for Ki67 (95% CI)	1.0	0.7 (0.4–1.4)	1.1 (0.6–2.2)	1.0	1.0 (0.6–1.6)	1.0	1.0 (0.6–1.6)
HR adjusted for HER2 (95% CI)	1.0	0.9 (0.5–1.6)	1.2 (0.6–2.3)	1.0	1.0 (0.6–1.7)	1.0	1.0 (0.6–1.6)

CI confidence interval, *HR* hazard ratio

[a]*MRPS23*/CEP17 < 2 and mean *MRPS23* < 6

[b]*MRPS23*/CEP17 ≥ 2 and/or mean *MRPS23* ≥ 6

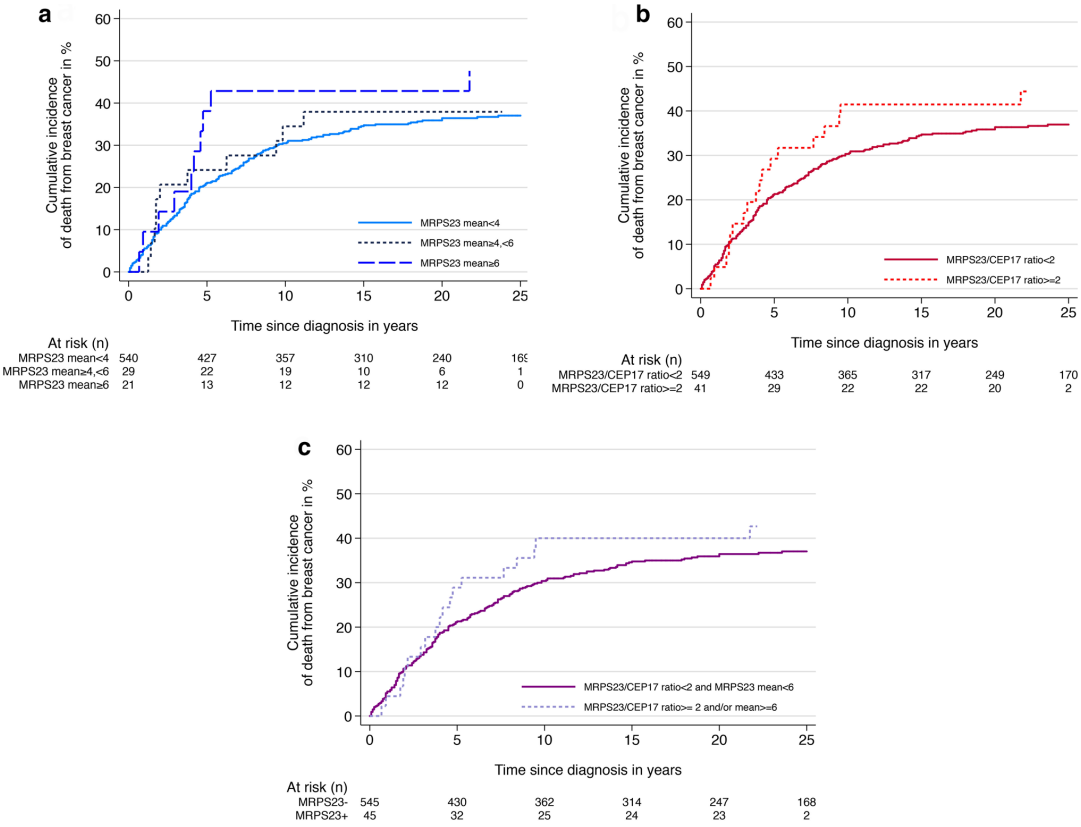


Fig. 3 Cumulative incidence of death from breast cancer according to *MRPS23* copy number status based on **a** mean *MRPS23* copy number ($p = 0.47$), **b** *MRPS23*/CEP17 ratio ($p = 0.29$) and **c** *MRPS23* amplification status ($p = 0.39$)

Table 5 Absolute and relative risk of dying from breast cancer according to HER2 status and amplification of *MRPS23*

	MRPS23−[a]/HER2−	*MRPS23*−/HER2+[c]	*MRPS23*+[b]/HER2−	*MRPS23*+/HER2+
Cum. risk after 5 years (%) (95% CI)	18 (15–22)	44 (32–58)	12 (4–32)	53 (33–76)
Cum. risk after 10 years (%) (95% CI)	28 (25–32)	47 (36–61)	23 (11–44)	63 (42–83)
HR, unadjusted (95% CI)	1.0	1.6 (1.1–2.4)	0.6 (0.3–1.4)	2.6 (1.5–4.5)
HR adjusted for age (95% CI)	1.0	1.7 (1.1–2.5)	0.6 (0.3–1.4)	2.6 (1.5–4.8)
HR adjusted for grade (95% CI)	1.0	1.2 (0.8–1.8)	0.5 (0.2–1.1)	2.3 (1.3–4.0)
HR adjusted for stage (95% CI)	1.0	1.9 (1.3–2.8)	0.6 (0.3–1.4)	1.8 (1.0–3.3)
HR adjusted for Ki67 (95% CI)	1.0	1.3 (0.9–1.9)	0.5 (0.2–1.0)	2.2 (1.3–3.0)
HR adjusted for HER2 (95% CI)	1.0	1.6 (1.1–2.4)	0.6 (0.3–1.4)	2.6 (1.5–4.5)

Cum. risk cumulative risk, *CI* confidence interval, *HR* hazard ratio

[a]*MPR23− : MPRS23*/CEP17 ratio < 2 and *MRPS23* mean < 6

[b]*MRPS23* + : *MRPS23*/CEP17 ratio ≥ 2 and/or *MRPS23* mean ≥ 6

[c]*HER2*/CEP17 ratio ≥ 2 or intense membranous staining (IHC 3+) in > 10% of tumour cells

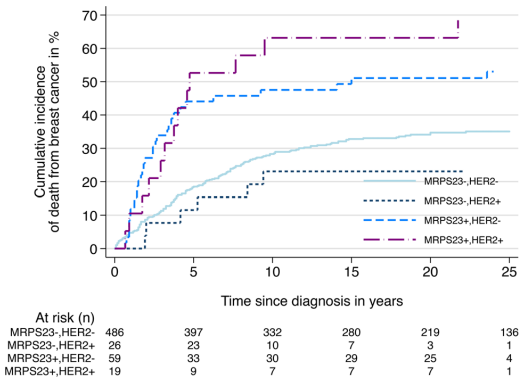


Fig. 4 Cumulative incidence of death from breast cancer according to *MRPS23* amplification status and HER2 status ($p < 0.01$)

found in the *MRPS23*+/HER2+ subtype (Table 5; Fig. 4). After 10 years of follow-up, patients with the *MRPS23*+/HER2+ subtype had a cumulative risk of death from breast cancer of 63% (95% CI 42–83). The corresponding risk for the *MRPS23*−/HER2+ subtype was 47% (95% CI 36–61). The lowest risk of death was found in the *MRPS23*+/HER2- subtype (23%, 95% CI 11–44). There were no clear differences in the rate of death from breast cancer between the *MRPS23*−/HER2+ and *MRPS23*+/HER2+ subtypes (HR 1.3, 95% CI 0.7–2.6). Separate analysis of prognosis was also done for all luminal cases, with similar results as for all cases combined (data not shown).

Cohort 2

MRPS23 expression according to molecular subtypes

In the METABRIC dataset, mean age at diagnosis was 61.1 (SD 12.4, range 22–96) years, and mean follow-up after diagnosis was 8.1 (SD 4.9) years. Characteristics of the study population are given in Table 6. By the end of follow-up, 506 (26%) patients had died from breast cancer, and 384 (20%) had died from other causes.

High *MRPS23* expression levels were associated with the Luminal B subtype (Fig. 5).

After 10 years of follow-up, cases with low *MRPS23* expression levels had a cumulative risk of death from breast cancer of 26% (95% CI 23–29%), compared to 24% (95% CI 20–30%) among cases with high *MRPS23* expression (Cut-off upper quartile; Fig. 6, Table 7). Comparing the rates of death from breast cancer, there were no significant differences between cases with gene expression levels in the upper quartile compared to the rest (HR 0.9, 95% CI 0.7–1.2, Fig. 6, Table 7). Similar results were obtained when analysis of prognosis was done for each quartile separately. Adjustments for age and histological grade did not influence the results.

Discussion

We identified *MRPS23* amplification in 8% of primary tumours and 9% of lymph node metastases in a large population of Norwegian breast cancer patients. The highest proportion of amplified cases was found within Luminal B (HER2+), HER2 type and Luminal B (HER2−) tumours. None of the amplified tumours were triple negative (5NP/BP). *MRPS23* amplification was associated with high Ki67 and high histological grade. No clear association between

Table 6 Characteristics of Cohort 2 (METABRIC) (Normal-like and basal-like subtypes excluded from the analyses)

	Total study population				Discovery cohort	Validation cohort
	Total	Mean probe *MRPS23*		χ^2		
		Quartile 1–3[a]	Quartile 4[b]			
N (%)	1434	1069	365		804 (56)	630 (44)
Mean age at diagnosis, years (SD)	63.2 (12.4)	63.6 (12.4)	62 (12.4)		62.3 (12.5)	64.3 (12.2)
Mean follow-up, years (SD)	8.3 (4.9)	8.5 (5)	7.8 (4.5)		8.1 (4.7)	8.5 (5.1)
Deaths from BC (%)	354 (25)	273 (25)	81 (22)		201 (25)	153 (24)
Deaths from other causes (%)	311 (22)	240 (22)	71 (19)		156 (19)	155 (25)
Histologic grade (%)						
I	139 (10)	116 (11)	23 (6)	< 0.001	61 (8)	78 (12)
II	635 (44)	504 (47)	131 (36)		375 (47)	260 (41)
III	598 (42)	392 (17)	206 (56)		368 (46)	230 (37)
Unknown	62 (4)	57 (5)	5 (1)		0	62 (10)
Lymph node metastasis (%)						
Yes	684 (48)	491 (46)	193 (53)	0.030	383 (48)	301 (48)
No	750 (52)	578 (54)	172 (47)		421 (52)	329 (52)
Tumour size (%)						
≤ 2 cm	613 (43)	472 (44)	141 (39)	0.170	351 (42)	262 (42)
> 2 cm, ≤ 5 cm	751 (52)	542 (51)	209 (57)		421 (52)	330 (53)
> 5 cm	69 (5)	54 (5)	15 (4)		32 (4)	37 (6)
Uncertain	1 (0)	1 (0)	0		0	1 (0)
Stage (%)						
I	104 (7)	84 (8)	20 (5)	0.075	0	104 (17)
II	177 (12)	141 (13)	36 (10)		0	177 (28)
III	27 (2)	20 (2)	7 (2)		0	27 (4)
IV	1 (0)	0	1 (0)		0	1 (0)
Unknown	1125 (78)	824 (77)	301 (82)		804 (100)	321 (50)
PAM50 subtype (%)						
Luminal A	709 (49)	606 (57)	103 (28)		445 (56)	255 (40)
Luminal B	488 (34)	276 (26)	212 (58)		266 (33)	222 (35)
HER2-type	237 (17)	187 (17)	50 (14)	< 0.001	84 (10)	153 (25)

N number of patients, *SD* standard deviation, *BC* breast cancer

[a]Mean probe *MRPS23* ≤ 8.31

[b]Mean probe *MRPS23* > 8.31

MRPS23 amplification and prognosis was seen. The proportion of HER2 positive cases was higher among *MRPS23* amplified cases, compared to non-amplified. *MRPS23*+/HER2+ had the poorest prognosis.

In the METABRIC dataset, Luminal B tumours had the highest level of *MRPS23* gene expression. We found no statistically significant associations between *MRPS23* expression levels and prognosis.

This study is based on a well-described cohort of breast cancer patients with long-term follow-up, and data from the METABRIC dataset. In the Norwegian cohort, the majority of patients have been followed until death [12]. Since relapse may occur even decades after the primary diagnosis, long-term follow-up is of particular value in breast cancer research. Molecular subtyping was performed in the same laboratory, using the same algorithm and antibodies in all cases [12]. Using FISH, gene copy number can be assessed while observing the morphology of the tumour, ensuring that only invasive tumour cells were examined.

MRPS23 copy number in primary tumours and lymph node metastases was assessed in TMAs. In the primary tumours, tissue for TMAs was taken from the tumour periphery. Previous studies have shown good correlation between TMAs and corresponding whole sections [20, 21]. Nevertheless, TMAs represent a small portion of each tumour, and, while copy number changes were observed throughout the tissue in amplified cases, intra tumour heterogeneity may not

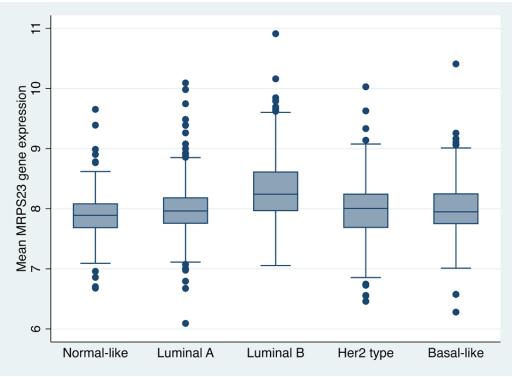


Fig. 5 *MRPS23* gene expression according to molecular subtype in 1971 patients from the METABRIC dataset

be captured. It would therefore be of interest to validate our findings in a study of whole sections.

There are no established guidelines as to how *MRPS23* amplification should be defined. According to *HER2* ISH guidelines, both *HER2* copy number and *HER2*/CEP17-ratio are taken into consideration [18]. *MRPS23* and *HER2* are both located on the long arm of chromosome 17 [4, 5, 22], and we chose to define *MRPS23* amplification according to *HER2* ISH guidelines, including both mean *MRPS23* copy number and *MRPS23*/CEP17 ratio in our definition. Previous studies have shown a high frequency of abnormalities on chromosome 17, but rarely true polysomy [23]. The number of amplified cases was increased when including *MRPS23*/CEP17 ratio in addition to mean *MRPS23* in the definition of amplification. Hence, our definition may have led to overestimation of *MRPS23* amplified cases. Nevertheless, we found that 8% of tumours were *MRPS23* amplified, whereas 20% and 33% of the tumours were amplified in the two datasets included in Gatza et al. In the latter two cohorts, only luminal (defined as non-basal) cases were included. When excluding the BP and 5NP in our in-house cohort, the proportion of amplified cases was still 8%. In accordance with other studies we found that amplification of *MRPS23* was associated with higher proliferation [3, 11]. However, contrary to others, we found no clear associations between *MRPS23* copy number increase and a poorer prognosis [3]. Our study demonstrates the importance of validating biomarkers identified by high-throughput genomic analyses. Validation analyses of single biomarkers with FISH, performed in FFPE tissue, indicate the marker's prognostic potential when assessed in a routine diagnostic setting.

MRPS23 has previously been found to be amplified exclusively in highly proliferative luminal tumours [3]. In that study, PAM50 was used for molecular subtyping, and "luminal" was defined as all tumours that were not basal [3]. This definition of luminal was based on a study showing that breast tumours could be separated into two main groups, one group containing luminal and HER2-positive tumours and the other group comprising basal-like tumours [17]. In our study, cohort 1 was divided into six subtypes based on IHC and ISH. Although it has been shown that surrogate markers can be used for molecular subtyping [24–27], there is a discrepancy between molecular subtype defined by surrogate markers and subtypes defined by gene expression analyses [28, 29]. Nevertheless, similar to Gatza et al., we only found *MRPS23* amplified cases among the non-basal tumours.

Contrary to others [3], we found no clear associations between *MRPS23* amplification and prognosis in our cohort of Norwegian breast cancer patients. We used a different method for assessment of gene copy number in

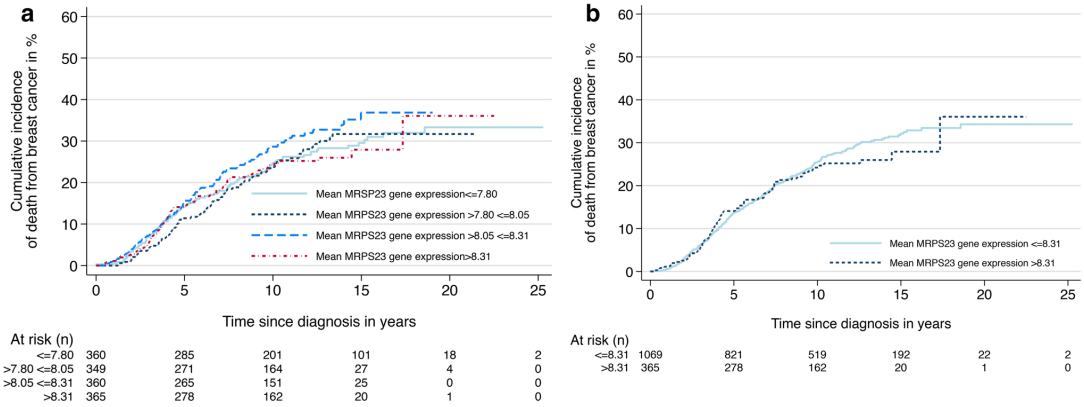


Fig. 6 Cumulative incidence of death from breast cancer according to *MRPS23* gene expression divided into **a** quartiles 1–4 ($p=0.4$), and **b** quartile 1–3 vs. quartile 4 ($p=0.6$)

Table 7 Absolute and relative risk of death from breast cancer according to gene expression levels of *MRPS23*, METABRIC data

	Mean probe MRPS23, quartiles				Mean probe MRPS23	
	Quartile 1[a]	Quartile 2[b]	Quartile 3[c]	Quartile 4[d]	Quartile 1–3	Quartile 4
Cum. risk after 5 years (%) (95% CI)	14 (11–18)	11 (8–15)	15 (12–20)	14 (11–18)	14 (12–16)	14 (11–18)
Cum. risk after 10 years (%) (95% CI)	25 (20–30)	23 (19–29)	29 (24–34)	24 (20–30)	26 (23–29)	24 (20–30)
HR, unadjusted (95% CI)	1	1.0 (0.7–1.3)	1.2 (0.9–1.6)	1.0 (0.7–1.3)	1	0.9 (0.7–1.2)
HR, adjusted for age (95% CI)	1	1.0 (0.7–1.4)	1.2 (0.9–1.6)	1.0 (0.7–1.4)	1	0.9 (0.7–1.2)
HR, adjusted for grade (95% CI)	1	1.0 (0.7–1.4)	1.2 (0.9–1.6)	0.9 (0.7–1.2)	1	0.8 (0.7–1.1)

Cum. risk cumulative risk, *CI* confidence interval, *HR* hazard ratio; Luminal A, Luminal B and HER2-type included in analyses

[a]Mean probe *MRPS23* ≤ 7.80

[b]Mean probe *MRPS23* > 7.80, ≤ 8.05

[c]Mean probe *MRPS23* > 8.05, ≤ 8.31

[d]Mean probe *MRPS23* > 8.31

our study (FISH), and different methodologies could partly explain the divergent results. Furthermore, the number of amplified cases was low in our study population, and the results must be interpreted with caution. Patient age in Cohort 1 is high, and the majority of patients were not given modern breast cancer treatment either due to their age at diagnosis or time of diagnosis [12]. This enables us to follow the near-natural course of disease after surgery. It would, however, be interesting to perform a *MRPS23* FISH study in a cohort of younger patients treated according to current guidelines.

Interestingly, the *MRPS23*+/HER2+ tumours had the poorest prognosis. HER2 is recognized as an important prognostic marker in breast cancer, and interaction with *MRPS23* could potentially be of clinical importance. However, due to the low number of cases in some of the *MRPS23*/HER2 categories, the results should be interpreted with caution.

We found no correlation between *MRPS23* expression levels and prognosis in the METABRIC data set. In our analyses of prognosis, we excluded basal-like and normal-like cases [3, 16]. A correlation between *MRPS23* copy number status and gene expression has previously been found in the METABRIC dataset [3]. Since transcription is regulated by several mechanisms, good correlation between gene copy number and gene expression is infrequent [30–32]. Our analyses show that *MRPS23* expression levels were overlapping between the different molecular subtypes. Such overlap was also seen between the highly proliferative luminal tumours (Luminal B) and basal-like tumours, the latter shown to be non-amplified. This could possibly be due to other *MRPS23* up-regulating mechanisms, and potentially explain the lack of correlation between gene expression levels and prognosis in the METABRIC dataset.

Conclusion

Using FISH on a large cohort of breast cancer patients we found that *MRPS23* amplification is associated with higher tumour cell proliferation. Amplifications were only found in luminal and HER2 type tumours. We found no association between *MRPS23* amplification and prognosis. In the METABRIC dataset, gene expression levels were highest in Luminal B tumours. There was no correlation between *MRPS23* expression and prognosis.

Acknowledgements Open Access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital). The study received financial support from the Medical Student Research Program at NTNU and UNIFOR. The authors thank the Department of Pathology, St. Olav´s Hospital, Trondheim University Hospital for making the archives available for the study and the Cancer Registry of Norway for providing the patient data.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This study was approved, including dispensation from the requirement

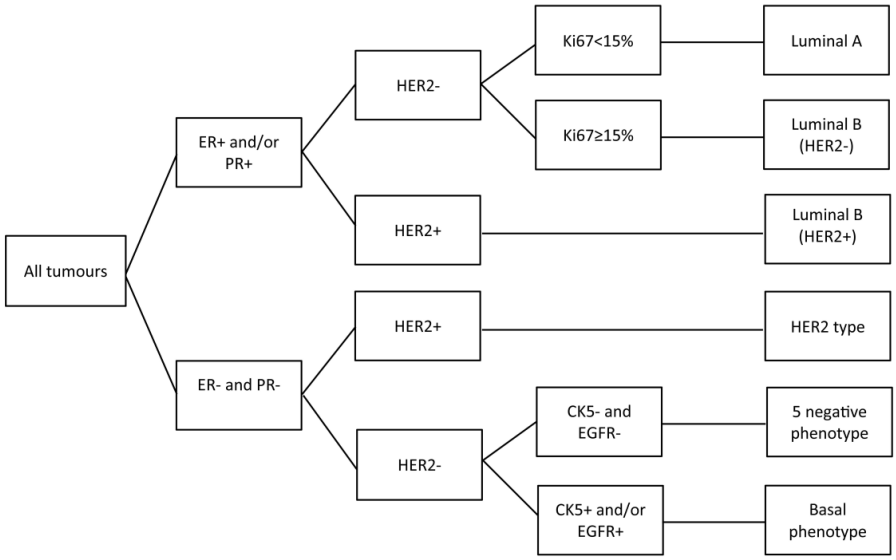
of patient consent, by the Regional committee for medical research ethics (REK 836-09).

References

1. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100(1):57–70
2. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144(5):646–674. https://doi.org/10.1016/j.cell.2011.02.013
3. Gatza ML, Silva GO, Parker JS, Fan C, Perou CM (2014) An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. Nat Genet 46(10):1051–1059. https://doi.org/10.1038/ng.3073
4. MRPS23 mitochondrial ribosomal protein S23 *Homo sapiens* (human) (2019) NCBI (National Center for Biotechnology Information). https://www.ncbi.nlm.nih.gov/gene/51649. Accessed 9 Jan 2019
5. Kenmochi N, Suzuki T, Uechi T, Magoori M, Kuniba M, Higa S, Watanabe K, Tanaka T (2001) The human mitochondrial ribosomal protein genes: mapping of 54 genes to the chromosomes and implications for human disorders. Genomics 77(1):65–70
6. Staub E, Gröne J, Mennerich D, Röpcke S, Klamann I, Hinzmann B, Castanos-Velez E, Mann B, Pilarsky C, Brümmendorf T, Weber B, Buhr HJ, Rosenthal A (2006) A genome-wide map of aberrantly expressed chromosomal islands in colorectal cancer. Mol Cancer 5:37. https://doi.org/10.1186/1476-4598-5-37
7. Li B, Zhang YL (2002) Identification of up-regulated genes in human uterine leiomyoma by suppression subtractive hybridization. Cell Res 12:215. https://doi.org/10.1038/sj.cr.7290127
8. Lyng H, Brøvig RS, Svendsrud DH, Holm R, Kaalhus O, Knutstad K, Oksefjell H, Sundfør K, Kristensen GB, Stokke T (2006) Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. BMC Genomics 7(1):268
9. Kurokawa Y, Matoba R, Nakamori S, Takemasa I, Nagano H, Dono K, Umeshita K, Sakon M, Monden M, Kato K (2004) PCR-array gene expression profiling of hepatocellular carcinoma. J Exp Clin Cancer Res 23(1):135–142
10. Pu M, Wang J, Huang Q, Zhao G, Xia C, Shang R, Zhang Z, Bian Z, Yang X, Tao K (2017) High MRPS23 expression contributes to hepatocellular carcinoma proliferation and indicates poor survival outcomes. Tumour Biol 39(7):1010428317709127. https://doi.org/10.1177/1010428317709127
11. Gao Y, Li F, Zhou H, Yang Y, Wu R, Chen Y, Li W, Li Y, Xu X, Ke C, Pei Z (2017) Down-regulation of MRPS23 inhibits rat breast cancer proliferation and metastasis. Oncotarget 8(42):71772–71781. https://doi.org/10.18632/oncotarget.17888
12. Engstrom MJ, Opdahl S, Hagen AI, Romundstad PR, Akslen LA, Haugen OA, Vatten LJ, Bofin AM (2013) Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. Breast Cancer Res Treat 140(3):463–473. https://doi.org/10.1007/s10549-013-2647-2
13. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486(7403):346
14. Kvale GHI, Eide GE (1987) A prospective study of reproductive factors and breast cancer. Am J Epidemiol 126(5):831–841
15. Lakhani SR, International Agency for Research on Cancer (2012) WHO classification of tumours of the breast. World Health Organization classification of tumours, 4th edn. IARC, Lyon
16. Prat A, Perou CM (2011) Deconstructing the molecular portraits of breast cancer. Mol Oncol 5(1):5–23. https://doi.org/10.1016/j.molonc.2010.11.003
17. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MD, Niu B, McLellan MD, Uzunangelov V (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell 158(4):929–944
18. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, Bilous M, Ellis IO, Fitzgibbons P, Hanna W, Jenkins RB, Press MF, Spears PA, Vance GH, Viale G, McShane LM, Dowsett M (2018) Human Epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. Arch Pathol Lab Med. https://doi.org/10.5858/arpa.2018-0902-SA
19. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2006) REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 100(2):229–235. https://doi.org/10.1007/s10549-006-9242-8
20. Batistatou A, Televantou D, Bobos M, Eleftheraki AG, Kouvaras E, Chrisafi S, Koukoulis GK, Malamou-Mitsi V, Fountzilas GJ (2013) Evaluation of current prognostic and predictive markers in breast cancer: a validation study of tissue microarrays. Anticancer Res 33(5):2139–2145
21. Dekker TJ, Ter Borg S, Hooijer GK, Meijer SL, Wesseling J, Boers JE, Schuuring E, Bart J, van Gorp J, Mesker WEJBCR (2012) Determining sensitivity and specificity of HER2 testing in breast cancer using a tissue micro-array approach. Breast Cancer Res 14(3):R93
22. ERBB2 erb-b2 receptor tyrosine kinase 2 [Homo sapiens (human)] (2019) NCBI (National Center for Biotechnology Information). https://www.ncbi.nlm.nih.gov/gene/2064. Accessed 24 June 2019
23. Marchiò C, Lambros MB, Gugliotta P, Di Cantogno LV, Botta C, Pasini B, Tan DS, Mackay A, Fenwick K, Tamber N, Bussolati G, Ashworth A, Reis-Filho JS, Sapino A (2009) Does chromosome centromere copy number predict polysomy in breast cancer? A fluorescence in situ hybridization and microarray-based CGH analysis. J Pathol. https://doi.org/10.1002/path.2574
24. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn H-J, Members P (2011) Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2011. Ann Oncol 22(8):1736–1747
25. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS (2009) Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst 101(10):736–750
26. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, Perou CM, Nielsen TO (2008) Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res 14(5):1368–1376
27. Blows FM, Driver KE, Schmidt MK, Broeks A, Van Leeuwen FE, Wesseling J, Cheang MC, Gelmon K, Nielsen TO, Blomqvist C

(2010) Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med 7(5):e1000279
28. Kim HK, Park KH, Kim Y, Park SE, Lee HS, Lim SW, Cho JH, Kim JY, Lee JE, Ahn JS, Im YH, Yu JH, Park YH (2018) Discordance of the PAM50 intrinsic subtypes compared with IHC-based surrogate in breast cancer patients: potential implication of genomic alterations of discordance. Cancer Res Treat. https://doi.org/10.4143/crt.2018.342
29. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, Davies SR, Snider J, Stijleman IJ, Reed J (2010) A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor–positive breast cancer. Clin Cancer Res 16:5222
30. Geiger T, Cox J, Mann M (2010) Proteomic changes resulting from gene copy number variations in cancer cells. PLoS Genet 6(9):e1001090
31. Fritz G, Brachetti C, Bahlmann F, Schmidt M, Kaina B (2002) Rho GTPases in human breast tumours: expression and mutation analyses and correlation with clinical parameters. Br J Cancer 87(6):635
32. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S (2014) Proteogenomic characterization of human colon and rectal cancer. Nature 513(7518):382

Supplementary Figure 1



Paper II

Breast Cancer Research and Treatment (2021) 186:65–77
https://doi.org/10.1007/s10549-020-06035-0

PRECLINICAL STUDY

ZNF703 gene copy number and protein expression in breast cancer; associations with proliferation, prognosis and luminal subtypes

Elise Klæstad[1] · Joanna Ewa Sawicka[2] · Monica Jernberg Engstrøm[1,3] · Borgny Ytterhus[1] · Marit Valla[1,4] · Anna Mary Bofin[1]

Received: 9 September 2020 / Accepted: 25 November 2020 / Published online: 2 January 2021

Abstract

Purpose Amplification of 8p12 is frequent in breast cancer and associated with poor prognosis in luminal subtypes. *ZNF703* has been identified as the driver gene of proliferation in the A1 amplicon situated in 8p12. In this study, the aims were to investigate associations between *ZNF703* copy number alterations and molecular subtypes, proliferation and prognosis, and using immunohistochemistry, examine associations between *ZNF703* copy number and ZNF703 protein expression.

Methods Copy number alterations in 702 primary breast tumours and corresponding lymph node metastases were examined using fluorescence in situ hybridization with probes for *ZNF703* and centromere 8. In addition, protein expression was studied in 869 tumours from the same cohort. Associations between copy number alterations and protein expression and tumour characteristics were assessed using Pearson chi square test. The prognostic impact of *ZNF703* copy number increase and protein expression was assessed estimating cumulative incidence of breast cancer death and hazard ratios.

Results We found mean *ZNF703* copy number ≥ 6 in 7% of tumours, most frequently in Luminal B subtypes. We found a positive association between increased copy number, and high proliferation, high histological grade, and poor prognosis. Luminal A tumours with high copy number had high histological grade and poor prognosis (borderline significant). We found positive nuclear staining in 76% of primary tumours. There was an association between copy number status and protein expression, but no association between protein expression and prognosis.

Conclusions In breast cancer, high *ZNF703* copy number is associated with increased proliferation, Luminal B subtypes and poor prognosis.

Keywords ZNF703 · Breast cancer · Proliferation · Copy number · Amplification · Luminal B

Introduction

Amplification of the chromosomal region 8p12 is frequent in breast cancer occurring in 10–15% of cases. It is associated with poor prognosis [1, 2]. Four amplicons, A1–4, have been identified in the chromosomal region 8p11–12 [1, 2]. The Zinc finger protein 703 gene (*ZNF703*) is located in the A1 amplicon which is the shortest and most telomeric of the four amplicons, and has been identified as the most likely driver gene in the A1 amplicon [1, 3–5]. In addition to breast cancer, high expression of *ZNF703* has been identified in gastric cancer [6], non-small cell lung cancer [7, 8], head and neck cancers [9–11] and cholangiocarcinoma [12]. *ZNF703* has been shown to differentially regulate mammary progenitor cells favouring luminal progenitor cells [13]. In breast cancer, it is most frequently expressed in the Luminal B subtype [3, 4, 13, 14]. Overexpression of *ZNF703* is

✉ Elise Klæstad
elisekl@stud.ntnu.no

[1] Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Erling Skjalgssons Gate, 7030 Trondheim, Norway

[2] Department of Pharmacology, Oslo University Hospital, Oslo, Norway

[3] Department of Breast and Endocrine Surgery, St. Olav's Hospital, Trondheim University Hospital, 7006 Trondheim, Norway

[4] Department of Pathology, St. Olav's Hospital, Trondheim University Hospital, 7006 Trondheim, Norway

associated with increased proliferation in breast cancer [3, 13, 15] and reduced survival in luminal tumours [13]. The gene has been suggested to play a role in oestrogen receptor (ER) signalling and cell cycle progression [3, 11, 13, 15].

A study of luminal breast cancer suggested that overexpression of the ZNF703 protein modulates cell cycle progression by shortening the G1 phase. *ZNF703* amplification tips the cancer cell in direction of self-renewal rather than differentiation [3]. In addition, overexpression of *ZNF703* mRNA in ER positive breast cancers is associated with reduced response to endocrine therapy (tamoxifen), proposedly through activation of the Atk/mTOR pathway, and down-regulation of the ERα pathway [16].

Luminal subtypes are by far the most common breast cancer subtypes and may be subdivided into Luminal A and B according to their expression of proliferation-associated genes [17, 18]. Protein biomarkers (Ki67, ER, and progesterone receptor (PR)) can be used as surrogates for gene expression analyses, to classify breast tumours as Luminal A and Luminal B. Luminal B may be further subdivided into Luminal B (human epidermal growth factor receptor 2 ($HER2)^+$) and Luminal B ($HER2^-$) depending on HER2 status [19, 20]. Clear differences in prognosis have been demonstrated between the different luminal subtypes [18, 19, 21, 22]. Identification of patients with an excellent prognosis, even without chemotherapy, is important in order to avoid unnecessary treatment of patients with low risk of recurrence. Equally, it is important to identify those who require more aggressive treatment. Both of these categories are found among luminal breast cancers [23].

The aims of the present study were to investigate the frequency of *ZNF703* copy number change in a well-characterized cohort of women with breast cancer [19] and to assess whether *ZNF703* copy number increase was associated with specific molecular subtypes, proliferation and prognosis. Furthermore, we assessed whether there was an association between *ZNF703* copy number and ZNF703 protein expression, and if ZNF703 protein expression was related to prognosis.

Materials and methods

Study population

Between 1956 and 1959, women in the county of Nord Trøndelag, Norway, were invited to take part in a population-based survey for the early detection of breast cancer [24]. A total of 25,727 women born between 1886 and 1928 were followed for breast cancer occurrence from January 1st, 1961, until December 31st, 2008 through linkage with the Norwegian Cancer Registry. Information on time and cause of death was obtained from the Norwegian Cause of Death Registry. In total, 1393 new cases of breast cancer were registered during follow-up. Of these, 909 were previously reclassified into the following molecular subtypes using immunohistochemistry (IHC) and chromogenic in situ hybridization (CISH): Luminal A (ER^+ and/or PR^+, $HER2^-$, Ki67 < 15%); Luminal B ($HER2^-$) (ER^+ and/or PR^+, $HER2^-$, Ki67 ≥ 15%); Luminal B ($HER2^+$) (ER^+ and/or PR^+, $HER2^+$); HER2 type (ER^- and PR^-, $HER2^+$); 5 negative phenotype (5NP) (ER^-, PR^-, $HER2^-$, cytokeratin 5 $(CK5)^-$, and epidermal growth factor receptor 1 $(EGFR)^-$); and Basal phenotype (BP) (ER^-, PR^-, $HER2^-$, $CK5^+$ and/or $EGFR^+$) [19]. After diagnosis, patients were followed until time of death or until December 31st, 2015.

Specimen characteristics

For the present study tissue microarrays (TMAs) were used. From each case, three 1 mm cores from the tumour periphery were included. TMAs were constructed using Tissue Arrayer MiniCore with TMA Designer2 software (Alphelys) [19]. Sections (4 µm) were cut from TMA blocks and stored on Superfrost slides at − 20 °C until use.

Of the 909 primary tumours, 856 were available in TMAs for fluorescence in situ hybridization (FISH). Of the available tumours, 115 were excluded due to unsuccessful FISH and 39 because of lack of tumour tissue, thus a total of 702 tumours were included in the FISH analyses. Of the 702 primary tumours, 241 had axillary lymph node metastases. Of these, 216 were available in TMAs. Of the available metastases, 27 were excluded due to unsuccessful FISH and 13 due to lack of tissue in the TMA. Thus, 176 lymph nodes metastases were included in the analyses.

FISH was done in accordance with the manufacturer's guidelines using DAKO Histology FISH Accessory Kit K 579911. Sections were heated at 60 °C for 1–2 h. Slides were de-waxed and rehydrated, and washed in DAKO wash buffer (2 min). The *ZNF703* (3 µL, Empire Genomics) and CEP8 (3 µL, Empire Genomics) probes were mixed with hybridizing buffer (9 µL) and applied to TMA slides. Slides were coverslipped, sealed, and dried in alufoil for 20 min. Denaturation was performed at 83 °C for 3 min, followed by hybridization overnight using DAKO hybridizer. Post-hybridization wash was done in DAKO Stringent wash buffer at 72 °C (2 min.) and DAKO wash buffer at room temperature (1 min). Slides were then dried at 37 °C for 15 min. DAPI II VYSIS (15 µl, no 06J50-001) was applied, and the slides were coverslipped and stored at − 20 °C.

For IHC, 867 of the 909 tumours were included in TMAs. Of these, 38 were excluded due to lack of tumour tissue in the TMA section. Thus, 829 tumours were included in the analyses. TMA sections were dried at 37 °C overnight and then at 60 °C for an hour. Sections were de-waxed and rehydrated in TissueClear (3 × 5 min.) and ethanol (100%,

96% and 80% 2 × 3 min. in each concentration) and rinsed in water (5 min.). Pre-treatment was done using PT-Link (DAKO). Immunostaining for ZNF703 was done in a DAKO Autostainer Plus using a monoclonal mouse antibody (clone 0654, Abcam), concentration 0.5 mg/mL, dilution 1:50. Counterstaining and dehydration was done in a Tissue-Tek Prisma: Hematoxylin (30 s) followed by dehydration in ethanol (80%, 96% and 100%, 2 × 3 min in each solution) and rinsing in TissueClear (3 × 3 min). Slides were covered with Tissue-Tek Prisma Glass g2 and sealed with TissueMount (Sakura).

Scoring and reporting

ZNF703 and CEP8 FISH

A fluorescence microscope (Nikon Eclipse 90i) was used for counting *ZNF703* and CEP8 copy number. For each case, all available tissue spots were examined and the numbers of fluorescent signals for *ZNF703* and CEP8 were counted in 20 well-preserved, non-overlapping tumour cell nuclei. Mean *ZNF703* and mean CEP8 copy number/tumour cell were estimated. The resultant mean copy numbers were then divided into three categories: <4; $\geq 4 < 6$; and ≥ 6. The REMARK criteria for tumour marker studies were followed [25].

ZNF703 immunohistochemistry

The proportion of tumour cells with nuclear ZNF703 immunostaining was assessed in a bright-field microscope. Each case was assessed independently by two pathologists. Cases were then divided into two groups based on the proportion of tumour cells with positive nuclear staining: $<50\%$ and $\geq 50\%$ ($ZNF703^-$ and $ZNF703^+$, respectively). Consensus was reached in cases with interobserver disagreement. In addition, cases were divided into four categories based on copy number status and nuclear staining; *ZNF703* copy number <6 and $ZNF703^-$; *ZNF703* copy number <6 and $ZNF703^+$; *ZNF703* copy number ≥ 6 and $ZNF703^-$; and *ZNF703* copy number ≥ 6 and $ZNF703^+$.

Statistical analyses

Pearson's chi square test was used to compare tumour characteristics across categories of *ZNF703* mean copy number and nuclear immunostaining. To compare copy numbers in primary tumours and corresponding lymph node metastases marginal homogeneity test was used. Cumulative incidence of death from breast cancer was estimated, and Gray's test was used to compare equality between cumulative incidence curves. Cox proportional hazard analyses were used to estimate hazard ratios (HR) of breast cancer death with 95% confidence intervals (CI). The analyses were adjusted for age (≤ 49, 50–59, 60–64, 65–69, 70–74, ≥ 75), stage (I–IV), histological grade (I-III), Ki67 status ($</\geq 15\%$), HER2 status and molecular subtype.

Results

FISH

Patient and tumour characteristics of the 702 women included in the FISH study are presented in Table 1. Mean age at diagnosis was 73.3 years and mean follow-up after diagnosis was 9.8 years. By the end of follow-up, 39% of the women had died from breast cancer, and 53% had died of other causes.

ZNF703 in primary tumours (FISH)

We found that 51 (7%) of the primary tumours had *ZNF703* copy number $\geq 4 < 6$, and 48 (7%) had copy number ≥ 6 (Table 1, Fig. 1a, b). *ZNF703* copy number ≥ 6 was found in all molecular subtypes except the Basal phenotype.

CEP8 in primary tumours

Of the tumours included in the FISH analyses, 674 (96%) had CEP8 copy number <4 and 26 (4%) had copy number $\geq 4 < 6$. Only two cases had CEP8 copy number ≥ 6. Among the 48 cases with *ZNF703* copy number ≥ 6, 43 (90%) had mean CEP8 <4, four (8%) had mean $\geq 4 < 6$ and one (2%) had mean ≥ 6.

ZNF703 in lymph nodes

In total, 176 of the cases were examined for *ZNF703* copy number status in the corresponding lymph node metastases (Table 2). There were no significant differences in *ZNF703* copy number status in the primary tumours and the corresponding lymph node metastases (marginal homogeneity test, $p = 0.7$). Of the 15 cases with mean *ZNF703* ≥ 6 in the primary tumour, 13 (97%) also had mean ≥ 6 in the lymph node metastases.

ER status, molecular subtypes and histological grade

Of the 702 cases in the FISH analyses, 593 (85%) of primary tumours were ER positive (1% cut-off level) (Table 1). Of these, 506 (85%) had *ZNF703* copy number <4, 42 (7%) had copy number $\geq 4 < 6$ and 45 (8%) had copy number ≥ 6. Among the ER^- tumours, 97 (89%) had *ZNF703* copy number <4, 9 (8%) had copy number $\geq 4 < 6$ and three (3%) cases

Table 1 Characteristics of the study population according to *ZNF703* mean copy number

	Total study population	Categories defined by mean *ZNF703*			
		< 4	≥ 4, < 6	≥ 6	*p* value (χ^2)
N (%)	702 (100)	603 (86)	51 (7)	48 (7)	
Mean age at diagnosis (SD)	73.3 (9.9)	73.7 (9.8)	71.1 (9.9)	71.2 (11)	
Mean follow-up after diagnosis (SD)	9.3 (8.0)	9.4 (8)	10.5 (8.6)	7.8 (7.5)	
Deaths from BC (%)	274 (39)	229 (38)	19 (37)	26 (54)	
Deaths from other causes (%)	369 (53)	323 (54)	26 (51)	20 (42)	
Grade (%)					
I	82 (12)	81 (14)	1 (2)	0	< 0.001
II	380 (54)	340 (56)	21 (41)	19 (40)	
III	240 (34)	182 (30)	29 (57)	29 (60)	
Lymph node metastasis					
Yes	241 (34)	203 (34)	17 (33)	21 (44)	0.4
No	283 (40)	249 (41)	17 (33)	17 (35)	
Unknown histology	178 (26)	151 (25)	17 (33)	10 (21)	
Tumour size					
≤ 2 cm	327 (46)	285 (47)	22 (43)	20 (42)	0.5
> 2 cm, ≤ 5 cm	96 (14)	83 (14)	4 (8)	9 (19)	
> 5 cm	10 (1)	10 (2)	0	0	
Uncertain, but > 2 cm	111 (16)	91 (15)	11 (22)	9 (19)	
Uncertain	158 (23)	134 (22)	14 (28)	10 (21)	
Stage					
1	346 (49)	296 (49)	30 (59)	20 (42)	0.2
2	277 (40)	243 (40)	13 (25)	21 (44)	
3	42 (6)	35 (6)	5 (10)	2 (4)	
4	34 (5)	27 (5)	3 (6)	4 (8)	
Unknown	3 (0)	2 (0)	0	1 (2)	
ER status					
< 1	109 (15)	97 (16)	9 (18)	3 (6)	0.04
≥ 1 < 10	23 (3)	23 (4)	0	0	
≥ 10 < 50	23 (3)	21 (4)	0	2 (4)	
≥ 50 < 90	86 (12)	67 (11)	12 (24)	7 (15)	
≥ 90	458 (65)	393 (65)	30 (58)	35 (73)	
Unknown	3 (0)	2 (0)	0	1 (2)	
Molecular subtype (%)					
Luminal A	349 (50)	324 (54)	16 (31)	9 (19)	< 0.001
Luminal B (HER2-)	185 (26)	135 (22)	23 (45)	27 (56)	
Luminal B (HER2 +)	51 (7)	40 (7)	2 (4)	9 (19)	
HER2 type	44 (6)	39 (7)	3 (6)	2 (4)	
5NP	23 (3)	20 (3)	2 (4)	1 (2)	
BP	50 (7)	45 (7)	5 (10)	0	
Histologic subtype					
Ductal	497 (71)	422 (70)	41 (80)	34 (71)	0.1
Lobular	87 (12)	81 (13)	2 (4)	4 (8)	
Tubular	2 (0)	2 (0)	0	0	
Mucinous	32 (5)	31 (5)	0	1 (2)	
Medullary	16 (2)	11 (2)	3 (6)	2 (4)	
Papillary	29 (4)	22 (4)	2 (4)	5 (10)	
Metaplastic	10 (1)	10 (2)	0	0	
Other	29 (4)	24 (4)	3 (6)	2 (4)	

Table 1 (continued)

	Total study population	Categories defined by mean *ZNF703*			
		< 4	≥ 4, < 6	≥ 6	*p* value (χ^2)
Ki67 high/low (%)					
Ki67 < 15%	395 (56)	364 (60)	17 (33)	14 (29)	< 0.001
Ki67 ≥ 15%	307 (44)	239 (40)	34 (67)	34 (71)	
Mitoses/10 HPF, median (IQR p25, p75)	5 (2,13)	5 (1,11)	13 (6,22)	13 (7,20)	
Mitoses/10HPF, quartiles (%)					
≤ 2	228 (32)	216 (36)	9 (18)	3 (7)	< 0.001
> 2, ≥ 5	124 (18)	113 (19)	3 (6)	8 (17)	
> 5, ≤ 13	181 (26)	153 (25)	14 (27)	14 (29)	
> 13	167 (24)	119 (20)	25 (49)	23 (48)	
Unknown	2	2 (0)	0	0	

N number of patients, *SD* standard deviation, *BC* breast cancer, *ER* Oestrogen receptor, *HER2* human epidermal growth factor receptor 2, *5NP* 5 negative phenotype, *BP* basal phenotype, *HPF* high-power field, *IQR* inter quartile range

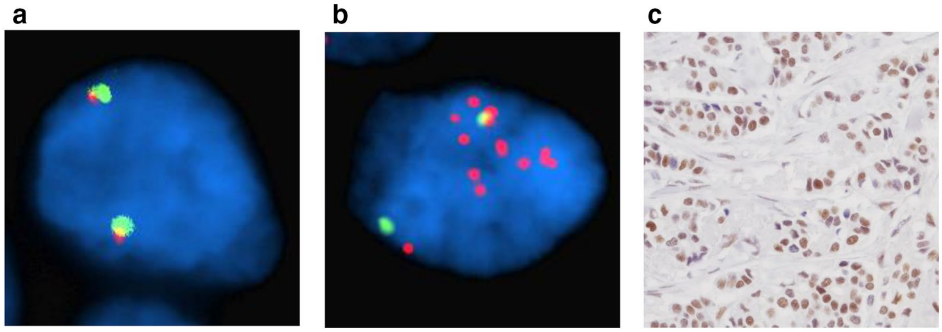


Fig. 1 **a** Breast cancer cell nucleus with two copies of both *ZNF703* and CEP8. **b** Breast cancer cell nucleus with increased *ZNF703* copy number. **c** Immunohistochemical staining showing positive nuclear staining for ZNF703 in breast cancer (×400 magnification)

Table 2 *ZNF703* status in primary tumours and lymph node metastases according to *ZNF703* mean copy number

	Mean *ZNF703*/tumour cell, primary tumours				
	< 4	≥ 4, < 6	≥ 6	Total	Marginal homogeneity test
Mean *ZNF703*/tumour cells, lymph nodes					
< 4	142 (97)	7 (46)	2 (13)	151	$p = 0.7$
≥ 4, < 6	3 (2)	4 (27)	0	7	
≥ 6	1 (1)	4 (27)	13 (87)	18	
Total	146	15	15	176	

had copy number ≥ 6. Within the molecular subtypes, mean *ZNF703* copy number ≥ 6 was found in nine (3%) of Luminal A cases, 27 (15%) of Luminal B ($HER2^-$), nine (18%) of Luminal B ($HER2^+$), two (5%) of HER2 type, two (4%) of 5NP and none of the BP (Table 1). Furthermore, 60% of tumours with mean copy number ≥ 6 were histological grade III, compared to 30% of tumours with mean copy number < 4 ($p < 0.01$). Of the nine Luminal A tumours with mean *ZNF703* copy number ≥ 6, four (44%) were histological grade II and five (56%) were grade III.

Proliferation (Ki67 and mitosis)

There was an association between *ZNF703* copy number increase and proliferation. In total, 71% of tumours with mean copy number ≥ 6 had Ki67 ≥ 15%, compared to 40% of tumours with mean copy number < 4 ($p < 0.01$, Table 1). Mitotic count was also higher in tumours with copy number ≥ 6 compared to tumours with mean < 4 (48% and 20% in the upper quartile, respectively, $p < 0.01$).

ZNF703 copy number and prognosis

After ten years of follow-up, patients with mean copy number ≥ 6 had a cumulative risk of death from breast cancer of 48% (95% CI 35–63%), compared to 32% (95% CI 28–36%) in patients with copy number < 4 (Table 3, Fig. 2, $p = 0.04$). In the Cox regression analysis, patients with *ZNF703* copy number ≥ 6 had a higher rate of death compared to those with copy number < 4 (HR 1.6 (95% CI 1.1–2.5). Adjustments for age, stage, grade, Ki67 and HER2 gave similar results.

Mean *ZNF703* in luminal subtypes

Among Luminal A cases, patients with *ZNF703* mean copy number ≥ 6 had a cumulative incidence of death from breast cancer of 56% (95% CI 28–71) after 10 years of follow-up, compared to 24% (95% CI 20–29) among cases with mean copy number < 4 ($p = 0.09$, borderline significance) (Table 4, Fig. 3a). Luminal B (HER2$^-$) cases with copy number ≥ 6 had a cumulative risk of death from breast cancer of 48% (95% CI 31–68) after ten years of follow-up, compared to 33% (95% CI 26–41) in cases with *ZNF703* mean < 4 ($p = 0.3$) (Table 4, Fig. 3b).

IHC

Patient and tumour characteristics of the 829 cases included in the IHC analyses are presented in Table 5. Mean age at diagnosis was 72.1 years, and mean follow-up after diagnosis was 10.4 years. By the end of follow-up,

Table 3 Absolute and relative risk of death from breast cancer according to *ZNF703* mean copy number

	ZNF703 mean		
	< 4	≥ 4, < 6	≥ 6
Cum incidence after 5 years (%) (95% CI)	22 (19–25)	29 (19–44)	31 (20–46)
Cum incidence after 10 years (%) (95% CI)	32 (28–36)	31 (21–46)	48 (35–63)
HR unadjusted (95% CI)	1	0.9 (0.6–1.5)	1.6 (1.1–2.4)
HR adjusted for age (95% CI)	1	0.9 (0.6–1.4)	1.6 (1.0–2.4)
HR adjusted for stage (95% CI)	1	1 (0.6–1.6)	1.6 (1.1–2.5)
HR adjusted for grade (95% CI)	1	0.8 (0.5–1.2)	1.3 (0.9–2)
HR adjusted for Ki67 (95% CI)	1	0.8 (0.5–1.3)	1.3 (0.9–2)
HR adjusted for HER2 (95% CI)	1	0.9 (0.6–1.5)	1.4 (0.9–2.1)

Cum incidence cumulative incidence, *HR* hazard ratio, *CI* confidence interval

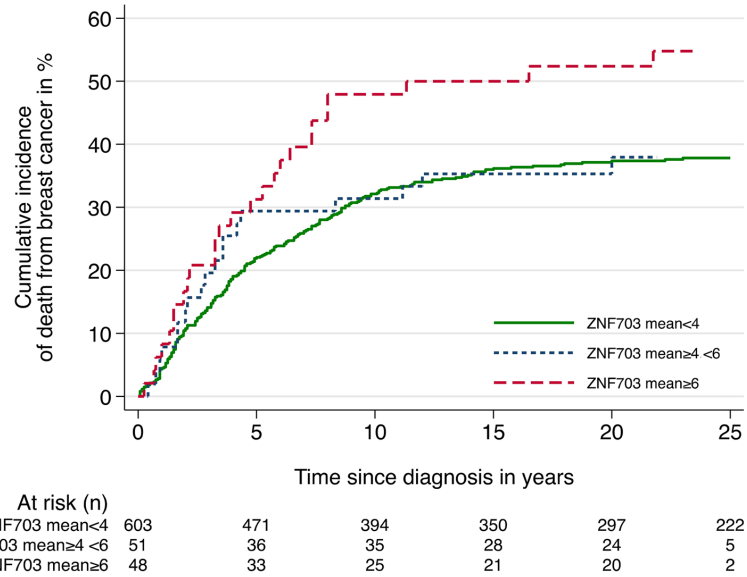


Fig. 2 Cumulative incidence of death from breast cancer according to *ZNF703* mean copy number ($p = 0.04$)

Table 4 Cumulative incidence of death from breast cancer according to *ZNF703* mean copy number in Luminal A and Luminal B (HER2$^-$) tumours

	Luminal A			Luminal B (HER2$^-$)		
	ZNF703 mean			*ZNF703* mean		
	<4	≥4, <6	≥6	<4	≥4, <6	≥6
Cum incidence after 5 years (%) (95% CI)	14 (11–18)	25 (10–54)	33 (12–72)	19 (14–27)	22 (10–45)	30 (16–51)
Cum incidence after 10 years (%) (95% CI)	24 (20–29)	25 (10–54)	56 (28–72)	33 (26–42)	26 (13–50)	48 (31–68)

Cum incidence cumulative incidence, *CI* confidence interval, *HER2* human epidermal growth factor receptor 2

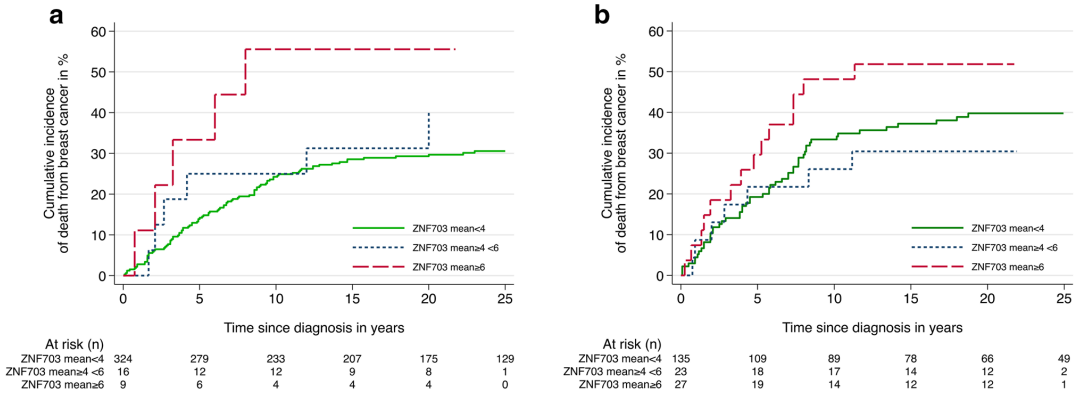


Fig. 3 Cumulative incidence of death from breast cancer according to *ZNF703* mean copy number in primary tumours of **a** Luminal A subtype ($p=0.09$) and **b** Luminal B (HER2−) subtype ($p=0.3$)

40% of the women had died from breast cancer, and 52% had died from other causes.

ZNF703 protein expression in primary tumours

We found positive nuclear staining (≥50%) in 628 (76%) of the primary tumours (Table 5, Fig. 1c). Positive staining was seen in all molecular subtypes of breast cancer. The luminal subtypes had the highest proportion of cases with positive staining, with 77% of Luminal A, 89% of Luminal B (HER2$^-$) and 83% of Luminal B (HER2$^+$). Among the non-luminal subtypes, 49% of HER2 type tumours, 56% of 5NP and 49% of BP showed positive staining (Table 5). Positive staining was seen in all categories of *ZNF703* copy number (Table 6). However, there was an association between copy number increase and positive IHC staining (70% and 94% were IHC positive among cases with mean *ZNF703* < 4 and mean ≥ 6, respectively ($p<0.01$)).

ZNF703 protein expression and prognosis

The cumulative risk of death from breast cancer for cases with positive nuclear staining (≥50%) was 32% (95% CI 29 – 36) after ten years of follow-up. For cases with negative staining the cumulative risk was 38% (95% CI 32–45) (Table 7, Fig. 4). Similarly, cases with positive staining had a lower rate of death compared to cases with negative staining (HR 0.8 (95% CI 0.6–1.0). Adjustments for age, stage, grade, Ki67 and HER2 did not influence the results significantly.

Copy number status, protein expression and prognosis

After ten years follow-up, cases with copy number ≥ 6 and positive IHC staining had the highest cumulative risk of death from breast cancer (47%, (95% CI 33–62)) (Table 8, Fig. 5, $p=0.01$). The lowest risk was seen among cases with mean < 6 and positive staining (29% (95% CI 25–34)). The

Table 5 Characteristics of the study population according to proportion of ZNF703 nuclear staining

	Study population	Nuclear ZNF703 staining		
		< 50%	≥ 50%	*p* value (χ^2)
N (%)	829	201 (24)	628 (76)	
Mean age at diagnosis, years (SD)	72.1 (10.4)	74 (9.6)	71.5 (10.6)	
Mean follow-up after diagnosis, years (SD)	9.5 (8.4)	7.7 (7.1)	10.1 (8.7)	
Deaths from BC (%)	332 (40)	85 (42)	247 (39)	
Deaths from other causes (%)	438 (52)	99 (49)	339 (54)	
Grade (%)				
I	97 (12)	36 (18)	61 (10)	< 0.001
II	440 (53)	85 (42)	355 (56)	
III	292 (35)	80 (40)	212 (34)	
Lymph node metastasis (%)				
Yes	289 (35)	77 (38)	212 (34)	0.5
No	324 (39)	78 (39)	246 (39)	
Unknown histology	216 (26)	46 (23)	170 (27)	
Tumour size (%)				
≤ 2 cm	371 (45)	85 (42)	286 (46)	0.05
> 2 cm, ≤ 5 cm	101 (12)	36 (18)	65 (10)	
> 5 cm	10 (1)	3 (2)	7 (1)	
Uncertain, but > 2 cm	141 (17)	31 (15)	110 (18)	
Uncertain	206 (25)	46 (23)	160 (25)	
Stage (%)				
1	409 (49)	89 (44)	320 (51)	0.3
2	326 (40)	89 (44)	237 (38)	
3	49 (6)	11 (5)	38 (6)	
4	41 (5)	12 (6)	29 (5)	
Unknown	4 (0)	0	4 (1)	
Molecular subtype (%)				
Luminal A	400 (48)	94 (47)	306 (49)	< 0.001
Luminal B ($HER2^-$)	222 (27)	25 (13)	197 (37)	
Luminal B ($HER2^+$)	63 (8)	11 (5)	52 (8)	
HER2 type	55 (7)	28 (14)	27 (4)	
5NP	30 (4)	13 (6)	17 (3)	
BP	59 (7)	30 (15)	29 (5)	
Histologic subtype (%)				
Ductal	582 (70)	132 (66)	450 (72)	< 0.001
Lobular	111 (13)	22 (11)	89 (14)	
Tubular	3 (0)	0	3 (0)	
Mucinous	35 (4)	10 (5)	25 (4)	
Medullary	20 (2)	10 (5)	10 (2)	
Papillary	31 (4)	7 (3)	24 (4)	
Metaplastic	12 (1)	9 (4)	3 (0)	
Other	35 (4)	11 (5)	24 (4)	
Ki67 high/low (%)				
Ki67 < 15%	465 (56)	124 (62)	341 (54)	0.07
Ki67 ≥ 15%	364 (44)	77 (38)	287 (46)	
Mitoses/10 HPF, median (IQR p25, p75)	6 (2, 13)	5 (1, 14)	6 (2, 12.5)	
Mitoses/10HPF, quartiles (%)				
≤ 2	267 (32)	71 (35)	196 (31)	0.2
> 2, ≥ 6	180 (22)	36 (18)	144 (23)	
> 6, ≤ 13	181 (22)	42 (21)	139 (22)	
> 13	200 (24)	51 (25)	149 (24)	
Unknown	1 (0)	1 (1)	0	

N number of patients, *SD* standard deviation, *BC* breast cancer, *HER2* human epidermal growth factor receptor 2, *5NP* 5 negative phenotype, *BP* basal phenotype, *HPF* high-power field, *IQR* inter quartile range

Table 6 *ZNF703* mean copy number and nuclear staining in primary tumours

	Mean copies *ZNF703*				
	< 4	≥ 4, < 6	≥ 6	Total	χ^2
Proportion of tumour cells with positive ZNF703 nuclear staining					
< 50%	176 (30)	8 (16)	3 (6)	187	< 0.001
≥ 50%	414 (70)	43 (84)	45 (94)	502	
Total	590	51	48	689	

Table 7 Absolute and relative risk of death from breast cancer according to proportion of ZNF703 nuclear staining in primary tumours

	Proportion of nuclear staining	
	< 50%	≥ 50%
Cum incidence after 5 years (%) (95% CI)	26 (23–35)	23 (20 -27)
Cum incidence after 10 years (%) (95% CI)	38 (32–46)	32 (29–36)
HR unadjusted (95% CI)	1	0.8 (0.6–1.0)
HR adjusted for age (95% CI)	1	0.8 (0.6–1.0)
HR adjusted for stage (95% CI)	1	0.8 (0.6–1.0)
HR adjusted for grade (95% CI)	1	0.8 (0.6–1.0)
HR adjusted for Ki67 (95% CI)	1	0.7 (0.6–1.0)
HR adjusted for HER2 (95% CI)	1	0.8 (0.6–1.0)

Cum incidence cumulative incidence, *HR* hazard ratio, *CI* confidence interval, *HER2* human epidermal growth factor receptor 2

number of cases with mean copy number ≥ 6 and negative staining was too small to be included in the analyses.

Discussion

In this study, 7% of primary tumours had high *ZNF703* copy number (mean ≥ 6). High *ZNF703* copy number was associated with Luminal B subtypes, increased proliferation, high histologic grade and poor prognosis. Copy number increase was found within all molecular subtypes. *ZNF703* copy number increase was not accompanied by an increase in CEP8 copy number. Using IHC, we found ZNF703 positive nuclear staining in 76% of primary tumours. ZNF703 positive nuclear staining was seen in all molecular subtypes; however, the highest proportion was seen among Luminal B tumours. There was a positive correlation between *ZNF703*

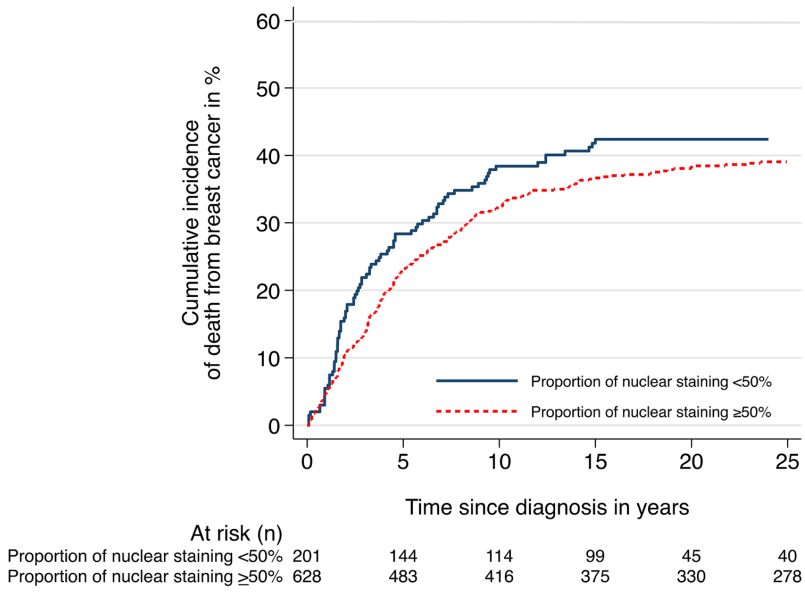


Fig. 4 Cumulative incidence of death from breast cancer according to proportion of ZNF703 nuclear staining in primary tumours ($p = 0.2$)

Table 8 Cumulative incidence of death from breast cancer according to *ZNF703* mean copy number and proportion of ZNF703 nuclear staining in primary tumours

	Mean < 6 and < 50%	Mean < 6 and ≥ 50%	Mean ≥ 6 and < 50%	Mean ≥ 6 and ≥ 50%
Cum incidence after 5 years (%) (95% CI)	28 (22–35)	20 (17–24)	33 (6–94)	31 (20–47)
Cum incidence after 10 years (%) (95% CI)	38 (32–46)	29 (25–34)	–	47 (33–62)

Cum incidence cumulative incidence, *CI* confidence interval

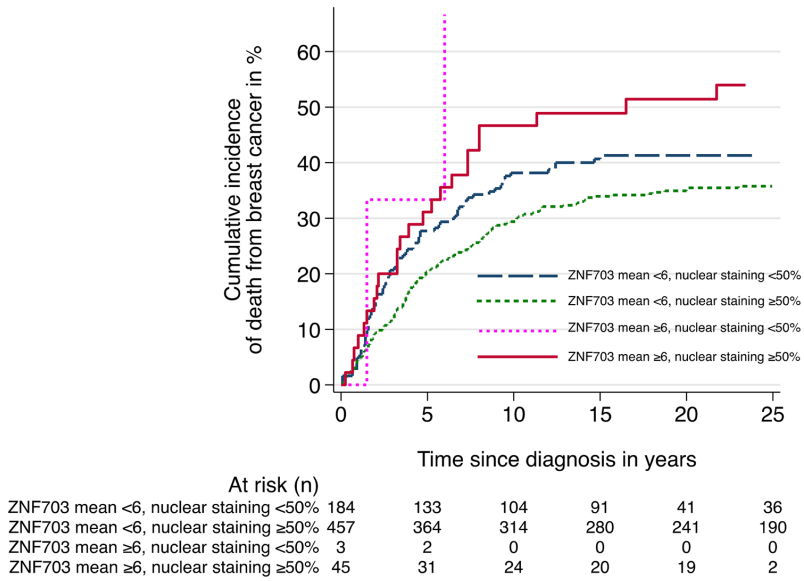


Fig. 5 Cumulative incidence of death from breast cancer according to *ZNF703* mean copy number and proportion of nuclear staining in primary tumours ($p = 0.01$)

copy number increase and positive immunostaining. However, we found no significant association between ZNF703 nuclear staining and prognosis.

This study is based on a well-described cohort of Norwegian breast cancer patients with long-term follow-up. Data on patient outcome was acquired from high quality national registries [26]. Most patients were followed until time of death [19]. Since breast cancer patients may experience relapse many years after the primary diagnosis [19, 21], long follow-up is of particular importance in breast cancer research and clinical management. Molecular subtyping was done in the same laboratory, using the same algorithm, and the same biomarkers and cut-off levels for all cases. In situ methods such as FISH and IHC ensure that only invasive tumour cells are included for assessment. For ZNF703 IHC, each case was evaluated by two pathologists independently, with consensus agreement in cases of discrepancy. There are also some limitations to this study. The cases in the study were diagnosed over a period of 40 years and fixation and preanalytical procedures may have varied over time possibly affecting the level and quality of IHC staining, and rendering some samples unsuited for FISH analysis. Furthermore, treatment protocols varied considerably over this time span. The TMA method allows for targeted analyses of highly proliferative areas in the tumour periphery. However, intratumoural heterogeneity and representativity may represent challenges when working with TMAs. Furthermore, some of the subgroups had few cases and the results must be interpreted accordingly.

We identified high *ZNF703* copy number (mean ≥ 6) in 7% of cases. This is in concordance with others who, using gene expression analyses, found *ZNF703* amplification in 8% of breast tumours [4], and in 9–19% of ER^+ tumours [13, 27].

Using FISH, we found that 96% of tumours had CEP8 copy number < 4. Of the cases with *ZNF703* mean copies ≥ 6 only one case (2%) had CEP8 ≥ 6. This supports previous studies identifying *ZNF703* as part of an amplicon in the 8p chromogenic region that does not include CEP8 [1, 3, 13].

A previous study has indicated that increased copy numbers of *HER2* and *ZNF703* in luminal tumours are mutually exclusive [13]. However, we found that nine (19%) tumours with *ZNF703* copy number ≥ 6 were Luminal B ($HER2^+$). All Luminal B ($HER2^+$) cases in this study had their HER2 status defined by CISH. Holland et al. defined the intrinsic subtypes based on gene expression patterns using the PAM50 classifier [13]. Even though studies have shown good correlation between molecular subtypes defined by surrogate markers and by gene expression analyses, the methods are not identical [23, 28, 29].

Holland et al. found *ZNF703* copy number gain or amplification in 23.2% of ER^+ tumours and in 11.7% of ER^- tumours. In comparison, we found *ZNF703* copy number ≥ 6 in 7.5% of ER^+ tumours and 3% of ER^- tumours. In our study, ER status was determined using IHC, while Holland used the PAM50 classifier. Holland et al. found a positive association between mRNA levels and protein expression of ZNF703, both being significantly higher in ER^+ breast tumours [13].

Previous studies show that Luminal B tumours with *ZNF703* amplification have significantly poorer survival than non-amplified Luminal B tumours [3, 13]. In contrast to these studies, we did not find a significantly poorer prognosis among Luminal B tumours with increased copy number of *ZNF703,* compared to cases without. We found that Luminal A cases with increased *ZNF703* copy number had significantly poorer prognosis than those without copy number increase. However, the number of amplified cases was small in both subtypes and results must be interpreted with caution. Among Luminal A tumours with *ZNF703* copy number ≥ 6, the majority were of histological grade III. It is therefore possible that some of these tumours may represent misclassified Luminal B tumours [30].

In accordance with previous studies we found that ZNF703 was expressed mainly in the tumour cell nuclei [3, 16]. We also found a positive correlation between gene copy number and ZNF703 protein expression, as previously shown by others [3, 13]. However, the proportion of tumours with positive staining was much higher than the proportion of cases with copy number increase. Even though there was an association between *ZNF703* copy number increase and a poor prognosis, and between copy number and protein expression, we found no association between protein expression and prognosis.

In conclusion, *ZNF703* copy number increase was associated with increased proliferation, a poor prognosis, and the Luminal B subtypes. We identified a subgroup of Luminal A tumours with *ZNF703* copy number increase and poorer prognosis. Since luminal tumours comprise a great proportion of all breast cancers, identification of new prognostic markers and potential treatment targets within this large subgroup of patients could prove to be of clinical importance [23].

Acknowledgements The Medical Student Research Program, NTNU and Rakel and Otto Kr Bruun's legacy has given the study financial support. The authors thank the Department of Pathology, St. Olav´s Hospital, Trondheim University Hospital for making the archives available for the study and the Cancer Registry of Norway for providing the patient data.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest to declare.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This study was granted approval including dispensation from the general requirement of patient consent, by the Regional Committee for Medical and Health Research Ethics, Midt Norge (REK 836/2009). Permission for linkage between data from the Cancer Registry of Norway and the Cause of Death Registry has been granted by the Norwegian Data Inspectorate.

References

1. Gelsi-Boyer V, Orsetti B, Cervera N, Finetti P, Sircoulomb F, Rouge C, Lasorsa L, Letessier A, Ginestier C, Monville F, Esteyries S, Adelaide J, Esterni B, Henry C, Ethier SP, Bibeau F, Mozziconacci MJ, Charafe-Jauffret E, Jacquemier J, Bertucci F, Birnbaum D, Theillet C, Chaffanet M (2005) Comprehensive profiling of 8p11-12 amplification in breast cancer. Mol Cancer Res 3(12):655–667. https://doi.org/10.1158/1541-7786.MCR-05-0128
2. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10(6):529–541. https://doi.org/10.1016/j.ccr.2006.10.009
3. Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousselet E, Lonigro A, Adelaide J, Baudelet E, Esteyries S, Wicinski J, Audebert S, Charafe-Jauffret E, Jacquemier J, Lopez M, Borg JP, Sotirious C, Popovici C, Bertucci F, Birnbaum D, Chaffanet M, Ginestier C (2011) ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. Embo Mol Med 3(3):153–166. https://doi.org/10.1002/emmm.201100121
4. Reynisdottir I, Arason A, Einarsdottir BO, Gunnarsson H, Staaf J, Vallon-Christersson J, Jonsson G, Ringner M, Agnarsson BA, Olafsdottir K, Fagerholm R, Einarsdottir T, Johannesdottir G, Johannsson OT, Nevanlinna H, Borg A, Barkardottir RB (2013) High expression of ZNF703 independent of amplification indicates worse prognosis in patients with luminal B breast cancer. Cancer Med 2(4):437–446. https://doi.org/10.1002/cam4.88
5. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours

reveals novel subgroups. Nature 486(7403):346–352. https://doi.org/10.1038/nature10983
6. Yang G, Ma F, Zhong M, Fang L, Peng Y, Xin X, Zhong J, Yuan F, Gu H, Zhu W, Zhang Y (2014) ZNF703 acts as an oncogene that promotes progression in gastric cancer. Oncol Rep 31(4):1877–1882. https://doi.org/10.3892/or.2014.2997
7. Baykara O, Bakir B, Buyru N, Kaynak K, Dalay N (2015) Amplification of chromosome 8 genes in lung cancer. J Cancer 6(3):270–275. https://doi.org/10.7150/jca.10638
8. Baykara O, Dalay N, Kaynak K, Buyru NJCm, (2016) ZNF 703 Overexpression may act as an oncogene in non-small cell lung cancer. Cancer Med 5(10):2873–2878
9. Yang H, Jiang WQ, Cao Y, Sun YA, Wei J, An X, Zhang YC, Song M, Wang SS, Yuan ZY, Peng RJ, Chen TH, Li LR, Shi YX (2015) Elevated ZNF703 protein expression is an independent unfavorable prognostic factor for survival of the patients with head and neck squamous cell carcinoma. Dis Markers 2015:640263. https://doi.org/10.1155/2015/640263
10. Baltaci E, Karaman E, Dalay N, Buyru N (2018) Analysis of gene copy number changes in head and neck cancer. Clin Otolaryngol 43(4):1004–1009. https://doi.org/10.1111/coa.12686
11. Wang H, Deng X, Zhang J, Ou Z, Mai J, Ding S, Huo S (2017) Elevated expression of zinc finger protein 703 promotes cell proliferation and metastasis through PI3K/AKT/GSK-3beta signalling in oral squamous cell carcinoma. Cell Physiol Biochem 44(3):920–934. https://doi.org/10.1159/000485360
12. Li K, Wang J, Han J, Lan Y, Xie C, Pan S, Liu L (2016) Overexpression of ZNF703 facilitates tumorigenesis and predicts unfavorable prognosis in patients with cholangiocarcinoma. Oncotarget 7(46):76108–76117. https://doi.org/10.18632/oncotarget.12627
13. Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, Chin SF, Hurtado A, Bruna A, Ali HR, Greenwood W, Dunning MJ, Samarajiwa S, Menon S, Rueda OM, Lynch AG, McKinney S, Ellis IO, Eaves CJ, Carroll JS, Curtis C, Aparicio S, Caldas C (2011) ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. Embo Mol Med 3(3):167–180. https://doi.org/10.1002/emmm.201100122
14. Glodzik D, Purdie C, Rye IH, Simpson PT, Staaf J, Span PN, Russnes HG, Nik-Zainal S (2018) Mutational mechanisms of amplifications revealed by analysis of clustered rearrangements in breast cancers. Ann Oncol 29(11):2223–2231. https://doi.org/10.1093/annonc/mdy404
15. Slorach EM, Chou J, Werb Z (2011) Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. Genes Dev 25(5):471–484. https://doi.org/10.1101/gad.1998111
16. Zhang X, Mu X, Huang O, Xie ZQ, Jiang M, Geng MY, Shen KW (2013) Luminal breast cancer cell lines overexpressing ZNF703 are resistant to tamoxifen through activation of Akt/mTOR signaling. PLoS ONE 8(8):10. https://doi.org/10.1371/journal.pone.0072053
17. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. Nature 406(6797):747–752. https://doi.org/10.1038/35021093
18. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98(19):10869–10874. https://doi.org/10.1073/pnas.191367098
19. Engstrom MJ, Opdahl S, Hagen AI, Romundstad PR, Akslen LA, Haugen OA, Vatten LJ, Bofin AM (2013) Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. Breast Cancer Res Treat 140(3):463–473. https://doi.org/10.1007/s10549-013-2647-2
20. Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, Cheang MC, Gelmon K, Nielsen TO, Blomqvist C, Heikkila P, Heikkinen T, Nevanlinna H, Akslen LA, Begin LR, Foulkes WD, Couch FJ, Wang X, Cafourek V, Olson JE, Baglietto L, Giles GG, Severi G, McLean CA, Southey MC, Rakha E, Green AR, Ellis IO, Sherman ME, Lissowska J, Anderson WF, Cox A, Cross SS, Reed MW, Provenzano E, Dawson SJ, Dunning AM, Humphreys M, Easton DF, Garcia-Closas M, Caldas C, Pharoah PD, Huntsman D (2010) Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med 7(5):e1000279. https://doi.org/10.1371/journal.pmed.1000279
21. Valla M, Vatten LJ, Engstrom MJ, Haugen OA, Akslen LA, Bjorngaard JH, Hagen AI, Ytterhus B, Bofin AM, Opdahl S (2016) Molecular subtypes of breast cancer: long-term incidence trends and prognostic differences. Cancer Epidemiol Biomarkers Prev 25(12):1625–1634. https://doi.org/10.1158/1055-9965.Epi-16-0427
22. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100(14):8418–8423. https://doi.org/10.1073/pnas.0932692100
23. Prat A, Pineda E, Adamo B, Galvan P, Fernandez A, Gaba L, Diez M, Viladot M, Arance A, Munoz M (2015) Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast 24(Suppl 2):S26-35. https://doi.org/10.1016/j.breast.2015.07.008
24. Kvale GHI, Eide GE (1987) A prospective study of reproductive factors and breast cancer. Am J Epidemiol 126(5):831–841
25. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, Statistics Subcommittee of NCIEWGoCD (2006) REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 100(2):229–235. https://doi.org/10.1007/s10549-006-9242-8
26. Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, Møller B (2009) Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer 45(7):1218–1231. https://doi.org/10.1016/j.ejca.2008.10.037
27. Fumagalli D, Wilson TR, Salgado R, Lu X, Yu J, O'Brien C, Walter K, Huw LY, Criscitiello C, Laios I, Jose V, Brown DN, Rothe F, Maetens M, Zardavas D, Savas P, Larsimont D, Piccart-Gebhart MJ, Michiels S, Lackner MR, Sotiriou C, Loi S (2016) Somatic mutation, copy number and transcriptomic profiles of primary and matched metastatic estrogen receptor-positive breast cancers. Ann Oncol 27(10):1860–1866. https://doi.org/10.1093/annonc/mdw286
28. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, Davies SR, Snider J, Stijleman IJ, Reed J, Cheang MC, Mardis ER, Perou CM, Bernard PS, Ellis MJ (2010) A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clin Cancer Res 16(21):5222–5232. https://doi.org/10.1158/1078-0432.Ccr-10-1282
29. Viale G, Slaets L, Bogaerts J, Rutgers E, Van't Veer L, Piccart-Gebhart MJ, de Snoo FA, Stork-Sloots L, Russo L, Dell'Orto P, van den Akker J, Glas A, Cardoso F (2014) High concordance of protein (by IHC), gene (by FISH; HER2 only), and microarray readout (by TargetPrint) of ER, PgR, and HER2: results

from the EORTC 10041/BIG 03–04 MINDACT trial. Ann Oncol 25(4):816–823. https://doi.org/10.1093/annonc/mdu026

30. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thürlimann B, Senn HJ (2015) Tailoring therapies-improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol 26(8):1533–1546. https://doi.org/10.1093/annonc/mdv221

Paper III

This paper was included in the thesis as a manuscript. A revised version was published in BMC Cancer, December 2022.

Klæstad E, Opdahl S, Raj SX, Bofin AM, Valla M. **Long term trends of breast cancer incidence according to proliferation status.** *BMC Cancer*. 2022;22(1):1340. Published 2022 Dec 21. doi:10.1186/s12885-022-10438-1

Long term trends of breast cancer incidence according to proliferation status

Elise Klæstad[1], Signe Opdahl[2], Sunil Xavier Raj[3], Anna Mary Bofin[1], Marit Valla[1,4,5]

Affiliations

[1] Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway

[2] Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway

[3] Cancer Clinic, St. Olav's Hospital, Trondheim University Hospital, 7006 Trondheim, Norway

[4] Department of Pathology, St. Olav's Hospital, Trondheim University Hospital, 7006 Trondheim, Norway

[5] Clinic of Laboratory Medicine, St. Olav's Hospital, Trondheim University Hospital, 7006 Trondheim, Norway

Keywords: breast cancer, incidence trends, proliferation, Ki-67, mitotic count,

Corresponding author: Elise Klæstad, Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Erling Skjalgssons gate, 7030 Trondheim, Norway

elise.klastad@ntnu.no

Abstract

Background: Long-term breast cancer incidence trends according to proliferation status are poorly described. We studied time-trends in breast cancer incidence, using mitotic count and Ki-67 as markers of proliferation.

Methods: Among 83 298 Norwegian women followed for breast cancer occurrence 1961-2012, 2995 incident breast cancers were diagnosed. Ki-67 was assessed using immunohistochemistry on tissue microarrays and mitoses were counted on whole sections. We compared incidence rates according to proliferation status among women born 1886-1928 and 1929-1977, estimating age-specific incidence rate ratios. We performed multiple imputations to account for unknown proliferation status. Mean values of Ki-67 and mitotic counts were calculated, according to age and birth year. We performed separate incidence analyses for $HER2^+$ and triple negative breast cancers.

Results: Among women aged 40-69 years, incidence rates of tumours with low-proliferative activity were higher among those born in 1929 or later, compared to before 1929, according to Ki-67 and mitotic count. Incidence rates of tumours with high-proliferative activity were also higher in women born in 1929 or later compared to before 1929 according to Ki-67, but not according to mitotic count. Mean values of Ki-67 and mitotic count varied according to age and birth year. In subtype-specific analyses we found an increase of high-proliferative $HER2^+$ tumours according to Ki-67 in women born in 1929 or later, compared to before 1929.

Conclusions: There has been a temporal increase in both low- and high-proliferative breast cancers.

Introduction

In Norway, breast cancer incidence rates have doubled since the establishment of the Cancer Registry of Norway 60 years ago (1). Breast cancer mortality rates remained stable from 1965 until 1995 but have shown a steady decline from 1995 until today. Nevertheless, breast cancer is the second most common cause of cancer related deaths among Norwegian women (1). Worldwide, breast cancer is the most common cancer and the leading cause of cancer related death among women (2). The discovery of the molecular subtypes in the 2000s (3) gave new insight into the heterogeneity of breast cancer. These subtypes are associated with different risk factors (4-6), incidence trends (7), prognosis (7, 8) and treatment response (9). The molecular subtypes of breast cancer are defined by gene expression patterns (3, 10). Molecular subtyping can also be done using immunohistochemistry (IHC) and *in situ* hybridization (ISH) as surrogates for gene expression analyses (11-13). Studies of incidence trends have demonstrated an increase in oestrogen receptor $(ER)^+$ breast cancers (7, 14-18), and a decrease in ER^- tumours (15-18). The increase in ER^+ tumours, and the decrease in ER^- tumours has been described for all ages (16-18). There has also been an increase in Luminal A (7, 15, 19) and Luminal B (Human epidermal growth factor receptor 2 $(HER2)^-$) breast cancers (7). The increase in ER^+ tumours has been attributed in part to the use of menopausal hormone therapy (MHT) and the introduction of mammography screening programs (20-22).

Proliferation is one of the hallmarks of cancer (23-25). High proliferation is associated with poor prognosis in breast cancer (26-28), and can be measured using mitotic count, Ki-67 protein expression and gene expression assays (11, 29-32). High mitotic counts and Ki-67 levels are associated with reduced overall survival and disease free survival (33). Ki-67 can be used to select early-stage breast cancer patients for chemotherapy and to monitor treatment response (34, 35). Recently, adjuvant therapy with abemaciclib in combination with endocrine therapy was approved by the Food and Drugs Administration (FDA) in the USA as

a treatment option in high-risk, hormone receptor positive, HER2 negative, node-positive, early breast cancers with Ki67 ≥20% (36). In addition, Ki-67 can be used to separate $ER^+/HER2^-$ tumours into Luminal A and Luminal B ($HER2^-$) (12, 35). Due to interobserver and interlaboratory variations and lack of consensus on cut-off levels, the clinical use of Ki-67 has been debated (32, 34, 35). According to Norwegian guidelines, mitotic count is also routinely reported as a marker of tumour cell proliferation (32).

Several studies have described time trends in incidence and prognosis of breast cancers overall and according to hormone receptor status and molecular subtype. However, to the best of our knowledge, time trends in breast cancer incidence according to proliferation status have not previously been described. Our aim was to study long-term trends in incidence of high- and low-proliferative breast cancers in a population of Norwegian women born between 1886 and 1977, using mitotic counts and Ki-67 as markers of proliferation.

Materials and Methods

This follow-up study comprises three large cohorts of Norwegian women (Supplementary figure 1). Information regarding incident breast cancers was obtained from the Cancer Registry of Norway using national identity numbers to link person data.

Cohort 1: Between 1956 and 1959 all women in the northern part of Trøndelag County, Norway, born between 1886 and 1928 were invited to attend a clinical screening for early detection of breast cancer (37). These women (n=25 727) were followed for breast cancer occurrence from January 1st, 1961, until the date of breast cancer diagnosis, death from other causes, emigration, or December 31st, 2008 (8). Among these women, 1379 incident breast cancer cases were registered. After diagnosis, patients were followed until death from breast cancer or from other causes, or until December 31st, 2010. Formalin-fixed paraffin-embedded

(FFPE) tissue was available for 945 of the diagnosed cases, and 909 of these have previously been reclassified into molecular subtypes (8).

Cohort 2: Between 1995 and 1997, all women in the northern part of Trøndelag County, Norway, born between 1897 and 1977 were invited to participate in the HUNT2 study (38). A total of 34 221 women were followed for breast cancer occurrence from attendance until the date of breast cancer diagnosis, death from other causes, emigration, or December 31st, 2009 (7). Among these, 731 incident breast cancer cases were registered. After diagnosis, patients were followed until death from breast cancer or other causes, or until December 31st, 2015. Of the 731 cases, 653 have previously been reclassified into molecular subtypes (7, 8).

Cohort 3: All women born at E. C. Dahl's foundation in Trondheim (in the southern part of Trøndelag) between 1920 and 1966 were followed for breast cancer occurrence from January 1st, 1961, until the date of breast cancer diagnosis, death from other causes, emigration, or December 31st, 2012 (39). Of the 23 350 women included, 885 incident breast cancer cases were registered (40). Participants were followed until death from breast cancer or any other causes, or until December 31st, 2015. Of the 885 cases, 545 have previously been reclassified into molecular subtypes (40).

In this study we pooled data from the three cohorts. In the analysis of incidence rates, we used data from a total of 2995 breast cancers that occurred among 83 298 women. Participants in the three cohorts were included from parts of the county of Trøndelag and across an overlapping range of birthyear (Supplementary figure 1). Thus, some women were included in more than one cohort. By using a case specific identity number, 171 incident breast cancer cases overlapping between two or all three cohorts were identified (5.7%). Due to anonymization of study participants, identification of overlap in the healthy background

population was not possible. Therefore, to avoid underestimation of incidence rates we did not exclude duplicate incident cancers in the analyses.

Specimen characteristics

Tumours were previously classified into histopathological type and grade and reclassified into molecular subtypes by the Breast Cancer Subtypes research group (7, 8, 40). Chromogenic *in situ* hybridization (CISH), fluorescence *in situ* hybridization (FISH) and IHC on tissue micro arrays (TMAs) were used to classify tumours into molecular subtypes according to the algorithm presented in Table 1 (8). For TMA construction, three 1-mm-in-diameter tissue cores were taken from the tumour periphery. According to current guidelines, Ki-67 was counted in 500 epithelial tumour cells in hot spot areas and reported as the proportion of nuclei with positive IHC staining (32). Scoring and reporting of the other molecular markers used for molecular subtyping have previously been described in detail (7, 8, 40).

Table 1 – Reclassification of breast cancers into molecular subtype

Molecular subtype	**Molecular marker**
Luminal A	ER^+ and/or PR^+, $HER2^-$, Ki67<15%
Luminal B ($HER2^-$)	ER^+ and/or PR^+, $HER2^-$, Ki67≥15%
Luminal B ($HER2^+$)	ER^+ and/or PR^+, $HER2^+$
HER2 type	ER^-, PR^-, $HER2^+$
Basal phenotype	ER^-, PR^-, $HER2^-$, $CK5^+$ and/or $EGFR^+$
5 negative phenotype	ER^-, PR^-, $HER2^-$, $CK5^-$, $EGFR^-$

Abbreviations: *ER* oestrogen receptor; *PR* progesterone receptor; *HER2* human epidermal growth factor receptor 2; *CK5* cytokeratin 5; *EGFR* epidermal growth factor receptor.

Ki-67 was assessed by two independent observers, and mean Ki-67 values were calculated. In cases with only one Ki-67 assessment, this was used. Tumours were then subdivided into categories according to Ki-67 status. We used two different cut-off levels for Ki-67: a) </≥15%, and b) </≥30% positive cells.

Mitoses were counted manually by two independent pathologists in ten high power fields in whole sections of breast cancer (7, 8, 40). According to current Norwegian guidelines the number of mitoses was then recalculated to number of mitoses/mm^2 (32, 41) and mean values were calculated. In cases with only one observation, this was used. According to WHO guidelines for histological grading mitotic counts are assigned a score from 1 to 3 based on thresholds for mitoses/mm^2 (32, 41). Based on these thresholds, we used two different cut-off levels for mitotic count: a) ≤/>3.6 mitoses/mm^2 (mitotic score 1 versus mitotic score 2 and 3), and b) </≥7.7 mitoses/mm^2 (mitotic score 1 and 2 versus mitotic score 3).

Statistical analyses

In the incidence analyses, we used the same cut-off for birth cohort as in a previous study by our group (7), and separated the study population into two groups: women born before 1929 and women born in 1929 or later. Incidence rates for all cancers combined and for each category of Ki-67 and mitotic count were calculated and plotted according to birth year and age at diagnosis. Poisson regression was used to compare incidence rates between women born before 1929 and women born in 1929 or later.

To examine Ki-67 and mitotic count as continuous variables, mean values with standard deviation (SD) were calculated according to age groups and birth year.

To examine whether a change in proliferation status over time occurred within specific molecular subtypes, we also analysed incidence rates according to birth cohort, Ki-67 status, and mitotic counts in $HER2^+$ and Triple negative (TN) tumours. Due to limited statistical power, these analyses were restricted to women aged 50 to 80 years. The TN tumours comprised all basal phenotype (BP) and 5-negative phenotype (5NP) tumours, and the $HER2^+$ tumours comprised the HER2 type and luminal B ($HER2^+$) subtypes. Analyses were not performed for luminal A and luminal B ($HER2^-$) tumours as they were already defined by Ki-67 status.

For some breast cancer cases, Ki-67 and/or mitotic count was unavailable. Ki-67 status was missing in 920 (31%) of the tumours, and mitotic count was missing in 886 (30%) of the tumours. Supplementary Table 1 gives an overview of tumour characteristics of cases with and without missing Ki-67 values. To compensate for missing values, multiple imputations (42, 43) were used to predict mitotic count and Ki-67 status for these cases. The imputation model included all information available: age (5-year categories) and calendar year at diagnosis (continuous), stage (I-IV, unknown), extent of disease (localized to the breast, local invasion, regional lymph nodes, distant lymph nodes or organ metastases, metastases detected, unknown) as reported by the Cancer Registry of Norway, year of birth (5-year categories), follow-up time after diagnosis (log transformed, continuous) and survival status (alive, death from breast cancer, death from other causes), with the assumption that data were missing at random (43). Incidence rates and incidence rate ratios (IRR) with 95% confidence intervals (CI) according to birth year and age were calculated based on 50 imputed datasets. STATA version 17 (STATA Corp.) was used for statistical analyses.

Results

Characteristics of study population

Mean age at baseline was 51.0 for women in cohort 1, 50.1 years for women in cohort 2 and 21.2 for women in cohort 3 (Table 2). Mean follow-up for breast cancer occurrence in the three cohorts was 29.7, 12.5 and 37.4 years, respectively. Mean age at diagnosis was 70.7 in cohort 1, 63.2 in cohort 2 and 54 years in cohort 3. Of the cases with known Ki-67 status, 538 (60%), 391 (60%) and 268 (50%) cases had Ki-67 <15%, in cohorts 1 to 3, respectively. Furthermore, 120 (14%), 108 (17%) and 123 (23%) cases had Ki-67 ≥30% in the three cohorts, respectively. Of the cases with known mitotic count, 539 (61%) in cohort 1, 428 (65%) in cohort 2 and 334 (61%) in cohort 3 had mitotic counts ≤3.6. Furthermore, 147 (16%), 97 (15%) and 92 (17%) cases had mitotic counts ≥7.7, respectively.

Table 2: Characteristics of the study populations used in estimation of breast cancer incidence

	Cohort 1	**Cohort 2**	**Cohort 3**
Women followed for breast cancer occurrence	**Women born 1886-1928**	**Women born 1897-1977**	**Women born 1920-1966**
Number of women	25 727	34 221	23 350
Mean age at baseline [a] (SD)	51.0 (11.6)	50.1 (17.5)	21.2 (3.7)
Mean follow up for BC occurrence [b] (SD)	29.7 (13.9)	12.5 (2.7)	37.4 (9.1)
Women with incident breast cancer			
Number of cases	1379	731	885
Mean age at diagnosis (SD)	70.7 (11.8)	63.2 (13.7)	54 (10.3)
Mean follow up after diagnosis (SD)	9.1 (8.9)	9.5 (4.8)	10.8 (7.5)
Death from breast cancer (%)	612 (44)	123 (17)	173 (20)
Death from other causes (%)	688 (50)	152 (21)	81 (9)
Ki67/500 tumour cells (%)			
Cases with missing status	496 (36)	84 (11)	340 (38)
<15	538 (39)	391 (53)	268 (30)
≥15, <30	225 (16)	148 (20)	154 (17)
≥30	120 (9)	108 (15)	123 (14)
Mitotic count (mitoses/mm^2) (%) [c]			

Cases with missing status	466 (34)	77 (11)	343 (39)
≤3.6	539 (39)	428 (58)	334 (38)
>3.6, <7.7	227 (16)	129 (18)	116 (13)
≥7.7	147 (11)	97 (13)	92 (10)
Molecular subtypes (%)			
Luminal A	433 (31)	354 (48)	236 (27)
Luminal B (HER2-)	248 (18)	157 (21)	178 (20)
Luminal B (HER2+)	71 (5)	48 (7)	65 (7)
HER2 type	62 (4)	33 (5)	24 (3)
5NP	33 (2)	19 (3)	6 (1)
BP	62 (4)	42 (6)	36 (4)
Unknown	470 (34)	78 (11)	340 (38)
Stage (%) [c]			
I	671 (49)	388 (53)	278 (31)
II	483 (35)	288 (39)	220 (25)
III	93 (7)	30 (4)	26 (3)
IV	116 (8)	25 (3)	31 (4)
Unknown	16 (1)	0	330 (37)
Extent of disease (%) [c]			
Disease localized to the breast	501 (36)	372 (51)	422 (48)
Local invasion	42 (3)	12 (1)	3 (0)
Regional lymph nodes	363 (27)	219 (30)	303 (34)
Distant lymph node or organ metastases	99 (7)	23 (3)	31 (4)
Metastases detected, unknown location	2 (0)	0	0
Unknown	372 (27)	105 (14)	126 (14)

[a)] At time of entry
[b] For women who were included prior to 20 years of age, follow up for breast cancer diagnosis started at their 20th birthday.
[c] As recorded by the Cancer registry of Norway. Information is based on histopathological and/or clinical examination.
Abbreviations: *SD* standard deviation; *BC* breast cancer; *5NP* 5 negative phenotype; *BP* basal phenotype,

Age-specific incidence rates according to year of birth

For women aged 40 to 69 years, age-specific breast cancer incidence rates were higher in women born in 1929 or later compared to those born before 1929 (Table 3).

Rates based on observed and imputed proliferation status followed the same patterns, with imputed rates being higher than observed rates. Imputed rates showed that the incidence of tumours with Ki-67 <15% and Ki-67 ≥15% in the ages 40-69 years was higher among women born in 1929 or later, compared to women born before 1929 (Table 3, Figure 1A and B). In the age group 70-79 years, there was no clear difference in breast cancer incidence according to Ki-67 at </≥15% cut-off. Using Ki-67 </≥30% as cut-off, we found an increase in imputed incidence rates of tumours with Ki-67<30% and ≥30% among women aged 40-69 years, but there was no clear difference in the age group 70-79 years (Supplementary Table 2 and Supplementary Figure 2A and B).

For mitotic count, imputed values showed that the incidence rates of tumours with ≤3.6 mitoses/mm^2 and <7.7 mitoses/mm^2 were higher in the age groups 40-69 years among women born in 1929 or later compared to women born before 1929 (Table 3, Figure 1C, Figure 1E). There was no clear difference in incidence rates of tumours with >3.6 mitoses/mm^2 or ≥7.7 mitoses/mm^2 when comparing women born before 1929 to women born in 1929 or later (Table 3, Figure 1D, Figure 1F). Differences in incidence rates when comparing women born in 1929 or later to women born before 1929 according to both Ki-67 and mitotic count are given in Table 3.

Table 3: Incidence rates, incidence rate differences (IRD) and incidence rate ratios (IRR) of proliferation markers Ki67 and mitotic count according to age at diagnosis and year of birth

		Observed					Imputed[a]				
		Incidence rate (cases/100 000 person-years)					Incidence rate (cases/100 000 person-years)				
	Age	Women born before 1929	Women born in 1929 or later	IRD	IRR	(95% CI)	Women born before 1929	Women born in 1929 or later	IRD	IRR	(95% CI)
Total[b]	40-49	78.3	127.4	49.1	1.6	(1.3-2.1)					
	50-59	110.0	214.4	104.4	1.9	(1.6-2.3)					
	60-69	170.6	298.5	127.9	1.7	(1.5-2.0)					
	70-79	246.5	240.6	-5.9	1.0	(0.8-1.2)					
Ki67/500 tumour cells (%)											
<15	40-49	16.9	34.6	17.7	2.1	(1.2-3.4)	31.9	50.2	18.3	1.6	(1.0-2.5)
	50-59	21.8	90.1	68.3	4.1	(2.9-5.9)	49.8	124.2	74.4	2.5	(1.9-3.3)
	60-69	58.9	146.3	87.4	2.5	(2.0-3.1)	92.3	186.7	94.4	2.0	(1.6-2.5)
	70-79	112.5	121.7	9.2	1.1	(0.8-1.5)	151.2	147.2	-4	1.0	(0.7-1.3)
≥15	40-49	15.9	51.5	35.6	3.2	(1.9-5.4)	46.1	77.1	31.0	1.7	(1.2-2.4)
	50-59	31.0	62.3	31.3	2.0	(1.5-2.8)	60.1	90.2	30.1	1.5	(1.1-2.0)
	60-69	46.9	83.8	36.9	1.8	(1.4-2.4)	78.1	111.6	33.5	1.4	(1.1-1.8)
	70-79	67.2	73.6	6.4	1.1	(0.7-1.7)	94.6	93.0	-1.6	1.0	(0.7-1.4)
Mitotic count (mitoses/mm^2)											
≤3.6	40-49	17.8	45.4	27.6	2.5	(1.5-4.2)	31.6	66.5	34.9	2.1	(1.4-3.2)
	50-59	20.0	101.1	81.1	5.0	(3.5-7.3)	43.2	140.6	97.4	3.3	(3.4-4.4)
	60-69	60.9	163.4	102.5	2.7	(2.1-3.4)	90.5	211.9	121.4	2.3	(1.9-2.9)

	70-79	110.0	138.7	28.7	1.2	(0.9-1.6)	145.3	166.6	21.3	1.1	(0.9-1.5)
>3.6	40-49	14.9	40.3	25.4	2.7	(1.6-4.6)	46.4	60.9	14.5	1.3	(0.9-1.9)
	50-59	34.6	51.4	16.8	1.5	(1.1-2.0)	66.5	73.8	7.3	1.1	(0.8-1.5)
	60-69	48.8	65.9	17.1	1.3	(1.0-1.8)	80.0	86.4	6.4	1.1	(0.8-1.4)
	70-79	73.8	59.4	-14.4	0.8	(0.5-1.3)	101.1	73.7	-27.4	0.7	(0.5-1.1)
<7.7	40-49	26.8	66.4	39.6	2.5	(1.7-3.7)	53.6	98.6	45	1.8	(1.3-2.5)
	50-59	38.3	132.0	93.7	3.5	(2.6-4.5)	78.8	185.4	106.6	2.4	(1.9-2.9)
	60-69	90.4	198.4	108	2.2	(1.8-2.6)	136.6	258.3	121.7	1.9	(1.6-2.2)
	70-79	160.0	181.2	21.2	1.1	(0.9-1.5)	203.5	216.6	13.1	1.1	(0.8-1.3)
≥7.7	40-49	5.9	19.3	13.4	3.2	(1.4-7.5)	24.3	28.8	4.05	1.2	(0.7-2.1)
	50-59	16.4	20.5	4.1	1.2	(0.8-2.0)	30.9	28.9	-2	0.9	(0.6-1.5)
	60-69	19.3	30.8	11.5	1.6	(1.0-2.5)	33.8	40.0	6.2	1.2	(0.8-1.9)
	70-79	29.9	17.0	-12.9	0.6	(0.2-1.3)	43.0	23.7	-19.3	0.6	(0.3-1.2)

[a] Based on 50 imputed datasets using age (5-year categories) and calendar year of diagnosis (continuous), stage (I, II, III, IV, unknown) and extent of disease (disease localized to the breast, local invasion, regional lymph nodes, distant lymph nodes or organ metastases, unknown) as reported by the Cancer Registry of Norway, year of birth (5-year categories), observation time after diagnosis (log-transformed) and survival status (alive, death from breast cancer, death from other causes).
[b] Breast cancer incidence from the Cancer Registry of Norway, including cases with unknown Ki67-status and mitotic count
Abbreviations: *IRD* incidence rate difference, *IRR* incidence rate ratio, *CI* confidence interval,

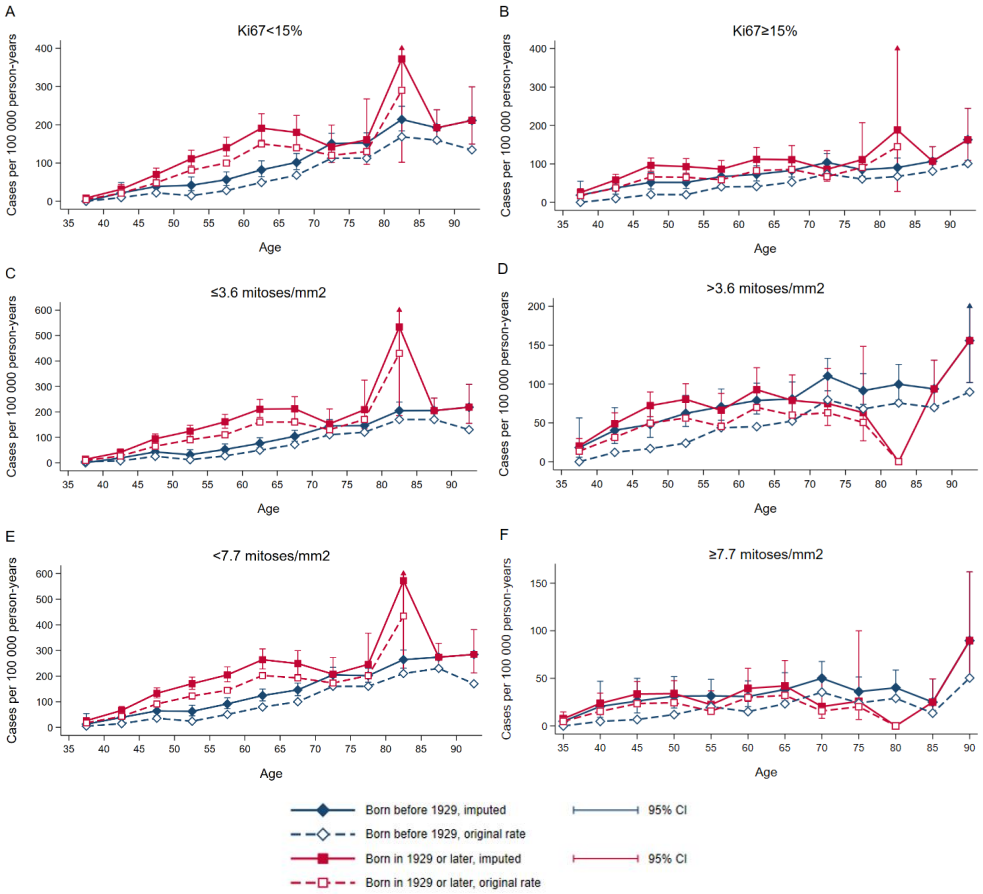


Figure 1: Incidence rates according to age, years of birth and proliferative marker status. Blue lines: women born before 1929. Red lines: Women born in 1929 or later. Dotted lines (red and blue) represent incidence rates of observed cases. Solid lines (red and blue) represent average incidence rates based on 50 imputed datasets with corresponding 95% CI. Figure A-F shows breast cancer incidence according to A) Ki-67 <15%, B) Ki-67 ≥15%, C) ≤3.6 mitoses/mm^2, D) >3.6 mitoses/mm^2, E) <7.7 mitoses/ mm^2 and F) ≥7.7 mitoses/ mm^2.

Mean values of proliferation markers according to age and birth year

We compared age-specific mean values for Ki-67 and mitotic count in women born before 1929 and women born in 1929 or later (Table 4). Differences in mean values according to birth cohort varied depending on age. Our analyses indicated that in the age group <49 years mean values of both Ki-67 and mitotic counts were higher among women born in 1929 or later compared to women born before 1929. In the age groups 50-64 and 70-74 years according to Ki-67 and in all aged >50 years according to mitotic count mean values were lower among women born in 1929 or later compared to those born before 1929.

Table 4: Mean Ki67 (%) and mean mitoses/mm^2 according to birth year and age at diagnosis

	Birth year in categories			
	Born before 1929		**Born in 1929 or later**	
Age	**Mean Ki67 (%) (SD)**	**Cases (n)**	**Mean Ki67 (%) (SD)**	**Cases (n)**
<45	14.9 (13.6)	8	27.5 (20.1)	133
45-49	16.8 (12.2)	25	22.6 (18.9)	166
50-54	24.7 (19.1)	26	19.2 (19.1)	180
55-59	23.2 (17.2)	60	15.9 (15.7)	156
60-64	19.0 (16.7)	91	16.1 (16.1)	163
65-69	16.2 (14.4)	129	16.5 (18.0)	105
70-74	16.4 (15.9)	215	14.9 (15.0)	47
75-79	13.9 (12.2)	191	16.2 (20.0)	22
80-84	13.7 (12.9)	206	15.7 (6.2)	3
Age	**Mean mitoses/mm^2 (SD)**	**Cases (n)**	**Mean mitoses/mm^2 (SD)**	**Cases (n)**
<45	6.3 (6.4)	8	6.7 (8.5)	132
45-49	4.1 (4.6)	25	5.5 (7.1)	166
50-54	7.5 (8.5)	27	4.7 (6.7)	180
55-59	6.2 (6.0)	62	3.1 (4.1)	156
60-64	5.3 (6.7)	95	3.7 (5.3)	163
65-69	4.6 (5.9)	133	4.1 (6.4)	105

70-74	5.0 (7.7)	222	3.4 (4.0)	47
75-79	3.7 (5.0)	202	2.7 (3.8)	22
80-84	3.0 (4.4)	213	1.5 (3.2)	3

Abbreviations: *SD* standard deviation.

Time trends in proliferative status within Triple Negative and HER2$^+$ tumours

We compared incidence rates of HER2$^+$ and TN tumours according to birth year, Ki-67-status, and mitotic count (Table 5). For HER2$^+$ tumours, we found that incidence rates of tumours with Ki-67≥15% increased from 11.4/100 000 person-years among women born before 1929 to 17.2/100 000 person-years among women born in 1929 or later (HR 1.6, 95% CI 1.1-2.4). The incidence rate of HER2$^+$ tumours with Ki-67≥30% was also higher among women born in 1929 or later, compared to women born before 1929 (HR 2.1 (95% CI 1.3-3.6)). For HER2$^+$ tumours with ≤3.6 mitoses/mm^2 incidence rates increased among women born in 1929 or later, compared to women born before 1929 (HR 1.9 (95% CI 1.1-3.2)). We found no clear changes in incidence rates among TN tumours.

Table 5: Incidence rates and hazard rates for subdivisions of molecular subtypes according to proliferation status and year of birth [a]

Molecular subtype	**Proliferation marker**	**Born before 1929**	**Born in 1929 or later**	**HR**	**95% CI**
	Ki67 (%)				
HER2$^+$	<15	7.0	7.5	1.1	(0.7-1.9)
	≥15	11.4	17.2	1.6	(1.1-2.4)
Triple negative	<15	3.7	2.7	1.1	(0.5-2.6)
	≥15	8.3	9.7	1.4	(0.9-2.3)
HER2$^+$	<30	13.2	14.8	1.2	(0.8-1.7)

	≥30	5.2	9.9	2.1	(1.3-3.6)
Triple negative	<30	6.1	4.3	1.1	(0.6-2.2)
	≥30	5.8	8.0	1.5	(0.9-2.5)
	Mitoses/mm^2				
HER2$^+$	≤3.6	6.0	9.7	1.9	(1.1-3.2)
	>3.6	13.3	15.1	1.2	(0.8-1.7)
Triple negative	≤3.6	3.6	2.4	1.0	(0.4-2.4)
	>3.6	9.3	10.0	1.3	(0.8-2.1)
HER2$^+$	<7.7	13.2	18.8	1.5	(1.0-2.1)
	≥7.7	6.2	5.9	1.1	(0.6-1.9)
Triple negative	<7.7	7.5	5.1	1.1	(0.6-1.9)
	≥7.7	5.5	7.3	1.5	(0.8-2.5)

[a] Analyses were limited to participants between 50 and 80 years.

[b] HER2$^+$ comprises Luminal B (HER2$^+$) and HER2-type tumours.

[c] Trippel Negative comprises basal phenotype and 5 negative phenotype tumours.

Abbreviations: *HR* hazard ratio; *CI* confidence interval.

Discussion

In this large population-based study of Norwegian women born between 1886 and 1977, age-specific breast cancer incidence rates were higher among women born in 1929 or later, compared to women born before 1929. Using Ki-67 as a marker of proliferation, we found an increase in incidence for both low- and high proliferative breast cancers. According to mitotic count, we found an increase in breast cancers with low-proliferative status. However, there was no increase in high-proliferative tumours according to mitotic count. Even though we found an increase in both high- and low-proliferative tumours, the increase was most prominent for low-proliferative tumours. We also did separate incidence analyses for HER2$^+$ and TN tumours. We found an increase in incidence of HER2$^+$ tumours with Ki-67 ≥15% and

Ki-67≥30%, while according to mitotic count, there was an increase in tumours with low mitotic count (≤3.6 mitoses/mm^2). There was no change in breast cancer incidence according to proliferation status among TN tumours.

Population based mammography screening was introduced in the county of Trøndelag in 2001 for all women between 50 and 69 years of age (44). Since the implementation of the mammography screening program in Norway the attendance in the program has been high (44). Of cancers detected among women attending mammography screening programs, approximately 25% are interval cancers (44). Studies have shown that mammography screening favours the detection of HER2 negative, luminal tumours (45, 46), particularly luminal A tumours (45). In addition, screening detected cancers are characterized by favourable traits, such as smaller size, and they are more often node negative compared to cancers detected outside of screening (45). Similar to our previous study of time trends in breast cancer incidence (7), we compared women born in 1929 or later to women born before 1929 in the present study. Women born before 1929 were not included in the mammography screening program, since they were older than 69 years of age at the time of implementation.

The use of Ki-67 as a predictive marker in breast cancer has been debated because of inter- and intra-observer variations, laboratory variations, and the lack of consensus regarding cut-off-values, scoring and reporting (34, 47, 48). Low reproducibility of Ki-67 results has, in particular, been demonstrated at levels >5% and <30% (34). Cheang *et al*. performed a study on breast cancer tumours in TMAs where they found that 13.25% was the optimal cut-off for Ki-67 IHC to separate luminal A from luminal B ($HER2^-$) tumours (12). This was subsequently incorporated into the St. Gallen guidelines of 2011, which defined 14% as the cut-off for Ki-67 on whole sections to separate luminal A from luminal B tumours (49), even though clinical Ki-67 assessment is usually performed on whole sections. A later study by Knutsvik *et al*. has shown that Ki-67 values differ depending on specimen type, and that

specimen-specific cut-off levels may be appropriate (50). We used 15% as Ki67 cut-off in our molecular subtyping. According to both national (32) and international guidelines (34), Ki-67 cut-offs should be determined based on local laboratory values. Since there is no international consensus on the appropriate Ki-67 cut-off levels, we used two different cut-offs in our incidence analyses: </≥15% and </≥30%. After the initiation of our study, the FDA approved the use of abemaciclib in combination with endocrine therapy as adjuvant therapy for patients with high risk, hormone receptor-positive, $HER2^-$, lymph node-positive and Ki-67≥20% early breast cancers (36), based on results from the monarchE study (51, 52). This study shows that in the era of multigene assays, Ki-67 still plays a role as a relevant prognostic marker in breast cancer.

In addition to Ki-67, we used mitotic count as a marker of proliferation. We used the established thresholds for mitotic counts included in histological grading (41).

The three cohorts included in this study comprise women from the northern and southern parts of Trøndelag county. There is overlap in birth period between the cohorts, and thus some women were included in more than one cohort. Breast cancer cases were identifiable through assigned case specific identity numbers, and we were able to identify the overlapping incident cases. Due to anonymization of study participants, we could not identify the corresponding overlap among the healthy background populations. Therefore, excluding the overlapping incident breast cancer cases would have led to an underestimation of breast cancer incidence rates. Overlap in the healthy background population could have been avoided by removing participants with overlapping birth year, however this would lead to exclusion of more than half of the study population. We assumed that the overlap among women who developed breast cancer was proportionally similar to the overlap among the other study participants, and therefore did not exclude duplicate incident cancers in the incidence analysis. It could nevertheless lead to some overestimation of precision.

Information about Ki-67 and mitotic count was unavailable for some tumours, mainly because these patients were diagnosed at other hospitals. To compensate for the missing data, we used multiple imputations to prevent underestimation of incidence rates. Our imputation model included all available data, such as year of birth, age and calendar year at diagnosis, stage and extent of disease, follow up time after diagnosis and survival status. Even though such clinical information from national registries was included in the imputation model to prevent biased results, it is difficult to assess how well the imputed rates reflects the true values. However, observed and imputed incidence rates followed the same age-specific patterns, and although weaker, we found that the differences in incidence rates persisted after imputation.

This large cohort of Norwegian women with long term follow-up for breast cancer occurrence gives a unique opportunity to examine time trends in breast cancer incidence. Tissue preparations and immunostaining were performed in the same laboratory using the same antibody for all cases. Mitotic count was evaluated independently by two pathologists, and Ki-67 was evaluated by two independent observers of whom at least one was a pathologist.

The study was performed on archival tissue from six decades; hence preanalytical conditions may have varied. Nevertheless, it is shown that archival tissue can give valuable information about time trends in molecular markers (53). Storage of archival tumour tissue in FFPE blocks may have limited impact on protein antigenicity (54), and according to recommendations published by the International Ki-67 in Breast Cancer Working Group in 2011 prolonged storage of FFPE tissue blocks in room temperature has little effect on Ki-67 (55). However, recent studies have found that antigenicity of Ki-67 stored in FFPE blocks decreases with tissue age (56, 57). In our study, we found high Ki-67 levels inn tumour tissue across all storage periods, and we found higher incidence of low-proliferative incident breast

cancers among women born in 1929 or later, compared to women born before 1929. Nevertheless, we cannot exclude a decrease in antigenicity as a result of prolonged storage in the oldest tumours. Reduced Ki-67 antigenicity could therefore have led to an underestimation of the increase in incidence of tumours with low proliferative activity in women born in 1929 or later, and a false relative increase in tumours with high proliferative activity in women born in 1929 or later. Tissue storage does not influence the number of mitoses, and according to mitotic count we only found an increase in tumours with low proliferative activity. Thus, the increase in high proliferative tumours according to Ki-67 may in part be explained by loss of Ki-67 antigenicity over time.

The molecular subtypes are associated with different levels of proliferative activity (11). Hence, incidence trends of molecular subtypes could reflect changes in breast cancer proliferation levels. Studies on breast cancer incidence trends according to hormone receptor status through the last decades in the US and several European countries have found increasing incidence rates of ER^+ tumours and decreasing rates of ER^- cancers (14-18). Hormone receptor positive tumours are in general less proliferative than hormone receptor negative tumours (11, 31), and can be further subdivided into molecular subtypes (Luminal A and Luminal B) based on their proliferative status (10, 35). Furthermore, we have previously demonstrated increasing incidence rates for the Luminal subtypes, especially Luminal A (7). Hence, our findings of increased incidence of tumours with low proliferative activity are in accordance with previously described incidence trends. Additionally, we also identified an increase in incidence for breast cancers with Ki-67 ≥15% and Ki-67≥30%. The increase in tumours with higher proliferative activity was however not identified according to mitotic count. To examine a possible change in proliferation within the molecular subtypes we made separate incidence analyses among $HER2^+$ and TN tumours. For $HER2^+$ tumours we found an increase of tumours with high proliferative status according to Ki-67 status, while

according to mitotic count there was an increase of tumours with low proliferative status. We found no difference in incidence of TN tumours according to proliferation status. The observed increase in tumours with high-proliferative activity according to Ki-67 while not according to mitotic count could partly be explained by reduced Ki-67 antigenicity in the oldest set of tumours due to storage. Furthermore, Ki-67 is expressed during all phases of the cell cycle except G0, while mitotic figures only occur during the M-phase (58, 59). Discrepant levels of Ki-67 and mitotic count have been described (60), and could explain the difference in incidence rates according to proliferative markers. Additionally, in our study Ki-67 was assessed on TMAs while mitotic count was assessed on whole sections. Even though good correlation between TMAs and whole sections has been demonstrated (61), intratumoural heterogeneity may be a challenge.

Our study of breast cancer incidence trends shows that there has been a change in proliferation status over time, with decreased mean values of mitoses and Ki-67 for most age groups. An increase in low-proliferative breast cancers over time may partly be explained by the introduction of mammography screening programs and MHT (46, 62, 63). Women born before 1929 were not included in the mammography screening program, and few would have received MHT. In Norway, MHT use decreased after 2002 (21), after the Womens Health Initiative study demonstrated an association with increased cardiovascular risk (64). However, increasing incidence over time for breast cancers with low-proliferative activity among women aged 40-49 years cannot be explained by mammography screening or MHT. Different risk patterns have been described for the molecular subtypes (4-6). Reproductive factors such as low parity, early menarche and late menopause are mainly associated with increased risk of Luminal A tumours (4, 62). Risk factors associated with the other molecular subtypes are not yet fully understood, and studies have shown inconsistent results (6, 62). The observed changes in breast cancer incidence according to proliferation status in our study

are most likely multifactorial and may be affected by both reproductive factors and lifestyle factors, such as obesity and alcohol consumption (4, 6, 65).

In conclusion, there has been an increase in incidence rates of tumours with low-proliferative activity, measured by Ki-67 and by mitotic count. We also found an increase in tumours with high-proliferative activity according to Ki-67, but not according to mitotic count.

List of abbreviations

BP: Basal phenotype; CI: Confidence interval; CISH: Chromogenic *in situ* hybridization; ER: Oestrogen receptor; FDA: Food and Drugs Administration; FFPE: Formalin-fixed paraffin-embedded; FISH: Fluorescence *in situ* hybridization; HER2: Human epidermal growth factor receptor 2; HR: Hazard ratio; IHC: Immunohistochemistry; IRD: Incidence rate difference; IRR: Incidence rate ratio; ISH: *In situ* hybridization; MHT: Menopausal hormone therapy; SD: Standard deviation; TMA: Tissue micro array; TN: Triple negative

Declarations

Ethics approval and consent to participate

This retrospective study of archival tissue samples was approved by the Regional Committee for Medical and Health Research Ethics Central Norway (REK 836/2009) (Regional Committees for Medical and Health Research Ethics | Forskningsetikk). In addition, the Regional Ethics Committee waived the requirement of written informed consent in compliance with the Norwegian Health Research Act (Lov om medisinsk og helsefaglig forskning (helseforskningsloven) - Lovdata)). Permission for linkage between data from the

Cancer Registry of Norway and the Cause of Death Registry has been granted by the Norwegian Data Inspectorate. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are not openly available due to sensitive personal data. However, anonymized data may be available from Professor Anna Mary Bofin (anna.bofin@ntnu.no) upon request and with permission of the regional ethics committee.

Competing interests

The authors declare that they have no competing interests.

Funding

This study received financial support from the Norwegian University of Science and Technology, Trondheim, Norway, and The Joint Research Committee between St. Olav's hospital and the Faculty of Medicine and Health Sciences, NTNU (FFU).

Authors contributions

AMB and MV contributed to the assessment of molecular markers and mitotic counts performed prior to this study. MV and SO conceptualized the study. EK performed the statistical analyses, and SO and MV contributed to the interpretation of the results. EK drafted the manuscript. MV, SO, AMB and SXR edited and reviewed the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

The authors thank the Department of Pathology at St. Olav's Hospital, Trondheim University Hospital, Norway, for making the archives available for the study, and the Cancer Registry of Norway, and the Norwegian Cause of Death Registry for providing the patient data.

References

1. Norway CRo. Cancer in Norway 2019 - Cancer incidence, mortality, survival and prevalence in Norway. 2020.
2. World Cancer Report: Cancer Research for Cancer Prevention. Wild CP WE, Stewart BW, editor. Lyon, France: International Agency for Research on Cancer; 2020.
3. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.
4. Tamimi RM, Colditz GA, Hazra A, Baer HJ, Hankinson SE, Rosner B, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. Breast cancer research and treatment. 2012;131(1):159-67.
5. Turkoz FP, Solak M, Petekkaya I, Keskin O, Kertmen N, Sarici F, et al. Association between common risk factors and molecular subtypes in breast cancer patients. The Breast. 2013;22(3):344-50.
6. Anderson KN, Schwab RB, Martinez ME. Reproductive risk factors and breast cancer subtypes: a review of the literature. Breast cancer research and treatment. 2014;144(1):1-10.
7. Valla M, Vatten LJ, Engstrom MJ, Haugen OA, Akslen LA, Bjorngaard JH, et al. Molecular Subtypes of Breast Cancer: Long-term Incidence Trends and Prognostic Differences. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2016;25(12):1625-34.
8. Engstrøm MJ, Opdahl S, Hagen AI, Romundstad PR, Akslen LA, Haugen OA, et al. Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. Breast cancer research and treatment. 2013;140(3):463-73.
9. Waks AG, Winer EP. Breast Cancer Treatment: A Review. Jama. 2019;321(3):288-300.
10. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(19):10869-74.
11. 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.
12. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009;101(10):736-50.
13. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res. 2004;10(16):5367-74.

14. Aromaa-Häyhä A, Auvinen P, Sarkeala T, Heinävaara S, Lamminmäki A, Malila N, et al. Improved diagnostics and change of tumour characteristics in breast cancer: a retrospective study over two decades. Acta Oncol. 2018;57(10):1331-8.
15. Mesa-Eguiagaray I, Wild SH, Rosenberg PS, Bird SM, Brewster DH, Hall PS, et al. Distinct temporal trends in breast cancer incidence from 1997 to 2016 by molecular subtypes: a population-based study of Scottish cancer registry data. British Journal of Cancer. 2020;123(5):852-9.
16. Mullooly M, Murphy J, Gierach GL, Walsh PM, Deady S, Barron TI, et al. Divergent oestrogen receptor-specific breast cancer trends in Ireland (2004–2013): Amassing data from independent Western populations provide etiologic clues. European Journal of Cancer. 2017;86:326-33.
17. Anderson WF, Katki HA, Rosenberg PS. Incidence of breast cancer in the United States: current and future trends. Journal of the National Cancer Institute. 2011;103(18):1397-402.
18. Anderson WF, Rosenberg PS, Petito L, Katki HA, Ejlertsen B, Ewertz M, et al. Divergent estrogen receptor-positive and-negative breast cancer trends and etiologic heterogeneity in Denmark. International journal of cancer. 2013;133(9):2201-6.
19. Acheampong T, Kehm RD, Terry MB, Argov EL, Tehranifar P. Incidence Trends of Breast Cancer Molecular Subtypes by Age and Race/Ethnicity in the US From 2010 to 2016. JAMA Network Open. 2020;3(8):e2013226-e.
20. Glass AG, Lacey JV, Jr., Carreon JD, Hoover RN. Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. J Natl Cancer Inst. 2007;99(15):1152-61.
21. Hofvind S, Sakshaug S, Ursin G, Graff-Iversen S. Breast cancer incidence trends in Norway--explained by hormone therapy or mammographic screening? Int J Cancer. 2012;130(12):2930-8.
22. Weedon-Fekjær H, Bakken K, Vatten LJ, Tretli S. Understanding recent trends in incidence of invasive breast cancer in Norway: age-period-cohort analysis based on registry data on mammography screening and hormone treatment use. BMJ (Clinical research ed). 2012;344:e299.
23. Hanahan D, Weinberg RAJc. The hallmarks of cancer. 2000;100(1):57-70.
24. Hanahan D, Weinberg RAJc. Hallmarks of cancer: the next generation. 2011;144(5):646-74.
25. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discovery. 2022;12(1):31-46.
26. van Diest PJ, van der Wall E, Baak JP. Prognostic value of proliferation in invasive breast cancer: a review. Journal of clinical pathology. 2004;57(7):675-81.
27. de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. British Journal of Cancer. 2007;96(10):1504-13.
28. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res. 2008;10(4):R65.
29. Burstein HJ, Curigliano G, Loibl S, Dubsky P, Gnant M, Poortmans P, et al. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. Annals of Oncology. 2019;30(10):1541-57.
30. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(8):1160-7.
31. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006;7(1):96.
32. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft. helsebiblioteket.no: Helsedirektoratet; 2022.
33. Stuart-Harris R, Caldas C, Pinder SE, Pharoah P. Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. Breast (Edinburgh, Scotland). 2008;17(4):323-34.
34. Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, et al. Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group. JNCI: Journal of the National Cancer Institute. 2021;113(7):808-19.
35. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, et al. Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Annals of oncology : official journal of the European Society for Medical Oncology. 2015;26(8):1533-46.
36. FDA approves abemaciclib with endocrine therapy for early breast cancer fda.gov: U.S Food & Drug Administration; 2021 [Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-abemaciclib-endocrine-therapy-early-breast-cancer.

37. Kvåle G, Heuch I, Eide GE. A prospective study of reproductive factors and breast cancer. I. Parity. Am J Epidemiol. 1987;126(5):831-41.
38. Holmen m.fl J. The Nord-Trøndelag Health Study 1995-97 (HUNT 2). Norsk Epidemiologi. 2011;13(1).
39. Sandvei MS, Lagiou P, Romundstad PR, Trichopoulos D, Vatten LJ. Size at birth and risk of breast cancer: update from a prospective population-based study. Eur J Epidemiol. 2015;30(6):485-92.
40. Sandvei MS, Opdahl S, Valla M, Lagiou P, Vesterfjell EV, Rise TV, et al. The association of women's birth size with risk of molecular breast cancer subtypes: a cohort study. BMC Cancer. 2021;21(1):299.
41. 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.
42. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ (Clinical research ed). 2009;338:b2393.
43. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59(10):1087-91.
44. Research-based evaluation of the Norwegian Cancer Screening Program Regjeringen.no: The Research Council of Norway; 2015 [Available from: https://www.regjeringen.no/contentassets/444d08daf15e48aca5321f2cefaac511/mammografirapport-til-web.pdf.
45. Crispo A, Barba M, D'Aiuto G, De Laurentiis M, Grimaldi M, Rinaldo M, et al. Molecular profiles of screen detected vs. symptomatic breast cancer and their impact on survival: results from a clinical series. BMC Cancer. 2013;13:15.
46. Sihto H, Lundin J, Lehtimäki T, Sarlomo-Rikala M, Bützow R, Holli K, et al. Molecular subtypes of breast cancers detected in mammography screening and outside of screening. Clin Cancer Res. 2008;14(13):4103-10.
47. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft, 2019. In: Helsedirektoratet, editor.: Helsedirektoratet; 2019.
48. Focke CM, van Diest PJ, Decker T. St Gallen 2015 subtyping of luminal breast cancers: impact of different Ki67-based proliferation assessment methods. Breast cancer research and treatment. 2016;159(2):257-63.
49. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Annals of oncology : official journal of the European Society for Medical Oncology. 2011;22(8):1736-47.
50. Knutsvik G, Stefansson IM, Aziz S, Arnes J, Eide J, Collett K, et al. Evaluation of Ki67 expression across distinct categories of breast cancer specimens: a population-based study of matched surgical specimens, core needle biopsies and tissue microarrays. PloS one. 2014;9(11):e112121.
51. Johnston SRD, Harbeck N, Hegg R, Toi M, Martin M, Shao ZM, et al. Abemaciclib Combined With Endocrine Therapy for the Adjuvant Treatment of HR+, HER2-, Node-Positive, High-Risk, Early Breast Cancer (monarchE). Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2020;38(34):3987-98.
52. Harbeck N, Rastogi P, Martin M, Tolaney SM, Shao ZM, Fasching PA, et al. Adjuvant abemaciclib combined with endocrine therapy for high-risk early breast cancer: updated efficacy and Ki-67 analysis from the monarchE study. Annals of oncology : official journal of the European Society for Medical Oncology. 2021;32(12):1571-81.
53. Dowsett T, Verghese E, Pollock S, Pollard J, Heads J, Hanby A, et al. The value of archival tissue blocks in understanding breast cancer biology. Journal of clinical pathology. 2014;67(3):272-5.
54. Engel KB, Moore HM. Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue. Arch Pathol Lab Med. 2011;135(5):537-43.
55. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in Breast Cancer: Recommendations from the International Ki67 in Breast Cancer Working Group. JNCI: Journal of the National Cancer Institute. 2011;103(22):1656-64.
56. Combs SE, Han G, Mani N, Beruti S, Nerenberg M, Rimm DL. Loss of antigenicity with tissue age in breast cancer. Laboratory Investigation. 2016;96(3):264-9.
57. Grillo F, Bruzzone M, Pigozzi S, Prosapio S, Migliora P, Fiocca R, et al. Immunohistochemistry on old archival paraffin blocks: is there an expiry date? Journal of clinical pathology. 2017;70(11):988.
58. Beresford MJ, Wilson GD, Makris A. Measuring proliferation in breast cancer: practicalities and applications. Breast Cancer Research. 2006;8(6):216.
59. Gerdes J, Lemke H, Baisch H, Wacker HH, Schwab U, Stein H. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. Journal of immunology (Baltimore, Md : 1950). 1984;133(4):1710-5.

60. Jalava P, Kuopio T, Juntti-Patinen L, Kotkansalo T, Kronqvist P, Collan Y. Ki67 immunohistochemistry: a valuable marker in prognostication but with a risk of misclassification: proliferation subgroups formed based on Ki67 immunoreactivity and standardized mitotic index. Histopathology. 2006;48(6):674-82.
61. Batistatou A, Televantou D, Bobos M, Eleftheraki AG, Kouvaras E, Chrisafi S, et al. Evaluation of current prognostic and predictive markers in breast cancer: a validation study of tissue microarrays. Anticancer research. 2013;33(5):2139-45.
62. Barnard ME, Boeke CE, Tamimi RM. Established breast cancer risk factors and risk of intrinsic tumor subtypes. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2015;1856(1):73-85.
63. Ellingjord-Dale M, Vos L, Tretli S, Hofvind S, Dos-Santos-Silva I, Ursin G. Parity, hormones and breast cancer subtypes - results from a large nested case-control study in a national screening program. Breast Cancer Res. 2017;19(1):10.
64. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. Jama. 2002;288(3):321-33.
65. Danaei G, Vander Hoorn S, Lopez AD, Murray CJL, Ezzati M. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. The Lancet. 2005;366(9499):1784-93.

Supplementary material

Supplementary table 1: Characteristics of breast cancer patients according to Ki-67-status (missing vs. non-missing)

	Born before 1929 (n=1641)		**Born in 1929 or later (n=1354)**	
	Non-missing	**Missing**	**Non-missing**	**Missing**
Number of women (%)	1100 (67)	541 (33)	975 (72)	379 (28)
Mean age at baseline (SD) [a]	50.5 (13.8)	52.1 (12.5)	35.1 (15.8)	25 (10.9)
Mean age at diagnosis (SD)	73.5 (10.6)	68.0 (13.1)	55.1 (9.8)	53 (9.6)
Mean year of diagnosis (SD)	1989 (11.7)	1979 (13.0)	2002 (6.5)	2002 (7.3)
Mean follow up for BC occurrence (SD)	23.4 (12.8)	16.5 (12.1)	20.5 (14.4)	28.5 (12.1)
Death from breast cancer (%)	427 (39)	271 (50)	145 (15)	65 (17)
Death from other causes (%)	570 (52)	246 (45)	82 (8)	23 (6)
Stage (%) [b]				
I	521 (47)	255 (47)	410 (42)	151 (40)
II	426 (39)	157 (29)	336 (35)	72 (19)
III	71 (6)	38 (7)	34 (3)	6 (2)
IV	54 (5)	73 (14)	23 (2)	22 (6)
Unknown	28 (3)	18 (3)	172 (17)	128 (34)
Extent of disease (%) [b]				
Disease localized to the breast	377 (34)	217 (40)	499 (51)	202 (53)
Local invasion	37 (3)	14 (3)	6 (1)	0
Regional lymph nodes	292 (27)	148 (27)	344 (35)	101 (27)
Distant lymph node or organ metastases	43 (4)	68 (13)	20 (2)	22 (6)
Metastases detected, unknown location	0	2 (0)	0	0
Unknown	351 (32)	92 (17)	106 (11)	54 (14)

[a] At time of entry
[b] As recorded by the Cancer registry of Norway. Information is based on histopathological and/or clinical examination.
Abbreviations: *SD* standard deviation

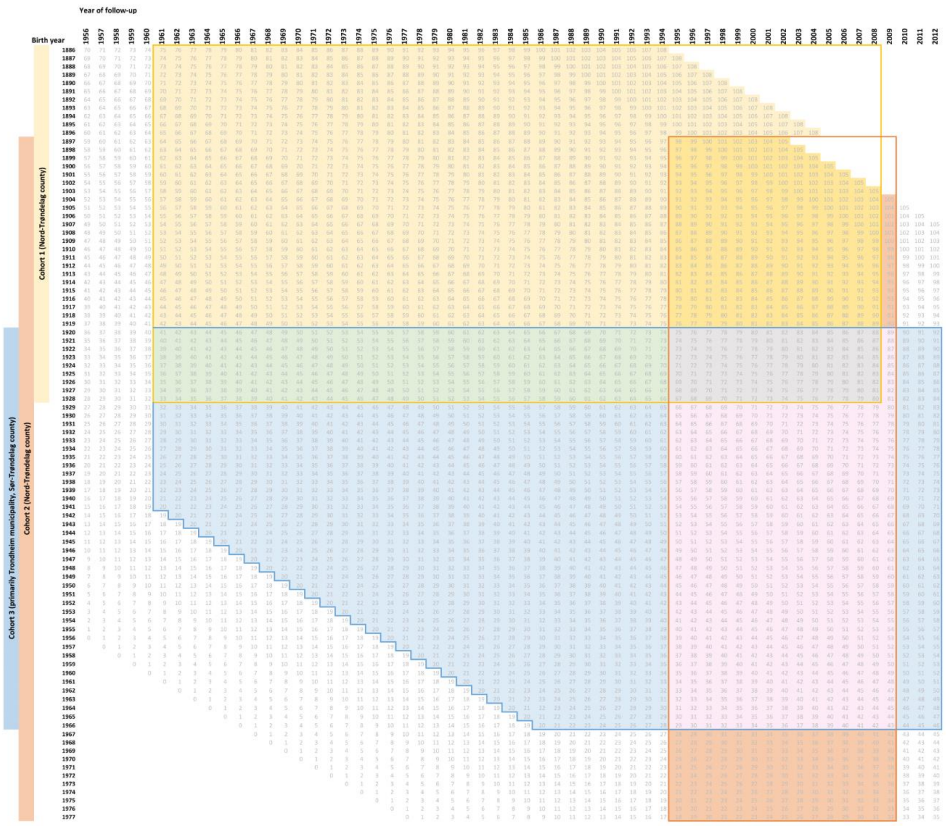
Supplementary table 2: Incidence rates, incidence rate differences (IRD) and incidence rate ratios (IRR) of proliferation marker Ki67

		Observed					Imputed[a]				
		Incidence rate (cases/100 000 person-years)					Incidence rate (cases/100 000 person-years)				
Ki67 (%)	Age	Women born before 1929	Women born in 1929 or later	IRD	IRR	(95% CI)	Women born before 1929	Women born in 1929 or later	IRD	IRR	(95% CI)
<30	40-49	27.8	60.7	32.90	2.2	(1.5-3.3)	58.4	89.7	31.30	1.5	(1.1-2.1)
	50-59	38.9	124.3	85.40	3.2	(2.4-4.1)	83.5	173.4	89.90	2.1	(1.7-2.6)
	60-69	88.4	193.3	104.90	2.2	(1.8-2.6)	140.8	248.9	108.10	1.8	(1.5-2.1)
	70-79	159.1	167.0	7.90	1.0	(0.8-1.4)	215.1	202.8	-12.30	0.9	(0.7-1.2)
≥30	40-49	5.0	25.4	20.40	5.1	(2.1-12.7)	19.1	37.6	18.50	2.0	(1.0-3.9)
	50-59	14.0	28.2	14.20	2.0	(1.3-3.3)	26.3	40.8	14.50	1.6	(1.0-2.4)
	60-69	17.4	36.8	19.40	2.1	(1.4-3.3)	29.6	49.5	19.90	1.7	(1.1-2.5)
	70-79	20.7	28.3	7.60	1.4	(0.7-2.7)	31.3	37.5	6.20	1.2	(0.6-2.3)

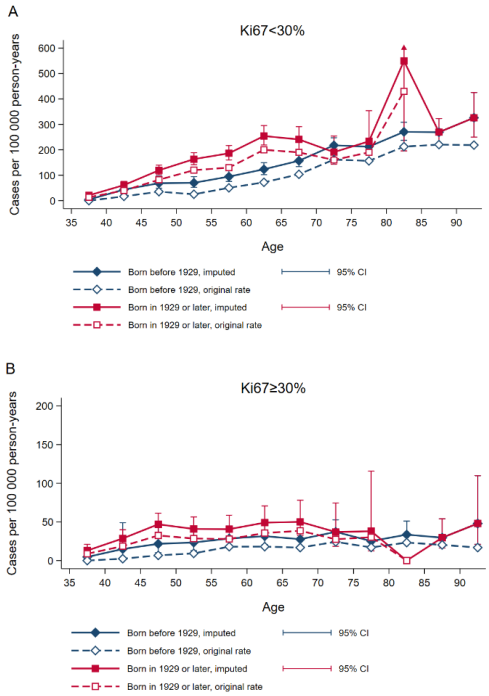
[a] Based on 50 imputed datasets using age (5-year categories) and calendar year of diagnosis (continous), stage (I, II, III, IV, unknown) and extent of disease (disease localized to the breast, local invasion, regional lymph nodes, distant lymph nodes or organ metastases, unknown) as reported by the Cancer Registry of Norway, year of birth (5-year categories), observation time after diagnosis (log-transformed) and survival status (alive, death from breast cancer, death from other causes).

[b] Breast cancer incidence from the Cancer Registry of Norway, including cases with unknown Ki67-status and mitotic count

Abbreviations: *IRR* incidence rate ratios, *CI* confidence interval,



Supplementary figure 1: Graphical display of birth year, age distribution and follow up period for the three cohorts. Cohort 1 is marked in yellow, cohort 2 in orange and cohort 3 in blue. Age according to birth year and year of follow up is listed in the background.



Supplementary figure 2: Incidence rates according to age, years of birth and proliferative marker status. Blue lines: women born before 1929. Red lines: Women born in 1929 or later. Dotted lines (red and blue) represent incidence rates of observed cases. Solid lines (red and blue) represent average incidence rates based on 50 imputed datasets with corresponding 95% CI.

ISBN 978-82-326-6026-1 (printed ver.)
ISBN 978-82-326-6155-8 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology