Precision Medicine in Breast Cancer: A Molecular Genomics and Diagnostics Approach

Dalal, Hina

2024

Document Version:
Publisher's PDF, also known as Version of record

Precision Medicine in Breast Cancer: A Molecular Genomics and Diagnostics Approach

HEENA DALAL
DEPARTMENT OF CLINICAL SCIENCES, LUND | FACULTY OF MEDICINE | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

Breast cancer remains one of the leading health challenges worldwide. This thesis presents innovative research on breast cancer within SCAN-B initiative, exploring novel biomarkers (ERβ; ESR2 and CITED1) and sophisticated digital PCR assays for precise HER2/ERBB2 DNA copy number quantification and HER2-low expression. Each study offers new insights into the heterogeneity of breast cancer and contributes to refining clinical approaches and enhancing therapeutic strategies.

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences, Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2024:90
ISBN 978-91-8021-585-5
ISSN 1652-8220

Precision Medicine in Breast Cancer: A Molecular Genomics and Diagnostics Approach

Heena Dalal

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.

To be defended at Segerfalksalen, BMC, Lund

Monday, June 17, 2024 at 14:00.

Faculty opponent

Professor Anita Grigoriadis, PhD

Professor of Molecular and Digital Pathology
Head of Comprehensive Cancer Centre,
School of Life Sciences and Medicine,
King's College London, UK

Organization: LUND UNIVERSITY
Faculty of Medicine
Department of Clinical Sciences, Lund
Division of Oncology
Document name: DOCTORAL DISSERTATION **Date of issue:** 2024-06-17

Author(s): Heena Dalal **Sponsoring organization**

Title and subtitle: Precision Medicine in Breast Cancer: A Molecular Genomics and Diagnostics Approach

Abstract:

Breast cancer is one of the most prevalent and diverse malignancies affecting women worldwide, with an estimated 2.3 million new cases globally in 2020 alone. In Sweden, breast cancer is a major societal issue with over 8600 new diagnoses and 1300 deaths per year. It presents significant challenges in diagnosis, treatment, and prognostication due to its complex biological, molecular, and clinical heterogeneity. The advent of molecular markers and advanced genomic technologies has opened new avenues for understanding and managing this disease more effectively. This thesis compiles five studies that underscore the importance of these advancements in refining breast cancer classification and contribute to new biomarker assessment which may ultimately improve patient outcomes. Study I explores the implications of estrogen receptor β (ERβ; *ESR2*) mRNA expression in breast cancer through a comprehensive transcriptomic analysis of a large cohort, SCAN-B. The findings indicate that higher *ESR2* expression correlates with improved overall survival, especially in cases receiving endocrine therapy and in triple-negative breast cancer, suggesting *ESR2*'s potential as a valuable prognostic marker and its potential role in immune modulation. Study II introduces a novel multiplex droplet digital PCR (ddPCR) assay for the accurate determination of HER2/*ERBB2* DNA copy number, and demonstrated high accuracy compared to traditional clinical methods. This study revealed an "ultrahigh" *ERBB2* copy number subgroup, as quantified by ddPCR, found to be associated with worse survival outcomes in patients treated with trastuzumab, emphasizing the assay's potential utility in refining treatment decision-making and its implications for tailored therapeutic strategies. Study III examined CITED1 as a predictive marker for anti-endocrine treatment efficacy, particularly in the context of tamoxifen therapy. The association between higher *CITED1* expression and favorable treatment outcomes in ER+ patients position *CITED1* as a promising biomarker for tailoring endocrine therapy. Study IV details the first 10-years of achievements of the Sweden Cancerome Analysis Network - Breast (SCAN-B) project, a population-based initiative that integrates genomic profiling and RNA-sequencing into clinical practice. By analyzing a large and diverse group of patients, SCAN-B aims to enhance personalized breast cancer care through detailed molecular analysis, demonstrating the project's contribution to advancing the field of personalized medicine in breast cancer. Study V addresses the classification and treatment implications of the "HER2-low" breast cancer subclass. By developing a refined ddPCR assay for *ERBB2* mRNA expression, this study provides a more nuanced understanding of *ERBB2* expression levels, offering insights into the potential benefits of emerging targeted therapies such as trastuzumab deruxtecan for this distinct group.

In conclusion, the research projects presented in this thesis demonstrate the critical role of molecular markers and genomic technologies in advancing our understanding and management of breast cancer. By enabling a more precise approach to diagnosis, treatment, and prognosis, these studies contribute to the ongoing advancement towards more individualized and effective breast cancer care.

Key words: breast cancer, precision medicine, RNA-seq, digital PCR, SCAN-B

Classification system and/or index terms (if any) Supplementary bibliographical information

Language English **Number of pages**: 142

ISSN and key title: 1652-8220
Lund University, Faculty of Medicine Doctoral Dissertation Series 2024:90
ISBN: 978-91-8021-585-5

Recipient's notes Price Security classification

Signature Date 2024-05-13

Precision Medicine in Breast Cancer: A Molecular Genomics and Diagnostics Approach

Heena Dalal

LUND
UNIVERSITY

Coverphoto by: © Lilian Fellner, Anatomical Artist

Faculty of Medicine
Department of Clinical Sciences, Lund

ISBN 978-91-8021-585-5
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2024

To my mentor

Lao H. Saal

कर्मण्येवाधिकारस्ते मा फलेषु कदाचन ।

मा कर्मफलहेतुर्भुर्मा ते संगोऽस्त्वकर्मणि ॥

-- Bhagavad Gita

Pursue your Karma with such passion that the outcome becomes immaterial!

Table of Contents

List of Original Studies

This thesis is based on the following five studies:

I. Clinical associations of *ESR2* (estrogen receptor beta) expression across thousands of primary breast tumors.

Hina Dalal, Malin Dahlgren, Sergii Gladchuk, Christian Brueffer, Sofia K. Gruvberger-Saal, and Lao H. Saal

Scientific Reports. 2022 Mar 18;12(1):4696.

II. Digital PCR quantification of ultrahigh *ERBB2* copy number identifies poor breast cancer survival after trastuzumab.

Pei Meng*, **Hina Dalal***, Yilun Chen, Christian Brueffer, Sergii Gladchuk, Miguel Alcaide, Anna Ehinger, and Lao H. Saal

NPJ Breast Cancer. 2024 Feb 19;10(1):14.

III. CITED1 as a marker of favourable outcome in anti-endocrine treated, estrogen-receptor positive, lymph-node negative breast cancer.

Malin Dahlgren*, Barbara Lettiero*, **Hina Dalal**, Kira Mårtensson, Alexander Gaber, Björn Nodin, Sofia K. Gruvberger-Saal, Lao H. Saal, and Jillian Howlin

BMC Research Notes. 2023 Jun 15;16(1):105.

IV. The Sweden Cancerome Analysis Network – Breast Initiative: 10-years follow-up of the multicenter multiomic SCAN-B initiative to improve personalization of breast cancer care.

Lao H. Saal, **Hina Dalal**, Pei Meng, Christian Brueffer, Sergii Gladchuk, Sofia K. Gruvberger-Saal, Anders Edsjö, Martin Malmberg, Christer Larsson, Lisa Rydén, Anna Ehinger, Niklas Loman, Cecilia Hegardt, Åke Borg, and Johan Vallon-Christersson

Manuscript

* Authors contributed equally.

V. Characterization of HER2-low and -ultralow using digital PCR for *ERBB2* expression profiling of breast cancer.

Pei Meng, **Hina Dalal**, Anna-Sophia Müller, Yilun Chen, Christian Brueffer, Fredrika Killander, Niklas Loman, Ana Bosch, Anna Ehinger, and Lao H. Saal

Manuscript

Author's contribution to the papers

I. Clinical associations of *ESR2* (estrogen receptor beta) expression across thousands of primary breast tumors.

As a first author, I contributed to the study's design and performed the analysis of SCAN-B and TCGA datasets, focusing on survival, differential gene expression, and gene set enrichment analyses. I assisted in drafting the manuscript, revising it in response to peer review, and addressing reviewers' comments. Additionally, I participated in the dissemination of our findings through poster presentations at three scientific events.

II. Digital PCR quantification of ultrahigh *ERBB2* copy number identifies poor breast cancer survival after trastuzumab.

As a co-first author, I contributed to the study's design and performed ddPCR wet lab experiments for approximately half of the patient samples included in this study. I analyzed the data including receiver operating curve and survival analyses of both the ddPCR and SCAN-B datasets. I was involved in drafting the manuscript and addressing reviewers' comments. I also presented the paper both as a selected talk and as a poster at an international conference.

III. CITED1 as a marker of favourable outcome in anti-endocrine treated, estrogen-receptor positive, lymph-node negative breast cancer.

As third author, I analyzed the TCGA and tissue microarray data, contributing significantly to the bioinformatics section of the manuscript. I also played an active role in revising the manuscript and addressing reviewers' comments regarding the data analysis.

IV. The Sweden Cancerome Analysis Network – Breast Initiative: 10-years follow-up of the multicenter multiomic SCAN-B initiative to improve personalization of breast cancer care.

As second author, I contributed to the structure of the study and analyzed the SCAN-B patient demographics and outcomes data and generated figures for biomarker distribution, treatments, and survival analyses. I assisted in writing and reviewing the manuscript.

V. Characterization of HER2-low and -ultralow using digital PCR for *ERBB2* expression profiling of breast cancer.

As second author, I contributed to the study's design and performed data analysis including survival analysis and RNA-seq validation analysis and helped generate some of the ddPCR data. I was also involved in writing and reviewing the manuscript.

Abstract

Breast cancer is one of the most prevalent and diverse malignancies affecting women worldwide, with an estimated 2.3 million new cases globally in 2020 alone. In Sweden, breast cancer is a major societal issue with over 8600 new diagnoses and 1300 deaths per year. It presents significant challenges in diagnosis, treatment, and prognostication due to its complex biological, molecular, and clinical heterogeneity. The advent of molecular markers and advanced genomic technologies has opened new avenues for understanding and managing this disease more effectively. This thesis compiles five studies that underscore the importance of these advancements in refining breast cancer classification and contribute to new biomarker assessment which may ultimately improve patient outcomes.

Study I explores the implications of estrogen receptor β (ERβ; *ESR2*) mRNA expression in breast cancer through a comprehensive transcriptomic analysis of a large cohort, SCAN-B. The findings indicate that higher *ESR2* expression correlates with improved overall survival, especially in cases receiving endocrine therapy and in triple-negative breast cancer, suggesting *ESR2*'s potential as a valuable prognostic marker and its potential role in immune modulation. **Study II** introduces a novel multiplex droplet digital PCR (ddPCR) assay for the accurate determination of HER2/*ERBB2* DNA copy number, and demonstrated high accuracy compared to traditional clinical methods. This study revealed an "ultrahigh" *ERBB2* copy number subgroup, as quantified by ddPCR, found to be associated with worse survival outcomes in patients treated with trastuzumab, emphasizing the assay's potential utility in refining treatment decision-making and its implications for tailored therapeutic strategies. **Study III** examined CITED1 as a predictive marker for anti-endocrine treatment efficacy, particularly in the context of tamoxifen therapy. The association between higher *CITED1* expression and favorable treatment outcomes in ER+ patients position *CITED1* as a promising biomarker for tailoring endocrine therapy. **Study IV** details the first 10-years of achievements of the Sweden Cancerome Analysis Network - Breast (SCAN-B) project, a population-based initiative that integrates genomic profiling and RNA-sequencing into clinical practice. By analyzing a large and diverse group of patients, SCAN-B aims to enhance personalized breast cancer care through detailed molecular analysis, demonstrating the project's contribution to advancing the field of personalized medicine in breast cancer. **Study V**

addresses the classification and treatment implications of the “HER2-low” breast cancer subclass. By developing a refined ddPCR assay for *ERBB2* mRNA expression, this study provides a more nuanced understanding of *ERBB2* expression levels, offering insights into the potential benefits of emerging targeted therapies such as trastuzumab deruxtecan for this distinct group.

In conclusion, the research projects presented in this thesis demonstrate the critical role of molecular markers and genomic technologies in advancing our understanding and management of breast cancer. By enabling a more precise approach to diagnosis, treatment, and prognosis, these studies contribute to the ongoing advancement towards more individualized and effective breast cancer care.

Popular Summary

Breast cancer is an enormous public health concern, having surpassed lung cancer as the most frequently diagnosed cancer worldwide. Breast cancer is not a single disease, but in fact there are several different types of breast cancer, each requiring different treatment approaches. Although the 5-year survival statistics for breast cancer is good, with generally more than 90% surviving, it is less commonly known that many breast cancer patients can have "late" relapses even after 5-years, leading to a 15-year survival rate of about 70%. In Sweden, it is the second most common cause of cancer death with over 1300 deaths per year. Although the number of breast cancer diagnoses has been increasing over the decades, the mortality rates show a general decrease, reflecting advancements in detection and treatment.

The concept of "precision medicine", particularly in the realm of molecular genomics and diagnostics, has improved the management of breast cancer. Recent research has been focusing on better understanding and stratifying different breast cancer subtypes and finding more precise ways to treat them. This thesis explores the application of precision medicine in breast cancer, focusing on the molecular characterization of tumors and the implications for diagnostics and treatment strategies using large scale analyses of breast cancer tissues and advanced sequencing methods and digital polymerase chain reaction (ddPCR) methods. Specifically, the five studies presented in this thesis explore the role of three cancer biomarkers, the 'estrogen receptor beta' (ERβ; encoded by the gene *ESR2*), HER2 (encoded by *ERBB2*), and CITED1 (encoded by *CITED1*) across very large subsets of breast tumor samples that have been collected within the Sweden Cancerome Analysis Network - Breast (SCAN-B) study.

In **Study I**, we analyzed over 3,000 breast tumors from SCAN-B and found that, although ERβ/*ESR2* is generally present at low levels, higher levels of this receptor are associated with better survival rates, especially in patients receiving specific hormone therapies or those with an aggressive form of breast cancer called "triple-negative breast cancer" (TNBC). Gene expression signatures indicative of the immune response were linked to *ESR2*-high tumors, and higher expression of *ESR2* was associated with improved overall survival and may influence immune response modulation. These findings

suggest ERβ's potential role as a biomarker for more favorable outcomes in patients receiving hormone therapy and in TNBC.

In **Study II**, we investigated the important biomarker HER2 (*ERBB2*) which is a protein that promotes the growth of cancer cells. In about 20% of breast cancers, excessive copies of the *ERBB2* gene are present leading to excessive HER2 protein, which can then be treated with anti-HER2 drugs which are now part of standard medical practice for over 20 years. Therefore, assessing HER2/*ERBB2* status is crucial for guiding treatment decisions in breast cancer, but current gold standard testing methods used in clinics for HER2 evaluation have their drawbacks and can show variable results. To improve this, we developed a new cutting-edge droplet digital PCR assay that can more precisely quantify the copy number (CN) of the *ERBB2* gene. Our ddPCR assay showed high accuracy and reliability in predicting HER2 status in breast cancer tissues from 909 patients. Interestingly, a group of patients with an "ultrahigh" *ERBB2* gene copy number were also discovered, who showed significantly worse survival outcomes despite treatment with the HER2-targeting drug trastuzumab (Herceptin). This finding, which was also confirmed in a larger SCAN-B cohort, suggests that ultrahigh *ERBB2* copy number levels could predict poorer long-term survival, highlighting the need for alternative therapeutic strategies in these cases.

CITED1 is a gene involved in how breast cancer cells respond to estrogen. **Study III** investigated the role of CITED1 as a biomarker for response to anti-hormone therapy in breast cancer and indicated that higher levels of CITED1 are linked with a better response to tamoxifen, a standard anti-estrogen treatment. This correlation was particularly strong in patients with estrogen receptor-positive, lymph node-negative breast cancer. Although early, this study suggests CITED1 as a potential predictive biomarker to identify which patients will benefit most from tamoxifen, and provides a potential mechanism that helps to mediate breast cancer response to hormone therapy.

The SCAN-B project is a large-scale population-based precision medicine initiative, initially started in Sweden in 2010, that aims to enroll all patients with primary (non-metastatic) breast cancer at all of the participating hospitals to build an infrastructure for breast cancer research and for dissemination of new diagnostic tools developed within SCAN-B. This effort has already led to the implementation of new molecular tests in healthcare in Skåne region of southern Sweden. In **Study IV** we present a status update of the SCAN-B project, summarizing the progress and patient inclusion for the project's first 10 full calendar years from late 2010 through the end of 2020. During this 10-year period, almost 14 thousand patients were enrolled in SCAN-B and over 9300 breast tumors were RNA-sequenced, making it one of the largest, if not the largest, breast cancer RNA-sequencing study in the world.

We present up-to-date clinical subtyping, treatment, and outcome results. The advantage of such population-based study is the ability to capture a representative cohort of patients, reflect real-world clinical outcomes, and identify rare subtypes of breast cancer, thereby enhancing the generalizability and applicability of research findings to a broader patient population.

Lastly, in **Study V** we again investigated HER2, but instead of analyzing ERBB2 DNA copies, we measured ERBB2 mRNA levels. This is because a new classification of breast cancer is emerging, termed "HER2-low", given that metastatic breast tumors with a moderate amount of HER2 appear to respond to a new class of drugs related to trastuzumab/Herceptin with an additional toxic linker molecule attached. Approximately 50% of all breast cancer cases are HER2-low, however the conventional semi-quantitative clinical tests for HER2 levels are imperfect and may not accurately evaluate low HER2 expression, due to differences in antibodies, protocols, and interpretations among pathologists. To address these concerns, we developed another new multiplex ddPCR assay to measure ERBB2 mRNA expression levels, which we validated on control samples and then analyzed over 1200 breast tumor samples from SCAN-B. This study aims to redefine HER2 expression categorizations in breast cancer, offering a more precise, cost-effective, and easily implementable tool based on ddPCR, which could improve diagnostic accuracy and also enhance treatment specificity, tailoring interventions to the molecular profile of each tumor.

In summary, these studies highlight the continuous efforts to understand breast cancer's molecular underpinnings and how this knowledge could lead to more personalized and effective diagnostics and treatments. Ultimately, through these efforts and many others, the goal is to improve patient stratification, leading to better informed treatment decisions, and improve the survival rates and quality of life for breast cancer patients.

Abbreviations

aa	amino acid(s)
ABC	advanced breast cancer
ABiM	All Breast Cancer in Malmö
ADC	antibody-drug conjugate
ADCC	antibody-dependent cellular cytotoxicity
AF	activating function domain
AI	aromatase inhibitor
AJCC	American Joint Committee on Cancer
AREG	amphiregulin
ASCO	American Society of Clinical Oncology
ASIR	age-standardized incidence rate
BC	breast cancer
BCI	Breast Cancer Index
BCS	breast-conserving surgery
BER	base excision repair
bHLH	basic-helix-loop-helix
BMI	body mass index
bp	base pair(s)
CAP	College of American Pathologists
cDNA	complementary DNA
CEP17	chromosome enumeration probe 17
CI	confidence interval
CISH	chromogenic in situ hybridization
CNA	copy number alteration
CNV	copy number variant
CT	chemotherapy
ctDNA	circulating tumor DNA
DBD	DNA binding domain
DCIS	ductal carcinoma in situ
ddPCR	droplet digital PCR
DDR	DNA damage response
DGE	differential gene expression
DMFS	distant metastasis-free survival
dMMR	DNA mismatch repair
dNTP	deoxynucleoside triphosphate

dPCR	digital PCR
EFS	event-free survival
ER	estrogen receptor
ERE	estrogen response element
ET	endocrine therapy
FFPE	formalin-fixed paraffin-embedded
FISH	fluorescence in situ hybridization
GSEA	Gene Set Enrichment Analysis
GTF	gene transfer format
HoR	homologous recombination
HR	hazard ratio
HRD	homologous recombination deficiency
HRT	hormonal replacement therapy
IBC	invasive breast cancer
IDC	invasive ductal carcinoma
IHC	immunohistochemistry
ILC	invasive lobular carcinoma
indel	insertion and deletion
ITH	intratumoral heterogeneity
KM	Kaplan-Meier
LBD	ligand-binding domain
LCIS	lobular carcinoma in situ
mAb	monoclonal antibody
MBC	metastatic breast cancer
ML	machine learning
MMR	mismatch repair
MRD	molecular residual disease / minimal residual disease
MRI	magnetic resonance imaging
MSI-H	microsatellite instability-high
NA	nucleic acid
NAC	neoadjuvant chemotherapy
NER	nucleotide excision repair
NGS	next-generation sequencing
NHEJ	non-homologous end-joining
NHG	Nottingham histological grade
OS	overall survival
pCR	pathological complete response
PCR	polymerase chain reaction
PFS	progression-free survival
PR	progesterone receptor
RFI	relapse-free interval
RFS	relapse-free survival

ROR	risk-of-recurrence
RPK	reads per kilobase
RT	radiation therapy
SBS	sequencing by synthesis
SCAN-B	Sweden Cancerome Analysis Network - Breast
sc-seq	single-cell sequencing
SERM	selective estrogen receptor modulator
SISH	silver in situ hybridization
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SSB	single stranded break
SV	structural variation
TAM	tamoxifen
TCGA	The Cancer Genome Atlas
T-DM1	trastuzumab emtansine
T-DXd	trastuzumab deruxtecan
TF	transcription factor
TGS	third-generation Sequencing
TKI	tyrosine kinase inhibitor
TMA	tissue microarray
TNBC	triple-negative breast cancer
TPM	transcripts per million
TSG	tumor suppressor gene
UTR	untranslated region
WES	whole-exome sequencing
WGS	whole-genome sequencing

1. Introduction

1.1 Hallmarks of Cancer

Cancer, a leading cause of death globally, represents a complex group of diseases characterized by many traits including unregulated cellular growth, proliferation, and invasion, with the “hallmarks of cancer” recently being updated to include now 14 key neoplastic capabilities [1–3]. The multifaceted nature of cancer, with a wide spectrum of types and subtypes, each with distinct molecular and histopathological features, presents a substantial challenge to healthcare systems worldwide.

Historically, the term 'cancer' traces its origin to the Greek word '*karkinos*' coined by Hippocrates, considered by many to be the father of medicine. The term was metaphorically used to describe the behavior of tumors, likening their spread to the legs of a crab. Throughout history, while the biological understanding and terminology of tumorigenesis have evolved, the word cancer has remained a constant representation of this complex and varied disease [4].

Epidemiologically, the incidence and mortality rates of cancer types vary significantly across different geographical regions, largely influenced by varying exposure to risk factors, hereditary predisposition, and disparities in healthcare resources [5]. Cancer is forecasted to become the leading cause of death worldwide within the 21st century and the greatest obstacle to extending life expectancy [6]. In many developed countries including Sweden, cancer represents a major public health challenge, with breast, prostate, lung, and colorectal cancers being the most common [7]. Lifestyle factors, including tobacco use, diet, alcohol consumption, and physical inactivity, are well-established contributors to cancer risk [8]. Additionally, environmental exposures, such as to chemical carcinogens or ionizing and non-ionizing radiation, along with genetic predispositions, play crucial roles in the development of cancer [9].

The advent of high-throughput genomic technologies has revolutionized the field of cancer research. In 2001, the first human genome draft sequence was published, a watershed moment in cancer research that paved the way for deeper understanding of the disease's genetic underpinnings [10]. This milestone and the comprehensive genomic analyses it has since enabled have

unearthed a plethora of genetic alterations and molecular mechanisms driving cancer development, significantly influencing the study of cancer by providing a comprehensive reference of alterations associated with each cancer type [11]. This knowledge has led to the identification of key driver mutations and disrupted signaling pathways, providing a more nuanced understanding of cancer biology [12]. Furthermore, these insights have identified new therapeutic targets and fueled the development of targeted therapies, marking a transition towards true personalized and precision medicine in oncology [13].

Cancers are broadly categorized based on their tissue of origin: carcinomas are the most common type, are derived from epithelial cells, and form solid tumors; sarcomas originate from connective tissues such as bone and muscle; myeloma develops from bone marrow plasma cells; leukemias and lymphomas originate from lymphocytes usually in the bone marrow or lymphatics, respectively; and mixed types exhibit characteristics of multiple categories [14]

Tumorigenesis (also called oncogenesis or carcinogenesis) is a multistep process involving the accumulation of genetic and epigenetic changes that confer a selective growth advantage to cells. These changes disrupt normal regulatory mechanisms governing processes such as cell proliferation, apoptosis, and differentiation, leading to the transformation of a normal cells towards a neoplastic state, eventually resulting in the development of cancer. All cancers were thought to originate from viruses 80 years ago [15]. However, a major paradigm-shift in 1960s revealed that some of our genes, termed, oncogenes, and then a decade later, tumor suppressor genes (TSG), could be causative culprits that go awry within our own cells and initiate and promote cancer [15]. Oncogenes and TSG are critical in promoting or inhibiting cell growth and proliferation, respectively. Mutations in these genes play a central role in cancer development. Oncogenes, initially normal genes or 'proto-oncogenes,' can become cancer-promoting when mutated or overexpressed by gain-of-function events. Examples include the HER2/*ERBB2* gene in breast cancer and the BCR-ABL fusion gene in chronic myeloid leukemia. The notion of Alfred Knudson's "two-hit hypothesis" also emerged in 1971, which indirectly led to the identification of TSGs. According to this hypothesis, a TSG (*RB1* in this case) requires both alleles to be inactivated, either through mutations or through epigenetic silencing, to cause a phenotypic change [16]. Other TSG were identified later, such as *TP53* and *BRCA1/BRCA2*, which function mainly to manage DNA damage [17], or *PTEN* which limits cell growth and proliferation [18]. Loss-of-function mutations or epigenetic changes that inactivate or suppress the expression of these tumor suppressor genes can lead to cancer development. The interplay between oncogenes and tumor

suppressor genes is a key aspect of the molecular basis of cancer [19]. However, some genes exhibits both oncogenic and tumor-suppressor functions under different cellular contexts [20]. For example, *TP53* gene mutations not only abolish the tumor suppressive functions but also equip the protein to acquire novel pro-oncogenic properties, by gain-of-function effect [21,22]. Another example is NOTCH receptors that are tumor-suppressors in squamous epithelial cells while plays an oncogenic role in T-lineage acute lymphoblastic leukemia [23,24].

A fundamental framework for understanding cancer was established by Hanahan and Weinberg, who delineated the hallmarks of cancer in their seminal works published in 2000 and later updated in 2011, 2022, and 2024 [1–3,25]. These hallmarks encapsulate the critical traits that distinguish cancer cells from normal cells, including sustained proliferative signaling, evasion of growth suppressors, resistance to cell death, replicative immortality, angiogenesis, and activation of invasion and metastasis. Later additions to these hallmarks include reprogramming of energy metabolism and evading immune destruction. This comprehensive framework is visually summarized in Figure 1.1.

Although these cancer hallmarks comprehensively delineate the framework encapsulating the various characteristics of cancer as a disease, a complex variety of mechanistic effects and phenotypes exists, both within the tumors and their micro-, macro- and immune- environments, which was recently expanded upon in the 2024 update to the hallmarks [25]. Briefly, the 13 additional systemic "clouds of complexity" includes tumor macro- and microenvironments, senescent cell states, aging systems and tissues, cancer cachexia, metabolic effects, dietary influences, pleiotropic immune responses, physiological prevention, thrombo-inflammation, obesity and physical activity, circadian clocks, and neural-cancer crosstalks. The authors also argued that these additional ''building blocks" of cancer, which govern the tumor initiation and promotion are yet incompletely understood.

The understanding of these hallmarks not only contributes to a deeper understanding of cancer biology but also guides the development of new therapeutic strategies targeting these specific characteristics of cancer cells. An example, which is also relevant to this thesis, is the "sustaining proliferative signaling" hallmark. In breast cancer (BC), this characteristic can be exemplified by the overexpression of the HER2 protein, a receptor tyrosine kinase that promotes cell growth and proliferation. Therapies such as trastuzumab (Herceptin) and pertuzumab, which are monoclonal antibodies (mAb) targeting the HER2 receptor, directly inhibit this proliferative signaling pathway [26–28]. This targeted approach has significantly improved outcomes for patients with HER2-positive breast cancer.

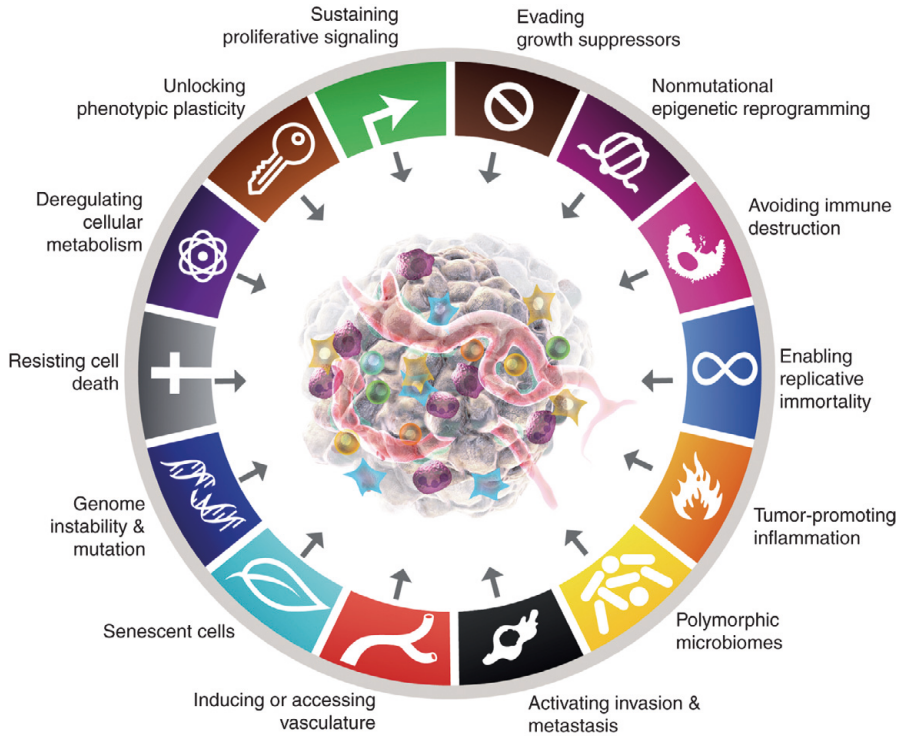

Figure 1.1: Hallmarks of Cancer from Hanahan & Weinberg [2]. Reproduced with permission from Elsevier.

Advances in screening, diagnostics, and treatment have improved survival rates, however cancer remains a significant burden and public health challenge, underscoring the need for continued research. Efforts must focus on improving early detection, understanding the complex molecular mechanisms underlying cancer progression, and developing effective, individualized treatment strategies. The ultimate goal of cancer research is to increase understanding and translate findings to enhance patient outcomes and reduce the global burden of cancer.

In this thesis, a multifaceted approach to breast cancer's complex biology was explored, by linking several key hallmarks of cancer, such as “sustaining proliferative signaling,” “resisting cell death,” “evading growth suppressors”, and “genomic instability and mutation”. For example, **Study I** links the exploration of the estrogen receptor beta (ERβ) and ER-mediated signaling to “resisting cell death.” In Study I, high ERβ expression was found to be significantly associated with improved survival, especially in contexts like endocrine therapy treated patients and triple-negative breast cancer (TNBC).

This suggests that it may contribute to "resisting cell death" by modulating immune responses, which may influence the tumor microenvironment's capacity of *ESR2*-high tumors making them more favorable to immune cell infiltration and less conducive to immune suppression. Next, in **Studies II** and **V**, HER2 signaling, crucial for "sustaining proliferative signaling," has been analyzed through ddPCR assay development that accurately quantifies HER2 copy number (CN) and its varied expression levels in breast cancer respectively. Additionally, in **Study II**, HER2/*ERBB2*-ultrahigh expression category, a subset with poor survival outcomes, emphasizes the complexity of "genomic instability and mutation" where this subgroup might exacerbate tumor aggressiveness and resistance to therapy. This precise quantification of CN and classification of HER2 expression at mRNA level into low, zero, and positive categories (**Study V**) may directly impact treatment decisions, especially with targeted therapies like trastuzumab and more recently FDA approved trastuzumab deruxtecan (T-DXd), an antibody-drug conjugate (ADC). In **Study III**, we showed that higher levels of CITED1 expression, linked to ER+ and luminal breast cancer subtypes, correlate with better outcomes in tamoxifen-treated patients, and appears to modulate the ER pathway in a way that enhances the effectiveness of anti-estrogen therapies. This finding underlines its potential role in "evading growth suppressors" by influencing ER signaling.

1.2 The Cancer Genome

Cancer is often referred to as a disease of the genome. While many factors can contribute to the onset and progression of cancer, in many cases genetics plays a critical role. In addition to one's inherited genetic variants, this can include somatic mutations that accumulate in cells over the course of a lifetime. Consequently, every cancer is in fact unique at the genome level, and understanding the specific drivers and vulnerabilities of an individual's cancer can give insights into how to treat it more effectively. The cancer genome and driver mutations are crucial areas of oncology research, impacting cancer development and treatments.

Germline and somatic mutations

Germline mutations, inherited and present in every cell, can predispose individuals with a higher risk of developing cancer, for example as seen with *BRCA1* and *BRCA2* mutations that are linked to breast and ovarian cancers. Women who inherit a deleterious *BRCA1* mutation face a 55-72% lifetime risk

of developing BC, while those with an inactivating *BRCA2* variant confront a 45-69% lifetime risk. These germline mutations also confer a heightened likelihood of cancer emerging in the contralateral breast following an initial BC diagnosis [29–31]. Other cancer-predisposing germline mutations in genes such as *TP53*, *CHEK2*, and *PTEN* also significantly contribute to cancer susceptibility. For example, *TP53* mutations are linked to Li-Fraumeni syndrome [32,33], also predisposing individuals to cancers of the breast [34], glioblastomas [35], sarcomas [36], and adrenal carcinoma [37]. *CHEK2* mutations are associated with an increased risk of colon cancer [38], prostate cancers [39], and double the risk for BC [40,41]. *PTEN* mutations are a hallmark of the autosomal dominant Cowden syndrome [42], and increases the risk for breast [43,44], thyroid [45], and endometrial carcinomas [46].

Somatic mutations, acquired in individual cells during a person's lifetime, are found in essentially all cancers. Somatic mutations can be 'driver' mutations or 'passenger' mutations. Mutations that provide a selective growth advantage, and thus promote cancer development, are termed drivers, and those that do not necessarily provide any significant advantage are termed passengers [47]. Acquisition of mutations in the *PIK3CA* oncogene, often found in various cancer types, is a typical example of driver somatic mutation. Given the inherent instability of cancer genomes, passenger mutations are quite abundant in cancer genomes, but their specific roles are less clear compared to driver mutations. A recent meta-analysis study examining 19 cancer types in The Cancer Genome Atlas (TCGA) dataset identified an average of 11 core driver mutation combinations per cancer. Each combination consisted of 2 to 6 alterations, highlighting the complexity and diversity of genetic changes driving various cancers. These finding signifies the complex genomic landscape of cancers and the importance of understanding these mutation combinations for targeted therapy strategies [48]. Another example driver mutation is the epidermal growth factor receptor (*EGFR*) mutations that are commonly observed in 43-89% of non-small cell lung cancer [49].

Solid tumors exhibit somatic mutations, with the amount varying significantly across different types of tumors. Somatic mutations can also play different roles at various stages of cancer development. For instance, some driver mutations arise early in the cancer's evolution, initiating the malignant transformation by providing a growth advantage to cells. However, as the tumor progresses, the dependence on these initial mutations may decrease, and they may no longer be essential for the survival of the cancer. A classic example is the driver gene that is often mutated early in the development of various cancers is *KRAS*. *KRAS* mutations are prevalent in about 30% of lung cancers and 90% of pancreatic cancers [50]. These mutations leads to tumor initiation by activating the KRAS protein, which triggers RAS/MAPK pathway

[51,52]. However, some studies have shown that as the tumor evolves, it can increase the expression or activity of other growth factor receptors or signaling molecules independent of KRAS, such as the Hedgehog signaling pathway, IGF-1R (insulin-like growth factor 1 receptor), or even rely on autocrine growth factor loops in pancreatic tumors [53,54]. This variation in mutational load is influenced by many factors such as the tumor's origin, environmental exposures to mutagens, intrinsic cellular processes like errors in DNA replication and repair, influence of metabolic activities, as well as various selective pressures (such as hypoxia, nutrient deprivation, and therapeutic interventions), cancer cells can adapt by upregulating or activating alternative pathways [55].

DNA damage and repair

To maintain cellular integrity and prevent oncogenesis, human cells have evolved highly attuned DNA repair pathways as well as DNA damage checkpoints to cope with the frequent challenge of endogenous and exogenous DNA insults. DNA damage can result from a multitude of factors, both internal and external, leading to changes that can disrupt the cell's genetic integrity. The DNA repair mechanisms are discussed below.

The causes of DNA damage include 1) oxidative damage in which the reactive oxygen species (ROS), byproducts of cellular metabolism, can oxidize nucleobases leading to modifications like 8-oxoguanine [56]; and 2) hydrolytic damage, where the DNA hydrolysis can cause deamination, depurination, or depyrimidination, creating sites without a base (apurinic/apyrimidinic sites, or AP sites) [57]. These changes are primarily repaired by the base excision repair (BER) pathway discussed below. 3) A third cause of DNA damage is ultraviolet radiation from the sun that can distort the DNA structure by disrupting the covalent bonding between adjacent pyrimidine bases, forming photoproducts such as thymine dimers and cyclobutane pyrimidine dimers and are typically repaired by nucleotide excision repair (NER) [58]. Fourth, 4) ionizing radiation can cause double-strand breaks (DSBs) and complex DNA damage, including crosslinks and clustered damage sites. DSBs are among the most lethal forms of DNA damage and are repaired through pathways such as homologous recombination (HoR) and non-homologous end joining (NHEJ) [59]. A fifth 5) cause is environmental chemicals and mutagens, such as polycyclic aromatic hydrocarbons and aflatoxins, which can add bulky groups to bases (adduct formation) or crosslink DNA strands. These are typically repaired by NER or, in the case of crosslinks, by a combination of NER and HoR [60].

DNA damage can manifest as either single-stranded breaks (SSBs) or DSBs, each presenting unique challenges and requiring different repair strategies. DNA SSBs are caused by oxidative damage, abasic sites, or DNA topoisomerase errors and can disrupt vital cellular processes like replication and transcription, potentially leading to cell death. SSBs can be corrected through BER, NER, and mismatch repair (MMR) [61,62]. During the replication phase of the cell cycle, DNA polymerases, an enzyme that copies DNA, also performs 'proofreading' by removing and replacing the incorrectly paired nucleotide right away, before continuing with DNA synthesis [63]. Following DNA replication, MMR removes and replaces mis-paired bases (ones that escaped correction during proofreading). MMR can also detect and correct small insertions and deletions that happen when the polymerases "slips" during replication, losing its correct positioning on the template [64]. Germline mutations in MMR genes such as *MSH2*, *MLH1*, and *MSH6* can lead to Lynch syndrome, a hereditary condition associated with an increased risk of colon, ovarian, and other cancers [65,66]. Malfunctioning of the MMR repair pathway also leads to the alterations of microsatellites (short, repeated sequences of DNA), and the overall mutational rate of a given cell increases. These kind of cancers are known as MSI-high (MSI-H), have treatment implications, and are most commonly are found in colorectal cancer, gastric cancer, and endometrial cancer [67,68].

BER corrects DNA by removing faulty bases through enzymes like OGG1, then fills the gaps with DNA polymerase and seals with ligase, ranging from short patch repair for single nucleotides to long patch repair for larger sections. One example of BER is the repair of uracil-containing DNA (caused by deamination of cytosine by hydrolysis, chemical agents, oxidative stress, or misincorporation of uracil by DNA polymerase during replication), where uracil is excised by DNA glycosylase, creating an abasic gap called an AP site. The gap is then cleaved by AP endonuclease and is filled using DNA polymerase and sealed with ligase [69].

NER corrects bulky DNA lesions caused by UV radiation or chemicals. It involves removing a damaged DNA segment and synthesizing a new strand. NER operates through two pathways: global genome NER (GG-NER) for repairing damage across the genome, and transcription-coupled NER (TC-NER) for lesions on the transcribed DNA strand [70]. Pre-clinical and clinical studies have revealed NER as a major resistance mechanism against cisplatin [71], and patients with alterations in the excision repair cross complement, *ERCC1*, had a high response to nivolumab due to their genetic instability [72]. Thus, NER pathway aberrations might predict the prognosis of cancer patients treated with immunotherapy. DNA DSBs, the most cytotoxic DNA lesions, initiate a comprehensive cellular DNA damage response (DDR). This process

involves the activation of the ATM kinase, part of the phosphoinositide 3-kinase-related protein kinase family, which rapidly associates with chromatin upon DSBs in coordination with the MRE11-RAD50-NBS1 (MRN) complex [73]. DNA DSB repair can be achieved by two broad categories depending on the use or not of a homologous DNA sequence as a template: HoR is a high-precision repair method utilizing a matching DNA sequence, often the sister chromatid, as a template. Active mainly in the S, G2, and M cell cycle phases, HoR involves creating single-stranded DNA, which is initially coated with RPA proteins and then replaced by Rad51 for pairing with a homologous template to facilitate repair [74]. Pathogenic mutations in a number of genes encoding HoR-related proteins are a well-studied cause of HoR deficiency (HRD), and, at the germline level, can confer risk for BC, but also occur somatically and contribute to sporadic breast cancer development, progression, and response to therapy [75].

NHEJ is a straightforward DNA repair mechanism that directly joins broken DNA ends without requiring a homologous template. Active across the cell cycle, this process involves proteins such as Ku70/Ku80, which protect DNA ends, as well as DNA-dependent protein kinase catalytic subunits, and XRCC4-LIG4 for end joining. NHEJ is quicker but tends to be more error-prone compared to HoR [73]. The tumor suppressors BRCA1 and BRCA2 play pivotal roles in HoR, maintaining a balance between HoR and NHEJ in healthy cells. However, when key DNA repair genes are deficient, this balance is disrupted. BRCA1's recruitment to DNA DSBs favors HoR repair, while the recruitment of 53BP1, another protein, promotes NHEJ, illustrating the delicate equilibrium in cellular DNA repair mechanisms [76,77].

Types of mutations

Across the genomic landscape, mutations may be classified at small or large genomic scales. Two types of small-scale mutations are prominent: 1) single nucleotide variants (SNVs; also known as point mutations) that involve the substitution of one nucleotide for another, and 2) small insertions and deletions, called indels, which introduce or remove short DNA segments of varying size but most commonly less than 50 base pairs (bp). These mutations have significant implications in genomic variation and disease processes, including cancer.

Single nucleotide variants and single nucleotide polymorphisms

Single nucleotide variants (SNVs) are central in cancer genomics, representing the most common mutation type. These mutations encompass six primary

substitution types, including transitions and transversions. Transitions, exchanges between purine to purine (G>A, A>G) or pyrimidine to pyrimidine (C>T, T>C), are more frequent due to structural compatibility compared to transversions, which is when a purine (A and G) and pyrimidine (C and T) interchanges – i.e., A>C, A>T, G>C, G>T, and their opposites C>A, T>A, C>G, and T>G.

SNVs can be classified based on their genomic location and impact on protein coding. If in the coding region of a protein, a SNV may be synonymous/silent (not altering amino acids) or non-synonymous (altering amino acids), and if outside of a coding region, then it is termed a non-coding SNV. A non-synonymous SNV can lead to different types of mutations in a protein. These include "missense" mutations, which change one amino acid to another, potentially altering mRNA and protein function, "nonsense" mutations which create a premature stop codon (UAA, UAG, or UGA in mRNA) resulting in a truncated protein, or a "read-through" or "stop loss" mutation, that leads to loss of a stop codon and thereby a longer than normal protein [78]. Changes in the amino acid structure and length of proteins often affects their normal functions, which can directly or indirectly contribute to cancer initiation and progression [79]. Another type of mutation called a regulatory region mutation can occur in promoters, enhancers, silencers, and insulators. For example, a substitution mutation could potentially disrupt the binding of transcription factors or other regulatory proteins, thereby affecting gene expression [80]. SNVs and regulatory region mutations are depicted in Figure 1.2. Non-coding and synonymous SNVs, once considered less impactful, are now recognized for their roles in oncogenesis. They may affect DNA binding domains and gene regulation, transcription efficiency, mRNA secondary structure, translation and splicing, for example as has been shown for specific mutations in *TP53* [81] and *KRAS* [82].

A germline SNV can be defined as a single nucleotide polymorphism (SNP) when it is found in a sufficiently large fraction of the population, typically >1% of the population. This threshold distinguishes SNPs as common normal variations in the human genome rather than rare mutations and are used as genetic markers that are important in studying genetic diversity, disease susceptibility, and even response to drugs. SNP genotyping, measurement of genetic variations at known SNP positions, can today be performed genome-wide at relatively low-cost using tools such as SNP microarrays.

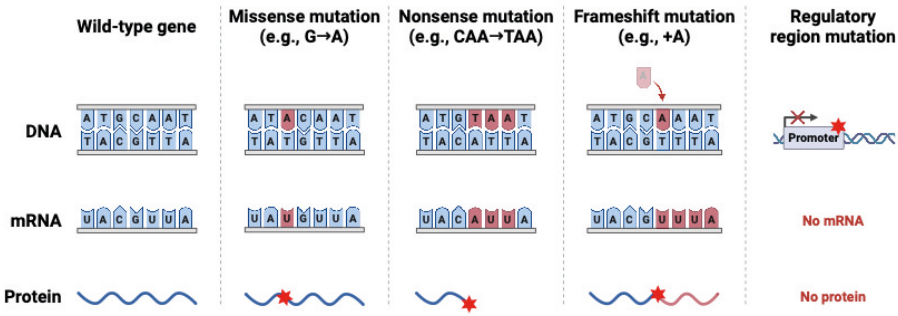

Figure 1.2: Types of SNV and indels. The effects of SNVs (missense and non-sense mutations) as well as indels (frameshift mutations) on downstream mRNA and protein product is shown. Regulatory region mutations can occur as either SNVs or indels. Created with BioRender.com.

Small insertions and deletions

Indels are pivotal in cancer genomics for both diagnostic and therapeutic implications because they can significantly impact gene function and are a common mechanism of kinase activation. These mutations, often comprising less than 50 bp, alter the reading frame of genes, leading to abnormal or dysfunctional proteins. Indels can be broadly categorized into two types: 1) frame-shift indels, which alter the DNA reading frame, leading to significantly different and often non-functional proteins (Figure 1.2); and 2) in-frame indels, which do not disrupt the reading frame but rather introduce or delete a small number of amino acids in the protein [83,84]. Accurate identification of indels in genomic studies poses a challenge due to presence of repeats, short interspersed elements, homopolymers/dimers and type of indel detection methods used [85]. Some NGS based studies have reported that indels are severely under-reported due to difficulties in accurate indel detection and it is estimated that, due to false-negative rates, up to one-third of the small indels in human genomes may be undetected [86].

Large chromosomal rearrangements: structural variations

Larger chromosomal aberrations, classified as structural variations (SV), are significant alterations defined by variations of at least 50 bp in size (typically much larger) and include large insertions and deletions, segmental duplications, amplifications, inversions, translocations, and chromosomal losses and gains, as well as more complex DNA rearrangement patterns (Figure 1.3A) [87]. SVs are prevalent in cancer, emerging during tumor development,

progression, and even in resistance to treatment [88]. SVs can lead to the amplification or disruption of crucial driver cancer genes or alter gene expression by rearranging coding or noncoding DNA elements to juxtapose regulatory regions or create fusion genes with new functions, making SVs highly significant in the study of genomic instability and cancer development. SVs often occur amidst periods of genomic or chromosomal instability, leading to intratumor heterogeneity and complex rearrangement events like chromothripsis [89] and chromoplexy [90]. The classic example of a clinically actionable gene fusion SV is the so-called Philadelphia chromosome, a translocation between chromosomes 22 and 9 creating the BCR-ABL1 fusion gene in chronic myelogenous leukemia, and which is treatable with specific tyrosine kinase inhibitors such as imatinib [91].

Pan cancer WGS data generated in large patient cohorts revealed that somatic SVs are the most common class of driver mutation in cancer, surpassing SNVs and indels in terms of proportion [92,93]. International initiatives like the Pan-Cancer Analysis of Whole Genomes (PCAWG) project, conducted by collaborations such as the International Cancer Genome Consortium (ICGC) and TCGA, have provided a comprehensive overview of the prevalence of somatic SVs in cancer. Their analyses reveal that across various cancers, SV – which include partially resolved SVs grouped with somatic copy-number alterations – account for approximately 55% of driver mutations, thus surpassing the number of point mutation drivers [93].

Despite several large pan-cancer analyses, compared to SNVs, SVs are rather underexplored in cancer genomes. Their discovery and characterization pose challenges, particularly with complex SVs being common in cancer, as they are difficult to detect with short-read sequencing data. Often, cancer studies categorize SVs under the broad term "somatic copy-number alterations" without deeply investigating their specific roles, structural details, or mechanism of formation. Interpreting somatic SVs is more complex than point mutations as they can impact multiple genes or entire chromosomes, with their effects on cancer gene expression frequently mediated by alterations in intergenic regulatory sequences [94].

Of particular interest to this thesis, copy number gains, a type of copy number variant (CNV) or copy number alteration (CNA), refers to a genomic change where additional copies of a section of DNA are present. This results in an increased number of certain gene copies within a cell, which if it is affecting an oncogene, it can lead to its overexpression and contributing to cancer development and progression. CNVs arise through diverse mutational processes, which include DNA recombination, replication, and repair mechanisms. As depicted in Figure 1.3B, CNVs can be classified as 1) recurrent CNV that have same size and show common breakpoints enriched in repeated

sequences such as low-copy repeats (LCRs; e.g., segmental duplications). LCRs and high-copy repeats (e.g., SINEs, LINEs), are often found near breakpoints, playing a critical role in CNV instability. These sequences are crucial for triggering key mechanisms for generation of recurrent CNV like non-allelic homologous recombination (NAHR) [95]. Secondly they can be classified as 2) non-recurrent CNVs which scattered breakpoints with different sizes but which may share a smallest region of overlap (SRO) between different patients [96] (see Figure 1.3B). They might be formed by mechanisms such as NHEJ and microhomology-mediated end joining (MMEJ) or replicative mechanisms such as replication slippage, fork stalling and template switching or microhomology-mediated break-induced replication [97].

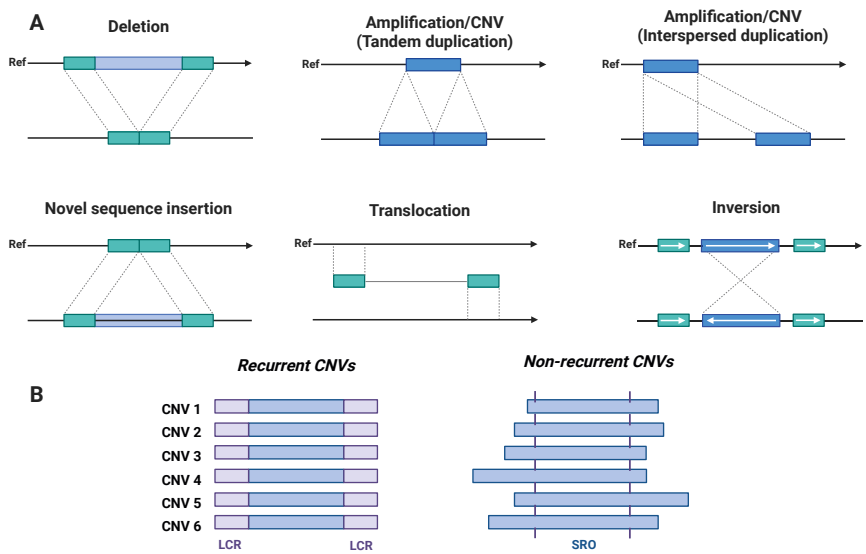

Figure 1.3: Types of chromosomal rearrangements. A. Different types of large structural rearrangements including large deletions, CNVs, large insertions, translocations, inversion are shown. B. Recurrent and non-recurrent rearrangements are shown. Recurrent CNVs have the same size and common breakpoints enriched in LCRs. Non-recurrent CNVs with different sizes may share SRO. Abbreviations: LCR: low-copy repeats; SRO: Smallest Region of Overlap; Ref: reference chromosome. Created with BioRender.com.

The locus-specific mutation frequency for CNV and other SV have been estimated to be two to four orders of magnitude greater than for point mutations [97]. CN gains are significant markers in cancer genomics for understanding tumor biology and guiding treatment strategies, precision

oncology and diagnostics [86,92,96,97]. In **Study II**, an absolute quantitative assessment of *ERBB2* CN was performed using a novel digital PCR (dPCR) assay, which was also shown to be predictive of outcome in early breast cancer patients treated with adjuvant trastuzumab.

Mutational signatures

Mutational signatures are distinctive patterns of mutations found in cancer genomes, reflecting the natural history of the mutations and their underlying mechanisms. First characterized by Alexandrov et al. using a non-negative matrix factorization (NMF) method, they determined a set of signature profiles and could estimate the contributions of each signature to each cancer genome [98]. They classified mutational signatures into different categories, single base substitutions (SBS), double base substitutions (DBS), multi-base substitutions (MBS), and indels, and made these signatures available via the Catalogue of Somatic Mutations in Cancer (COSMIC) database [99]. These signatures can be attributed to various factors, including exposure to environmental agents (e.g., smoking, UV), defects in DNA repair mechanisms such as BRCA1, activities of APOBEC cytidine deaminases, as well as to the physiological state of cells. Each category has distinct characteristics and implications for understanding cancer genomics. SBS involve single nucleotide changes, DBS are adjacent nucleotide pair changes, whereas MBS cover larger sequences, and indels involve insertions or deletions of DNA segments. These classifications help in understanding the complex mutational landscape of cancers and may provide valuable insights into the etiology of cancers and may provide predictive information independent of known clinical and molecular biomarkers [100].

Tumor heterogeneity

Tumor heterogeneity can be divided broadly into two types. Intratumoral heterogeneity (ITH) refers to the presence of genetically diverse cancer cells within a single tumor. This variability arises from mutations, genetic rearrangements, and epigenetic modifications that occur over the course of tumor development. As cancer cells proliferate, they can accumulate differences in their genome, leading to subpopulations of cells with distinct molecular and phenotypic characteristics. This diversity within a single tumor can result in significant challenges in treatment, as different subclones within the tumor may respond variably to therapies. For instance, some clones might be sensitive to a particular chemotherapy drug, while others might be resistant, leading to treatment failure and disease progression. ITH is also a major contributor to cancer relapse, as minor resistant subpopulations can

survive initial treatment and repopulate the tumor [101]. Conversely, inter-tumor heterogeneity describes the differences observed between separate tumors within the same patient, or in the most extreme sense, between tumors in different patients. This form of heterogeneity can occur between a primary tumor and its metastases or between multiple primary tumors in cases of multifocal diseases. Each tumor can develop unique genetic and epigenetic profiles, influenced by its microenvironment, which can affect its growth, metastatic potential, and response to therapy. For example, a primary breast cancer tumor might have a different genetic makeup compared to metastatic sites in the bone or brain, complicating devising the treatment strategies. This type of heterogeneity is particularly challenging in cases where metastatic disease needs to be treated differently from the primary tumor due to distinct characteristics that influence drug sensitivity and resistance [102].

ITH is considered one of the main causes for poor prognosis and outcomes in cancer patients because it leads to therapeutic resistance and treatment failure [103]. Moreover, ITH presents significant challenges in accurately understanding and treating a tumor due to issues related to tumor sampling. If the sampled tissue which was sequenced represents only a fraction of the tumor's cellular diversity, it may not capture the full spectrum of the genetic variability present. This incomplete view can lead to misunderstandings about the tumor's genetic landscape and potentially lead to ineffective treatment decisions [104]. Related to issues of cancer genome interpretation and ITH, it is worth mentioning several additional factors which can influence interpretation of cancer genomes. Good sample quality, the quality and quantity of DNA or RNA extracted from a tumor sample can affect the accuracy and completeness of sequencing results. Another factor is Tumor purity: the proportion of tumor cells in a sample can affect the sensitivity of sequencing assays. Samples with low tumor purity may require deeper sequencing to detect mutations with high confidence. Pre-analytical factors pose a challenge too, the way in which tumor samples are handled, processed, and stored prior to sequencing can affect the quality and quantity of DNA or RNA extracted. Analytical parameters used during sequencing data analysis, such as quality score thresholds, variant allele frequency cutoffs, and filtering criteria, can also influence the types and number of mutations that are identified. The type of sequencing analysis performed (i.e., whole genome sequencing or a gene panel) will also influence which portions of the genome are characterized, and how much data will be available about each of those regions. The type of software used for analysis and the chosen parameters will determine what type of variants (i.e., SNVs, CNVs, rearrangements) are identified and whether individual variants will be deemed significant or not. All these factors will influence the final list of variants, ITH and mutation burden.

Clonal diversity is a key aspect of ITH. Tumors can consist of multiple clones and subclones, each with distinct genetic and phenotypic characteristics. These subpopulations can evolve and adapt over time and understanding the clonal structure of tumors might be an important aspect for developing effective treatment strategies that target the diverse cell populations within a cancer [105]. For example, different subtypes of breast cancer, like HER2-positive, triple-negative, and hormone receptor-positive, exhibit distinct molecular profiles. Within a single tumor, genetic diversity can lead to various cell populations, affecting responses to treatment and contributing to drug resistance. This complexity poses challenges in treatment and underscores the need for personalized approaches [106]. In this thesis ITH has been touched upon in **Studies I**, **II**, and **IV**.

1.3 The Cancer Transcriptome

Genetic alterations and gene expression changes in cancer can be analyzed across at least three levels: genomic, transcriptomic, and proteomic. Each level offers unique insights and poses specific challenges. Understanding the complex interplay among these levels is crucial for personalized oncology, helping tailor treatments to individual genetic and molecular profiles.

Transcriptomics refers to the study of all the RNA transcripts in a sample, typically by using high-throughput technologies such as microarrays or RNA-sequencing (RNA-seq) [107]. As compared to the cancer genome, which represents the blueprint of the cell's genetic program, the cancer transcriptome is more dynamic because measurement of transcripts varies with biologic activity and cellular status over time. Transcriptomics can encompass assessments of complete set of RNA transcripts, including coding mRNAs and non-coding RNAs, alternative splicing, alternative polyadenylation, identification of fusion transcripts, explorations of noncoding RNAs (tRNA, miRNA, piRNA, siRNA, lncRNA, enhancer RNA), transcript annotation, and discovery of novel transcripts, but usually only a subset of these are investigated at any one time. Given its dynamic and comprehensive nature, transcriptomics serves as a valuable resource for understanding cancer mechanisms and identifying biomarkers [108]. While the genomic level can be regarded as the farthest from the cellular phenotype, transcriptomics can be seen as an intermediatory level, whereas the protein level may be seen as being closer to the cellular phenotype.

Coding mRNAs translate into proteins, which can include oncogenes and tumor suppressors, directly impacting cancer development and progression. Non-coding RNAs, like microRNAs and long non-coding RNAs, play crucial roles in

gene regulation and can influence cancer cell behavior [109]. For example, miRNAs can regulate gene expression post-transcriptionally by binding to the 3' untranslated regions (UTRs) of target mRNAs, leading to their degradation or inhibition of translation. A well-documented example of how miRNAs can influence cancer is miR-21, which is often upregulated in various cancers, such as gliomas, breast cancer, and colorectal cancer. miR-21 targets and downregulates the tumor suppressor PTEN [110]. Additionally, lncRNAs through interactions with proteins, DNA, and RNA, can modulate gene expression at multiple levels including by changing chromatin structure and the post-transcriptional processing of RNA. One notable mechanism is through the competing endogenous RNA (ceRNA) hypothesis, where lncRNAs act as molecular sponges to sequester miRNAs, reducing their regulatory impact on target mRNAs. An example of this is PTEN and its ceRNA the *PTEN* pseudogene (*PTENP1*), which can bind miRNAs that normally are targeting the PTEN mRNA. By sponging these miRNAs, *PTENP1* indirectly increases PTEN expression, thus impacting cell growth and apoptosis pathways [111].

Gene expression is regulated by a combination of genetic and epigenetic factors, including DNA methylation and histone modifications. Additionally, the epitranscriptome, which involves modifications on RNA molecules, adds another layer of complexity, influencing both gene expression and protein translation. In cancer, the delicate balance among these regulatory mechanisms is disrupted, resulting in altered gene expression patterns, splicing anomalies, and aberrant RNA modifications. Studying the transcriptome offers critical insights into the functional consequences of genetic alterations in cancer, bridging the gap between DNA and the proteome, and aiding in the identification of new therapeutic targets and a deeper understanding of cancer biology at a molecular level.

As the field of oncology continues to evolve, the integration of transcriptomics alongside genomics in clinical trials has proven to be a groundbreaking strategy for advancing precision medicine. This approach not only identifies clinically actionable molecular alterations but also deepens our understanding of the mechanisms underlying different cancers. An exemplary illustration of this advancement is the WINTHER precision medicine clinical trial, one of the first prospective trials in diverse solid malignancies to assess both genomics and transcriptomics to tailor treatments to specific molecular alterations. This trial demonstrated how transcriptomic analysis could significantly expand the number of patients effectively matched to drugs [112]. Other noteworthy precision medicine trials that incorporated gene expression analysis at coding and non-coding RNA level include PROVABES [113], INFORM [114] and PIPseq [115]. In this thesis, large scale transcriptomics analysis is performed in **Study I** and **IV**.

1.4 Breast Cancer

1.4.1 Mammary gland development

The mammary gland, or breast, distinguishes mammals from all other animals due to its unique anatomical structure. It evolved from apocrine sweat glands approximately 300 million years ago, highlighting a significant evolutionary adaptation that has played a crucial role in mammalian reproduction [116]. The structure and development of the mammary gland are highly conserved across mammals. Mammary gland development is a dynamic process influenced by hormonal, genetic, and environmental factors. It starts in embryogenesis, progresses through puberty, pregnancy, lactation, and finally involution. This complex organ comprises various cell types like epithelial cells that grow from the nipple into a fat pad, formed by adipocytes and infiltrated by vascular endothelial cells, fibroblasts and immune cells. Signals from the mesenchyme direct embryonic development, while circulating hormones released from the pituitary and ovary influence changes in puberty and adulthood.

The mammary epithelium consists of basal and luminal cells. Basal epithelium comprises myoepithelial cells forming its outer layer and a small stem cell population supplying various cell types. The luminal epithelium in the mammary gland is characterized primarily by the presence of hormone receptors for estrogen and progesterone. These hormone receptors are critical in defining the function and behavior of luminal cells, influencing their role in forming the ducts and secretory alveoli essential for milk production. These two epithelia create a bi-layered structure essential for lactation. Puberty triggers ductal growth and branching and is under hormonal influence, especially estrogen. During pregnancy, hormones such as progesterone and prolactin stimulate lobuloalveolar development, preparing for milk production. Post-lactation, the gland undergoes involution, reverting to a near-pre-pregnancy state [117]. The terminal duct lobular unit is the functional unit of the breast, and is made up of a cluster of alveoli connected to an adjoining ductule, as illustrated in Figure 1.4.

During menopause, the mammary gland typically undergoes involution, marked by the senescence and size reduction of the terminal duct lobular units. Incomplete and lobular involution which is a histological observation as well as mammographic breast density (MBD), a radiological assessment, have been found to be associated with an increased risk of developing breast cancer and may also serve as prognostic factors [118].

The female reproductive hormones, estrogens, progesterone, and prolactin are crucial for female reproductive processes and control postnatal mammary gland development and have a major impact on breast cancer. Estrogen synthesis mainly occurs in ovaries, but other tissues such as adipose tissues and osteoblasts also contribute. Four major types of estrogens exist: estrone (E1), estradiol (E2), estriol (E3) and estetrol (E4), with E2 being the most potent and prevalent form and dominant during fertile years. E1 is produced after menopause, whereas E3 and E4 are produced only during pregnancy. In the breast, estrogen drives duct development and influences prolactin secretion, while progesterone, aided by estrogen-induced progesterone receptor (PR) transcription, is vital for alveolar growth. These hormones play key roles in breast tissue development and function [119].

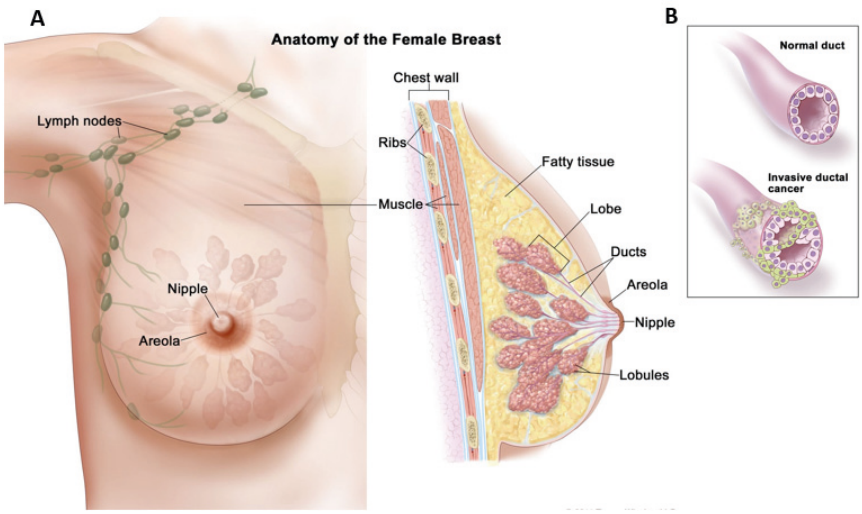

Figure 1.4: The female breast. A. The nipple, areola, lymph nodes, lobes and lobules, ducts, muscles, chest wall, ribs, and fatty tissue are illustrated. B. The normal duct architecture and invasive ductal carcinoma are depicted. Reprinted and adapted from National Cancer Institute © 2011, Terese Winslow LLC.

1.4.2 Breast cancer epidemiology

Incidence

Breast cancer is the most frequently diagnosed cancer among women globally, surpassing lung cancer as the most common cancer worldwide. It accounts for

approximately 11.7% of new cases, with an estimated 2.3 million women diagnosed annually. This is followed by lung (11.4%), colorectal (10.0%), prostate (7.3%), and stomach (5.6%) cancers. [5,120]. This rise in incidence may be attributed to improved detection, the availability and utilization of BC screening methods such as self-breast examinations and mammography, depending on a country's developmental index, where developed countries have the highest incidence [121]. Comparing cancer incidence rates is key for understanding cancer risk factors, guiding cancer control initiatives, and evaluating prevention programs. For instance, low and medium low and medium development index countries have also been experiencing increases in incidence due to a combination of factors including physical inactivity, excess body weight, delaying the onset of childbearing, shortening the breastfeeding period, and increasing the use of oral contraceptives that all contribute to BC development [122]. Although rare, men can develop breast cancer as well, with about 1% of all BC being diagnosed in men and invasive ductal carcinoma being the most common type among men, constituting approximately 90% of all male BC [123].

In Sweden, breast cancer is the most common type of cancer among women. In 2021, 8619 women were diagnosed with breast cancer and 1326 women died with breast cancer as the underlying cause of death [7,124]. In 2021, there was a 3.9% increase in the number of women diagnosed with BC as compared to the average of the three years prior to the start of the pandemic in 2020 [124]. Breast cancer affects women of all races and ethnicities, but there are notable differences in incidence and outcomes across these groups. Although the reasons behind racial and ethnic disparities are not fully understood, for example in the United States BC incidence is highest among white non-Hispanic women [125].

Mortality

In terms of mortality, breast cancer mortality rates show significant global variation, being higher in socio-economically disadvantaged areas due to limited access to early screening and treatment. In developed countries, some disparities are also evident, for example with higher mortality rates among black women compared to white women in the United States [126].

Globally, according to GLOBCAN 2020 estimates, BC is the fifth leading cause of mortality and the leading cause of death from cancer among women, accounting for 6.9% of all cancer-related deaths in 2020 [127]. One in 6 cancer-related deaths is due to BC, making it a leading cause of cancer mortality in 110 countries [121]. In the developing world, women have a 17% higher BC mortality rate due to a lack of defined early screening standards and poor or

delayed access to treatment [128]. Since the mid-1990s, Northern European countries have experienced a decline in mortality rates [129].

Despite a high number of total deaths, short- to mid-term survival rates for breast cancer patients are relatively favorable compared to other cancers, with a 98% survival rate at one year and 88.5% at five years [130]. However, it's important to note that a 5-year survival rate does not equate to a cure. In fact, the rate of relapse of BC approximates a nearly straight line, with about an equal number of relapses occurring before 5-years as the number that occurs after 5-years of diagnosis. Long-term survival rates decline significantly, with only 60% of patients surviving 15 years and 50% surviving 20 years, often due to the recurrence of the disease [131].

Age-standardized incidence rate and mortality rate (ASIR and ASMR, respectively), allows for meaningful comparisons across different demographic segments, aiding in understanding risk factors and shaping cancer control policies. ASIR with substantial regional differences is illustrated by data from the World Health Organization for 2020 in Figure 1.5.

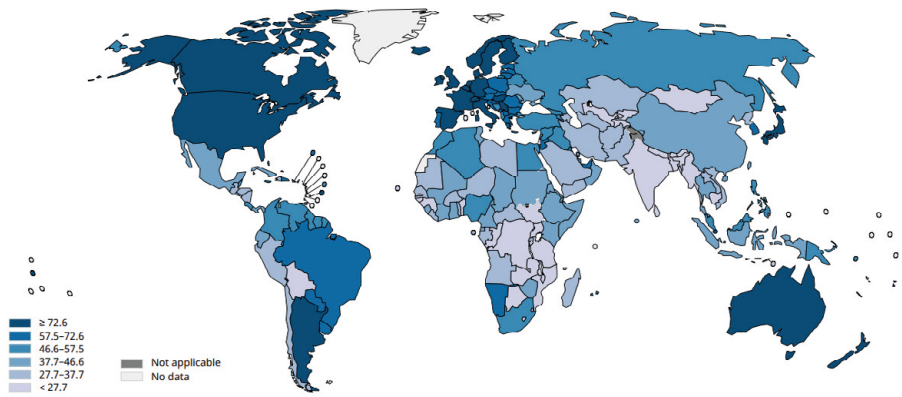

Figure 1.5: Global breast cancer incidence. Estimated ASIR for breast cancer per 100,000 women for the year 2020 are shown. Source: World Health Organization Global Cancer Observatory (https://gco.iarc.fr).

Early detection, enhanced awareness and advancements in treatment have notably lowered BC mortality in developed regions. Yet, in less affluent parts of the world, high mortality persists due to healthcare disparities [132]. Implementing organized screening, establishing regularly updated reliable breast cancer registries, and expanding mammography in these regions can boost early diagnosis, decrease mortality, and lessen the global impact of BC

[133]. That being said, it should be noted that BC has seen remarkable advancements in diagnosis, treatment, and patient survival, outpacing many other cancer types in the past decades [134]. This significant progress reflects a triumph in medical science and patient care.

1.4.3 Risk factors

A woman's likelihood or risk of developing BC is influenced by *non-modifiable* risk factors, such as sex, aging, race, *BRCA1* or *BRCA2* mutations, family history of breast cancer, and familial breast cancer syndromes, as well as by *modifiable* risk factors, such as alcohol use, smoking, obesity, and physical activity. Among non-modifiable risk factors, female gender is the strongest breast cancer risk factor, with BC diagnosis most common in middle-aged and older women, with an increasing cumulative risk of developing BC of 1.5% at age 40, 3% at age 50, and more than 4% at age 70 [135]. Globally, a woman has approximately a 1 in 8 chance, or about 12.5%, of developing breast cancer at some point during her life. This statistic can vary significantly depending on a range of factors including genetics, lifestyle, and environmental exposures, as well as regional healthcare practices and screening rates [120,136]. BC subtypes and age has also an observed link, where aggressive, resistant TNBC is more common in individuals under 40, while the luminal A subtype predominates in patients over 70 years of age [135]. This pattern reflects a broader trend where cancer risk increases with age, likely due to the accumulation of cellular alterations and prolonged exposure to carcinogens over time. Circulating estrogens and androgens are also positively associated with the risk for breast cancer in premenopausal women, with multiparity being protective [137]. Having a family history of breast and ovarian cancer (especially those characterized by *BRCA1* and *BRCA2* mutations) is a significant risk factor for breast cancer, with about 13-19% of breast cancer patients reporting a first-degree relative with the same disease [138,139]. Genetic mutations in *BRCA1* and *BRCA2*, mostly inherited in an autosomal dominant manner, are strongly linked to increased breast cancer risk, though sporadic mutations are also common [140]. As stated previously, women with a germline *BRCA1* mutation have a significantly elevated 55-72% risk of developing BC by age 70-80. For those with a germline *BRCA2* mutation, the risk ranges from about 45% to 69% [141,142]. Other key genes associated with high breast cancer risk include *TP53* [143], *CDH1* [144], *PTEN* [145,146], and *STK11* [147].

Reproductive history including early menarche, pregnancy, breastfeeding, menstruation, menopause, hormonal imbalance, along with duration and exposure to hormones such as E2 and progesterone, are crucial factors that influence the breast microenvironment and BC risk [148]. Pregnancy and

lactation or breast feeding are protective factors against breast cancer, where the risk is reduced by 4.3% for every 12 months of breastfeeding [149]. Higher breast tissue density correlates with increased risk for breast cancer, a trend observed both in pre- and post-menopausal women [150]. A personal history of breast cancer as well as non-cancerous breast alterations such as atypical hyperplasia or carcinoma in situ significantly increases the risk of new cancerous lesions in the breasts [151].

Among modifiable risk factors, physical activity among females with a family history of breast cancer, was found to be associated with a reduced risk of cancer, during both the premenopausal and postmenopausal periods [152]. The reason might be that being physically active and exercise might alter immune system responses, reduce the exposure to the endogenous sex hormones as well as lower IGF-1 levels [153]. Obesity and a higher body mass index (BMI ≥ 25 kg/m^2) are known risk factors that increase the incidence of BC [154]. Data shows that these factors are associated with poorer clinical outcomes and more aggressive tumor characteristics, including a higher percentage of lymph node metastasis and larger tumor sizes [155]. Obesity may be a reason for greater mortality rates and a higher probability of cancer relapse, especially in premenopausal women [156]. Alcohol consumption was observed to be associated with increased risk of BC, in particular for ER+ BC [157], perhaps due to direct and indirect carcinogenic effects of alcohol metabolites and alcohol-related impaired nutrient intake [158], or due to hormonal imbalances caused by increased levels of estrogens induced by alcohol intake [159]. Active as well as passive smoking significantly contributes to initiation of cancer by directly causing DNA damage, increasing the possibility of mutations within oncogenes and/or tumor suppressor genes [160]. Dietary supplementation with vitamin D has been shown to be associated with lower mortality rates in BC patients [161]. A diet high in ultra-processed foods, which are now classified as a Group 1 carcinogen, has been shown to increase the risk for not only BC but also for gastrointestinal cancers [162]. Chronic exposure to harmful chemicals such as pesticides, insecticides, hydrocarbons, and solvents may promote breast carcinogenesis by altering the tumor microenvironment and inducing epigenetic changes along with DNA damage [163,164]. Prolonged use of hormone replacement therapy (HRT) longer than 5 or 7 years also increases risk of BC [165,166] and lengthy use of specific drugs including a number of antidepressants, antibiotics, non-steroidal anti-inflammatory drugs, as well as statins, have been shown to be associated to an elevated risk of BC [167,168].

1.5 Breast Cancer Biomarkers

Breast cancers are clinically stratified based on key biomarkers that significantly influence treatment decisions and prognosis. These biomarkers include the expression of the estrogen receptor (ER), progesterone receptor (PR), and the overexpression or gene amplification of HER2 (encoded by the *ERBB2* gene). While ERα is acknowledged for its well-established significance in breast cancer pathology and treatment, ERβ is of particular focus in this thesis. In this section, the roles and actions of biomarkers ERβ, along with CITED1 and HER2, are detailed.

1.5.1 Estrogen receptors

Estrogens play a crucial role in the development of both normal and malignant mammary tissues. Their biological effects are mediated by estrogen receptors, specifically ERα and ERβ, which are part of the nuclear receptor superfamily of transcription factors. These receptors are characterized by highly conserved DNA-binding and ligand-binding domains [169].

Jensen and Jacobson first observed estrogen retention in hormone-responsive tissues in the 1960s [170]. Later, Jensen and colleagues discovered the existence of intracellular estrogen-binding receptors [171,172]. In 1986, the ERα gene was cloned by Chambon's group, revealing its DNA-binding domain with zinc finger motifs and a ligand-binding domain, typical of transcription factors [173,174]. They also identified the N-terminal activating function (AF-1) domain, crucial for ER's transcriptional activity through protein interactions. ERα protein consists of 595 amino acids (aa), encoded by gene *ESR1*, located on chromosome 6.

A decade later, in 1996, estrogen signaling was further refined and, adding to the mechanistic complexity, a second ER gene was discovered from rat prostate by Gustafsson and Kuiper and was named ERβ [175]. As depicted in Figure 1.6, ERβ is encoded by the *ESR2* gene located on chromosome 14, produced from eight exons, and comprising a 530 aa protein (full-length ERβ, the ERβ1 isoform) [176]. The genes for both estrogen receptors, ERα and ERβ, consist of eight exons and are alternatively transcribed into isoforms. These isoforms have different C-terminals but contain similar functional domains, exhibiting varying degrees of homology, as described below [177].

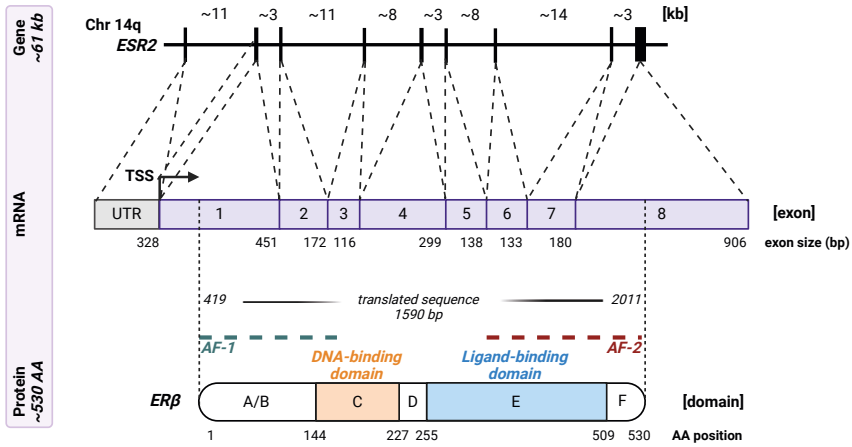

Figure 1.6: Schematic representation of *ESR2* gene. Mapping of gene organization to mRNA and its protein sequence are illustrated. Abbreviations: AF1: activating function 1; AF2: activating function 2; TSS: transcription start site. Figure modified from [178]. Created with BioRender.com.

A third known ER, G protein-coupled estrogen receptor 1, GPER (formerly known as GPR30), is a G protein-coupled receptor (GPCR) that was cloned in the late 1990s [179,180]. Unlike the classical ERs, GPER predominantly induces rapid, non-genomic estrogen signaling. Its activation leads to a multitude of downstream signaling events, including Ca^{2+} mobilization, cyclic adenosine monophosphate synthesis, and indirect activation of kinase pathways, including phosphatidylinositol-3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) pathways through trans-activation of EGFR [180,181].

ERβ signaling

ERα and ERβ share highly homologous protein domains, with different degrees of sequence conservation. The tertiary structure of ERα and ERβ includes an N terminal AF1 or A/B domain that mediates weak ligand-independent transcriptional activity. The DNA binding domain (DBD) or C domain is extremely well-conserved between ERα and ERβ with 97% homology and contains 2 functionally different zinc finger motifs, responsible for specific DNA binding to estrogen response elements (EREs) with high affinity, and mediates a weak receptor dimerization. The hinge domain, also known as the flexible D domain, includes a nuclear localization signal motif and a weak dimerization interface. The ligand-binding domain (LBD), or E domain, is highly conserved

and encompasses a ligand-binding pocket and the transactivation function (AF2). This domain also features a dimerization interface and is responsible for activating transcription in response to ligand binding. The DBD and LBD of the ERβ are 96% and 60% homologous with those of ERα, respectively, indicating that they may have similar but not identical functions [182]. The D domain, often referred to as the hinge domain, serves as a link between the DBD and LBD. The C-terminal or F domain, which is less conserved across different species, plays a role in negatively regulating ligand-dependent dimerization of both receptors [183–185].

ER is present in a cytosolic complex as a free unliganded form along with chaperones HSP70 and HSP90 and associated proteins. Binding of ligand to the AF2/LBD results in homo (α/α, β/β) and heterodimerization (α/β) with ERα and with other ERβ isoforms and dissociation from HSP70 and HSP90 as well as cytoplasmic relocalization of HSP90, and binding to EREs in promoters of estrogen-regulated genes [186,187]. While ERα and ERβ can bind to most EREs identically, transcription regulation depends on the ER subtype, tethering differential transcription factors and then modulating different target genes [177,188]. For example, in ERα, coactivators simultaneously bind to AF1 and AF2 to reach full activation, while ERβ has a lower activity of AF1 but a completely functional AF2 [189]. Generally, binding of agonist results in persistent nuclear localization, while that of anti-estrogens results in the formation of perinuclear clusters of ER-ligand complex [190]. Thus, the activation of ERα or ERβ can produce both unique and overlapping effects.

The natural ligands of ERα and ERβ includes estrogenic compounds including estradiol, estrone, and estriol, but relative to ERα, ERβ binds estriol and ring B unsaturated estrogens with higher affinity [191]. Phytoestrogens present in food that mimic estrogens in the body, have greater affinity for ERβ than ERα. For example, soybean isoflavones (genistein, daidzein, and biochanin A) are ERβ-selective agonists of transcriptional activation or repression, creating more stable the complexes with lower binding free energy [192,193].

ERα and ERβ transcriptional activity

ERα and ERβ exhibit multiple modes of action that influence their regulatory impact on gene expression. Upon ligand binding, these receptors translocate to the nucleus where they dimerize and engage with EREs in the DNA of target genes. This interaction initiates the transcriptional processes essential for executing estrogen's effects in cells [194]. Both receptors can directly bind to EREs, yet they display distinct binding dynamics and transcriptional outcomes [195]. The complexity of ERα and ERβ interaction with DNA is highlighted by the fact that EREs exist in multiple conformations. These include consensus EREs, characterized by palindromic ERE repeats, non-consensus EREs, which deviate from the typical sequence, single binding site EREs, multiple binding site EREs, and composite ERE sites which involve combinations of these elements [196].

ERα and ERβ compete for binding to these ERE sites. However, compared to ERα, ERβ has reduced binding affinity for non-consensus EREs, which constitute the majority of EREs and include those associated with genes such as C-fos, c-jun, pS2, and cathepsin D. Additionally, as shown in Figure 1.7, both ERs can interact with DNA indirectly by tethering to other DNA-bound transcription factors (TFs) such as AP-1, SP1, or NF-κB, as revealed by genome-wide mapping of ERβ binding sites that have extensive overlap with AP-1 binding sites [197]. Other non-genomic estrogen-dependent pathway signals are triggered by ligand binding to membrane ERs and mediated through signaling cascades (Akt, PKA, and ERK1/2) to activate STAT, CREB, NF-kB, and Jun TFs.

In a ligand-independent action in the absence of estrogen, ERβ can also activate expression of a unique sets of genes activating the pathways related to growth factor signaling via activated kinases (Figure 1.7). In an in vitro study, it was observed that the main difference in the specific activation of survival pathways mediated by ERα and apoptotic pathways mediated by ERβ is through their non-genomic membrane-initiated signaling [198]. Another in vitro study highlighted that while estrogen signaling through ERα promotes cell cycle entry, ERβ temporarily hinders it. The absence of ERβ expression is linked to increased cell transformation, suggesting distinct roles of ERα and ERβ in cellular processes [199]. Clinical studies on breast carcinomas align with in vitro findings, showing that whenever there is a shift in the ERα/ERβ ratio: ERα is often upregulated while ERβ is downregulated [200].

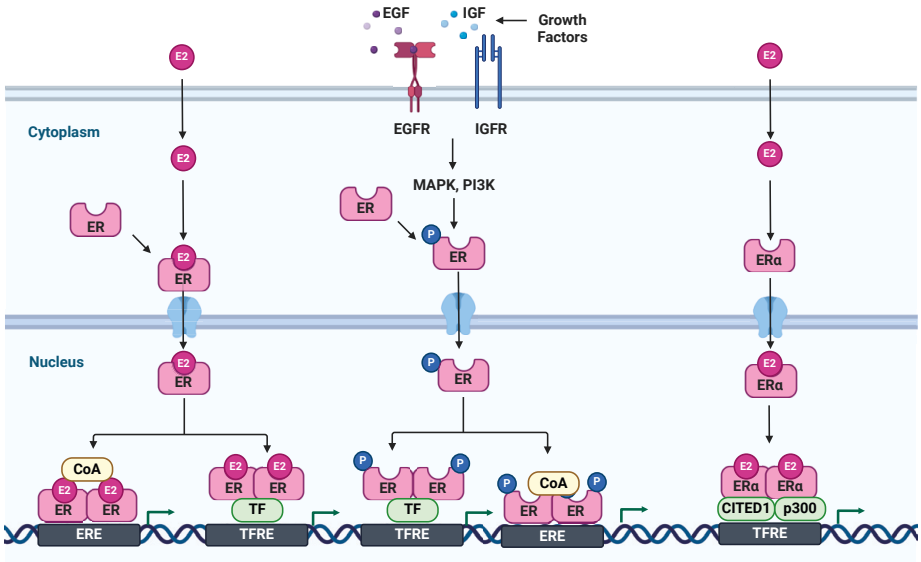

Figure 1.7: The estrogen Receptor (ER) signaling pathway. ER and CITED1 mechanisms for transcription activation are shown. After activation by ligand or upon phosphorylation, ER translocates to the nucleus, dimerizes and binds to ERE response elements either directly or in conjunction with transcriptional co-factor complexes. CITED1 can selectively coactivate the estrogen-dependent ER-mediated transcription of AREG and TGF-α. Abbreviations: CITED1: CBP/p300-interacting transactivator with Asp/Glu-rich C-terminal domain 1; CoA: co-activator; E2: estradiol; EGFR: epidermal growth factor receptor; ER: estrogen receptor; ERE: estrogen response element; IGFR: insulin like growth factor receptor; P: phosphoryl group; TF: transcription factor; TFRE: transcription factor response element. Created with BioRender.com.

Expression of ERs

In normal beast tissues, ERα is expressed in luminal epithelial cells but not in the stroma, whereas ERβ is present in luminal, myoepithelial, and stromal cells [201]. Though at lower levels than those found in reproductive tissues, ERα and ERβ are both expressed in other cell types as well including lymphocyte precursors [202], T and B cells [203], and in the brain [204].

In primary BC, tumors are classified as ER-positive or ER-negative based on the expression of ERα. At diagnosis, approximately 75% of BCs are ERα positive. [205], ERα expression defines luminal-type tumors, whereas basal-like BC are ER-negative [206]. ERα-positive tumors indicate a better prognosis due to their responsiveness to endocrine therapies and CDK4/6 inhibitors [207]. In contrast, ERα-negative tumors tend to be more aggressive and progress more rapidly [208]. Intriguingly, ERα expression is dynamic; approximately 50% of

ERα-positive breast cancers that recur after endocrine therapy lose expression of ERα [209].

The role of ERβ expression in breast cancer has been a subject of considerable debate and investigation, with conflicting reports regarding its prognostic value. For instance, some studies have suggested that high ERβ expression, regardless of ERα status, serves as a marker of positive treatment response in breast cancer patients undergoing chemotherapy [210–212] and endocrine therapy [210,213–215]. Conversely, other groups have shown that increased ERβ expression is associated with poor prognosis and reduced disease-free survival (DFS) in patients receiving endocrine therapy in postmenopausal primary BC [216,217] and reduced DFS in TNBC patients [215,218]. However, another study found no significant relationship between ERβ expression and patient outcomes, further complicating the narrative around ERβ's role in breast cancer [219].

In **Study I** of this thesis, large-scale transcriptomic analyses were performed that has demonstrated that *ESR2* (ERβ) is expressed at higher levels in ERα-negative breast tumors (basal-like subtype) and that *ESR2* showed a slight inverse correlation with *ESR1* expression. Also, *ESR2*-high tumors had favorable overall survival (OS), more pronounced in subgroups receiving endocrine therapy and TNBC [220]. These findings align with two other independent studies by Austin et al., [218] and Dey et al. [221] which further confirmed ERβ's role as a potential target for therapy in TNBC using TMA and in vivo studies, respectively.

Gene regulation of ERβ

Transcription factors that have basic-helix-loop-helix (bHLH) protein structural motifs bind to the evolutionally conserved E-box motif of the *ESR2* promoter, enhancing transcription of downstream genes. Many TFs, including c-jun, CREB, AP2α and AP2γ (AP2α and AP2γ also regulates *ESR1* transcription) binds to the *ESR2* promoter and regulates *ESR2* transcription [222,223]. ERβ, unlike ERα, was also shown to be modulated by circadian clock proteins, and the expression levels of these circadian regulators significantly impact ER signaling by controlling the intracellular concentrations of endogenous ERβ [224]. *ESR2* expression is also regulated by promoter methylation and histone modifications on the *ESR2* promoters known as 0K and 0N, controlling its transcription [225].

ERβ alternative splicing

To date, five experimentally validated functional ERβ isoforms termed ERβ1-ERβ5 have been described, which may be attributed to frame-shift mutations or alternative splicing of last coding exons, and found to be expressed in both normal and cancerous tissues [226]. The first cloned wild-type *ESR2* transcript, termed ERβ1, is the only isoform that retains full transcriptional activity. For the remaining isoforms, exons 1 to 7 are the same and only exon 8 is specific to a given isoform. These isoforms yield proteins of a lower molecular weight than the wild-type protein [227]. ERβ isoforms have different expression levels in different tissues, different associations with tumor characteristics, prognosis, and purportedly different roles in tumorigenesis. For example, *ERβ1* (which is abundantly expressed in most normal breast epithelial cells) was thought to be present in 20-30% of breast cancers [228]. In other research, ERβ1 expression by IHC in breast cancer was reported to be present in over 60% of cases [229,230]. ERβ4 and ERβ5, which do not bind estrogen but can bind to DNA, are expressed in the testis and placenta, respectively. ERβ2, ERβ4, and ERβ5 can heterodimerize with ERα and negatively regulate its transactivation activity [227,231]. ERβ1 overexpression predicts favorable prognosis and better survival in tamoxifen-treated BC patients [232] as well as TNBC [233]. ERβ2 was found to be associated with poor OS in BC patients [234] and ERβ5 was associated with poor outcome in HER2+ and TNBC [235]. Absent or decreased transcription of *ERβ1*, *ERβ2*, and *ERβ4* in some breast, ovarian, and prostate cancer tissues and cell lines may be attributed to methylation of CpG sites in the promoter 0N [236]. Moreover, it has been reported that miR-92 can reduce ERβ1 expression by direct targeting of the 3′-UTR of *ESR2* mRNA [237].

Clinical significance and prognostic value of ERβ in breast cancer

The exact role of ERβ in BC is controversial and was one subject that we addressed in this thesis work. ERβ expression has been reported in ERα-negative subtypes (basal and TNBC) [238,220], and majority of data from BC cell lines and clinical samples suggests that ERβ might have antiproliferative, pro-apoptotic, and tumor-suppressive functions [200,239–241]. In ERα-positive BC, some reports suggest that ERβ acts mainly as a tumor-suppressor [242,243].

In addition to studies stated above for ERβ transcripts and correlations to BC patient prognosis, it was also found that ERβ expression might be an independent predictor of response to tamoxifen in tamoxifen treated BC patients in adjuvant setting [244]. In another study, it was shown that ERβ regulates gene expression, proliferation, increases apoptosis in BC by

attenuating ERα-induced cell proliferation, reversing transcriptional activation and repression by ERα. The mechanism was suggested to be related to how ERα functionally incapacitates p53's tumor-suppressive actions by recruiting nuclear receptor corepressor (N-CoR) to repress p53-mediated transcriptional activation and prevent p53-dependent apoptosis (despite harboring wild-type p53). To reduce and destabilize this ERα-p53 binding, ERβ physically interacts with p53 and antagonizes ERα-p53-mediated transcriptional regulation, and activates ERα-repressed genes, supporting the notion that ERβ is a novel activator of the p53 pathway and can act as a tumor suppressor in BC cells [245]. In ERα-negative BC, studies have also supported the tumor-suppressive role of ERβ [218,246]

A multivariate analysis of 442 invasive BCs treated with adjuvant tamoxifen revealed that ERβ1 expression was associated with significantly better survival in patients with basal-like and TNBC tumors [234]. This study also showed that ERβ1 status was significantly associated with survival in postmenopausal, but not premenopausal, women, indicating an emerging role of ERβ as independent predictor of recurrence and mortality. Wang et al. showed that ERβ1 protein expression with negative pAKT predicted favorable prognosis in TNBC. The PI3K/AKT pathway, commonly activated in BC, is regulated by PTEN and controls proliferation, invasion, apoptosis, and hypoxia-related proteins. Increasing PTEN levels can reduce Akt signaling, increasing ERβ1 expression and blocking HER2/HER3 signaling in BC cells. Conversely, activation of PI3K/AKT leads to ERβ ubiquitination and degradation, influenced by Akt's interaction with MDM2 and CBP, promoting ERβ polyubiquitination [233]. Another study supports these results showing ERβ to be an independent marker for favorable prognosis after adjuvant tamoxifen treatment in ERα-negative BC tumors [214]. ERβ exerts its anti-proliferative role by reducing cell migration, invasion, and angiogenesis by inhibiting HIF-1α-mediated transcription by downregulating aryl hydrocarbon receptor nuclear translocator, ARNT, which further reduces the expression of VEGF, leading to the reduction of the number of intertumoral blood vessels [247,248]. These findings are further supported by a study where patients with ERα-negative BC, expressing high ERβ1 and ERβ2 protein, had significantly improved DFS and OS [249]. ERβ expression has also shown to inhibit cell growth by inducing G1 cell cycle arrest, which was further enhanced by 17β-estradiol treatment in TNBC cell lines [250].

Thus, a series of mounting evidence points to ERβ as having a tumor suppressive, pro-apoptotic and anti-proliferative role in BC, and its loss promoting breast carcinogenesis. In the future, ERβ may have clinical value regarding prediction, prognosis, and as a possible therapeutic target. Conversely, some studies have also reported that ERβ is not associated with

recurrence or survival in basal-like and TNBC [251]. The observed differences in the role of ERβ across various studies could be attributed to multiple factors. These include the co-expression of ERβ with its binding partners, the isolated expression of ERβ or its co-expression with ERα, the presence of different ERβ isoforms, the use of poorly specific antibodies, and variations in experimental approaches such as ectopic expression and gene knockdown [252,253]. Regarding gene knockdown, CRISPR/Cas genome editing technology presents a reliable alternative to traditional methods, helping to minimize artifacts and incidental effects. This technique was employed in a recent study to reassess the role of ERβ by creating an ERβ knockout mouse model. Observations from this model, which included the development of in situ ductal cancer in both the prostate and mammary gland, confirmed ERβ's oncosuppressive role [254]. To ensure the accuracy and reliability of research findings on ERβ's role in cancer, it is crucial to develop and utilize highly selective ERβ antibodies which should then be rigorously evaluated across a broad spectrum of well-characterized cancer samples. This approach will help clarify the discrepancies observed in previous studies using IHC and provide a more definitive understanding of ERβ's function in cancer biology. Thus, continued research is essential to precisely understand ERβ's roles, which could enhance diagnostic clarity and treatment for certain breast cancer subtypes such as TNBC, offering new management strategies.

1.5.2 CITED1

CITED1 is a transcriptional co-regulator of the CBP/p300-interacting transactivator with Asp/Glu-rich C-terminal domain (CITED) family of proteins [255]. It was originally named melanocyte-specific gene 1 (MSG1) following its discovery in highly pigmented murine melanocytes [256]. MSG1 expression was also identified in mammary epithelium, testes, the brain, embryonic tissues, and various tumors, leading to a broader understanding of its biological characteristics [257–259]. As a result of these findings and its expanded role, MSG1 was subsequently renamed “CITED1”. The *CITED1* gene is located on chromosome Xq13.1 and is highly conserved, comprising three exons and two introns. Its promoter region includes a TATA box and potential binding sites for various TFs such as USF, Brn-3, Brn-2, TFE3, Oct-1, AP-2, and Sp1. This region is crucial for initiating transcription of the *CITED1* gene, which translates into a 27-kDa nuclear protein [256]. CITED1 shares a conserved transcriptional activation region, CR2 (145–193 aa; Figure 1.8), with CITED2, CITED3, and CITED4 [260]. The CR2 region is rich in aspartic and glutamic acid residues which enhances its transcriptional activity by facilitating binding to the co-activators CBP/p300 as depicted in Figure 1.8 [255,256]. Several

proteins including heat shock cognate protein 70, hypoxia-inducible factor-1, β-catenin, and microphthalmia-associated transcription factor (MITF) can inhibit this interaction by competing for CBP/p300, thus suppressing transcription. Additionally, phosphorylation of CITED1 at five serine residues during the M-phase of the cell cycle (Ser16, Ser63, Ser67, Ser71, and Ser137) diminishes its ability to activate transcription by reducing its binding affinity for CBP/p300 [261,262].

As CITED1 lacks DNA-binding ability, it functions primarily as a facilitator that enhances the interactions between CBP/p300 and certain DNA-binding proteins such as SMAD4, ERα, and others (protein domains are shown in Figure 1.8). These complexes then translocate to the nucleus, bind to the promoters of specific genes, and amplify transcription. Notably, CITED1 interacts with the TF SMAD4 via its N-terminal SMAD4-interacting domain (SID; aa 30-60), enhancing TGF-β/BMP-induced transcription. It also interacts with ERα via its N-terminal CR2 region (aa 157–158), a connection that does not depend on its interaction with CBP/p300, thus supporting the coactivation of estrogen-dependent transcription [261].

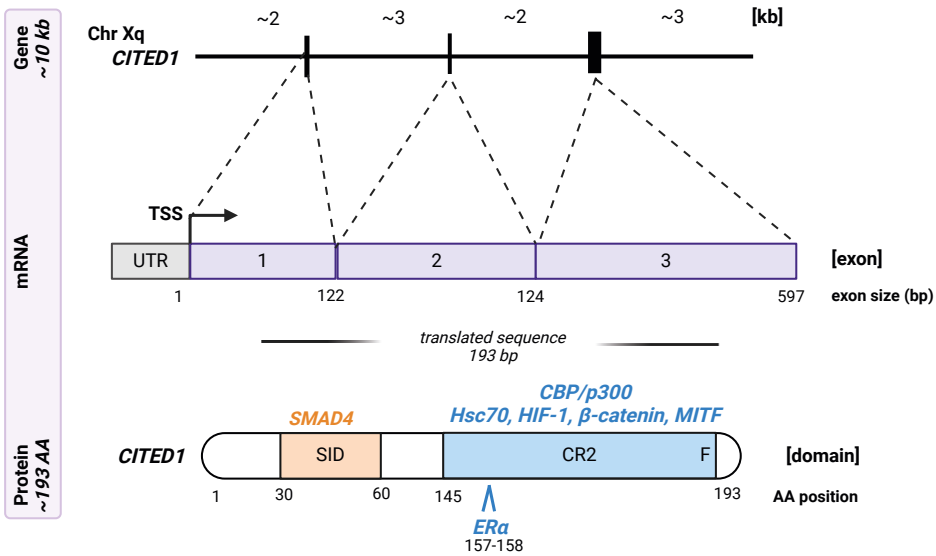

Figure 1.8: Schematic representation of *CITED1* gene. Mapping of genomic organization of mRNA and protein sequence are illustrated, and the protein domains are marked. Abbreviations: SID: SMAD4-interacting domain; CR2: conserved transcriptional activation region; TSS: transcription start site. Created with BioRender.com.

Role of CITED1 in estrogen signaling

CITED1 promotes the growth of mammary epithelium and the formation of mammary ducts by activating the transcription of amphiregulin (AREG) via the estrogen signaling pathway during pubertal mammary gland development (Figure 1.7) [263,258]. CITED1 can also coactivate the ER-mediated transcription of TGF-α in an estrogen-dependent manner, promoting the growth of MCF-7 cells in an autocrine manner [261]. CITED1 was also shown to enhance transcriptional activation of both ERα and ERβ in an estrogen-dependent manner, and functions as a selective coactivator for estrogen-dependent transcription in mammary epithelial cells [258,261].

In **Study III** of this thesis, *CITED1* mRNA and protein expression was analyzed using GOBO, TMA, and TCGA datasets. Notably, *CITED1* mRNA was found to be expressed in ER+ luminal breast cancer subtype and higher *CITED1* was found to be correlated with tamoxifen response, suggesting its role as a marker of better prognosis and favorable outcome in anti-endocrine treated, ER+, lymph-node negative BC patients. CITED1 protein expression was validated using TMA and was found to be associated with favorable outcome in ER+, tamoxifen-treated patients. It was also found that MCF-7 cells overexpressing CITED1 exhibit selective amplification of AREG but not TGF-α, and the maintenance of specific ERα/CITED1-mediated transcription is a good prognostic marker in patients with anti-endocrine-treated ER+/lymph node negative (LN-) breast cancer [264].

1.5.3 HER2 signaling

Human epidermal growth factor receptor 2 (HER2), also known as ErbB-2 or Neu, is one of the four members of the epidermal growth factor receptors with tyrosine kinase (TK) activity [265,266]. It is a proto-oncogene encoded by the *ERBB2* gene located on chromosome 17q12, was found to be amplified in a human BC cell line in 1980s, and has since then has been extensively studied in BC [267]. *ERBB2* mRNA is composed of 27 exons. The first exon codes for 30 aa, where the first 22 aa constitute a signal peptide which is cleaved during the migration of mature *ERBB2* mRNA to the cell membrane [268]. Full-length HER2 is a 185 kDa transmembrane receptor, a total 1255 aa in length, which is composed of a 620 aa extracellular domain, followed by a 23 aa transmembrane domain and a 490 aa intracellular domain with a tyrosine kinase activity as shown in Figure 1.9 [269,270]. HER2 is the only EGFR family member for which no ligand has been found, which can be explained by the unique structure of its extracellular domain, which is not favorable for ligand binding [271]. Yet, it is the preferred binding partner of all *ERBB* receptors

even as a monomer, because its extracellular domain is always in the open conformation [272]. It is now known that *ERBB2* gene amplification and HER2 protein overexpression is observed in 15-30% of breast cancers [273]. Out of the four receptors, all of which are transmembrane proteins with an intracellular TK domain, HER2 is the only one that does not bind to a ligand. HER2 overexpression and homo- and/or heterodimerization of this receptor with other EGFR family members leads to autophosphorylation of tyrosine residues initiating a cascade of signaling pathways, namely PI3K and MAPK signaling, which confer strong proliferative and tumor growth advantages [274]. *ERBB2* amplification and overexpression correlate with aneuploidy, lymph node metastasis, grade and size of tumor, and was found to be a significant predictor of patient outcome, poor prognosis (both overall survival and time to relapse), as well as therapy response in BC patients [265,275].

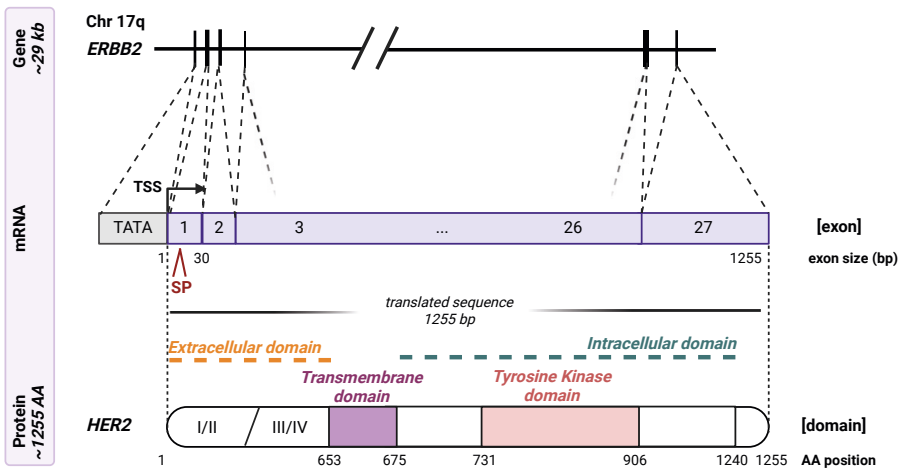

Figure 1.9: Schematic representation of *ERBB2* gene. Mapping of genomic organization of mRNA and protein sequence are illustrated, and the protein domains are marked. Abbreviations: SP: signal peptide; TSS: transcription start site. Created with BioRender.com.

HER2 signaling pathway is a complex network comprising various membrane receptors, their ligands, protein kinases and regulating genes, regulating various cellular functions such as cell proliferation, cell cycle progression, differentiation, migration, cell adhesion, and cell death [276]. As shown in Figure 1.10, four TK receptors, HER1 to HER4, and their multiple ligands form an input layer, exerting their effect via their TK activities upon ligand binding and subsequent dimerization (homo- in the case of HER2-to-HER2, and hetero-

for HER2-to-HER1/3/4). Upon dimerization, tyrosine residues on the intracellular domain of the receptors are auto- or trans-phosphorylated, serving as docking sites for various signaling molecules [277,278]. The most common downstream effectors of HER signaling include MAPK [279], the PI3K/Akt signaling pathway [278], and protein kinase C (PKC) activation (Figure 1.10). The signaling pathways activated in response depends on the ligand involved, the dimer formed as well as on cellular context. HER2 and HER3 are considered the preferred dimerization partners in the HER signaling pathway because HER2, despite having no known ligands, exhibits the strongest catalytic kinase activity among all family members. Conversely, HER3 lacks intrinsic TK activity, yet it can be activated upon binding with ligands such as heregulins [274,280].

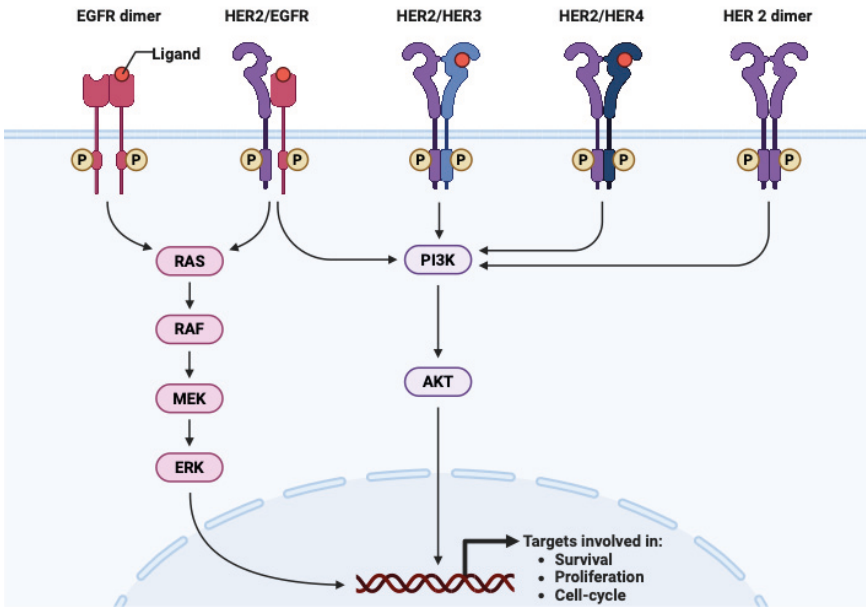

Figure 1.10: HER2 Signaling pathway. HER2 as well as the other members of the EGFR family which are located on the cell membrane and responds to a wide variety of ligands. Phosphorylation of the tyrosine kinase domain in the cytoplasm initiates downstream oncogenic signaling pathways such as PI3K/AKT pathway and Ras/MAPK pathway. Created with BioRender.com.

ERBB2 copy number

As discussed previously in section 1.2, gene amplification refers to an increase in the copy number of a specific chromosomal region and is commonly linked to overexpression of the affected gene(s). The *ERBB2* gene can be amplified up

to 25-50 or more copies in a BC genome, and there can be up to 40-100 fold increase in HER2 protein which corresponds to more than 2 million receptors expressed at the cell surface [281].

ERBB2 amplification is generally associated with a more aggressive form of breast cancer, and most in situ cancers with HER2 amplification are at a higher risk of progressing to invasive carcinoma [282,283]. While treatments with HER2-targeted therapies such as trastuzumab (discussed later) can lead to loss of HER2 expression, possibly due to selection of HER2-negative clones [284], it is also possible for HER2-negative tumors to become HER2-positive over time particularly post-ER-targeted therapy [285]. The observed changes in HER2 and ER expression may be due to their inverse relationship, where the regulation of one can influence the other. Specifically, blocking ER can lead to the upregulation of HER2, and conversely, targeting HER2 may induce ER expression. This interaction suggests a compensatory mechanism between the two pathways, as demonstrated in several studies [286,287]. *ERBB2*-amplified BC exhibit unique biological characteristics and clinical responses. These tumors typically show increased proliferation rates, a high histologic grade, and varied levels of ER and PR. They also demonstrate high aneuploidy, a propensity to metastasize to the CNS and viscera, and relative resistance to endocrine therapy, increased sensitivity to cytotoxic agents such as doxorubicin, and often present co-amplification of topoisomerase 2. Importantly, they respond well to HER2-targeted therapies [288,289]. The correct assessment of *ERBB2* CNA is of prime importance in BC clinically, as it further affects patient management and treatment decisions. This is the topic which is addressed in **Study II**.

Clinical assessment of HER2 status

To assess HER2 overexpression and/or *ERBB2* amplification, the updated 2018 American Society of Clinical Oncology and College of American Pathologists (ASCO/CAP) HER2 guidelines are followed in Sweden. Immunohistochemistry (IHC) and fluorescence or silver in situ hybridization (FISH or SISH) are the recommended methods for analyzing tumor tissues. These techniques, however, come with various challenges, as will be discussed later. HER2 expression can be scored as 0, 1+ (or 0/1+), 2+, and 3+ by IHC.

Briefly, BC is considered HER2 positive when there is evidence of HER2 overexpression with IHC score 3+ or gene amplification on ISH assay in at least one tumor sample, according to recent 2018 ASCO/CAP guidelines (Figure 1.11 and Table 1.1) [290–292]. In the case of IHC score 2+, reflex ISH testing is required to define HER2 status, with additional concomitant review of IHC slides in case of particular ISH results from groups 2, 3, and 4. In case of IHC 0

and 1+ or IHC 2+ with a negative ISH assay, the cancer is considered HER2 negative, and no HER2-targeted therapy is recommended, with the exception of dual-probe ISH testing group 3 results (HER2/CEP17 ratio < 2.0 with average *ERBB2* CN of 4.0 to 5.9 per cell) where tumors that are IHC 2+ are deemed HER2 positive (Figure 1.11) [290].

With IHC, results may vary with the use of lab-specific protocols including different fixation methods, fixation times, antigen retrieval solutions and methods, primary and secondary antibodies, as well as variation due to the subjective qualities of the observer. Altogether this leads to considerable issues with reproducibility of scoring. On the other hand, ISH is more labor intensive, time-consuming, and expensive, whereas the signals are easier to interpret. Both methods are affected by intra-tumoral heterogeneity, adding to variability in evaluation.

Table 1.1: 2018 ASCO/CAP HER2 dual ISH clinical subgroups and final determination based on integration with IHC.

Group	Description	HER2/CEP17 ratio	Mean HER2 CN	2018 ASCO/CAP recommendation
1	Classic HER2 amplified cancer	≥2.0	≥4.0	Positive
2	Monosomy 17	≥2.0	<4.0	Negative, unless concurrent IHC 3+
3	Co-amplification, previously polysomy 17	<2.0	≥6.0	Negative, unless IHC concurrent 2+ or 3+
4	Borderline/equivocal	<2.0	≥4.0 and <6.0	Negative, unless concurrent IHC 3+
5	Classic HER2 non-amplified cancer	<2.0	<4.0	Negative

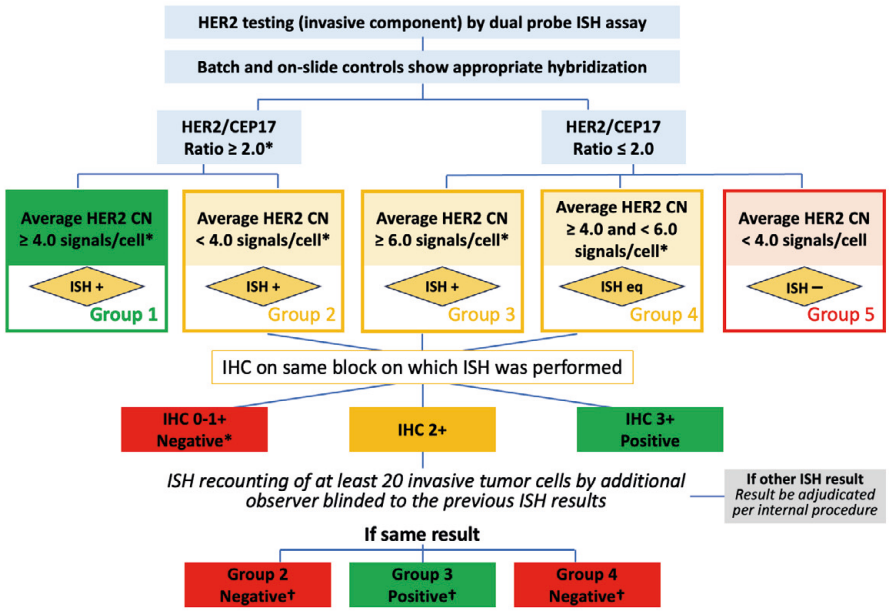

Figure 1.11: ASCO/CAP 2018 ISH determination algorithm. *Observed in a homogeneous and contiguous population. †An explanatory comment should be provided. Figure modified from [290,293]. Reprinted with permission.

Influence of polysomy 17

True polysomy is defined by the presence of extra copies of one or more whole chromosomes, and in case of chromosome 17, it provides an alternative mechanism for increasing HER2 gene dosage [294]. Chromosome enumeration probe 17 (*CEP17*) is a locus within the chromosome 17 centromere and is used to determine *ERBB2*/*CEP17* ratio in ISH experiments. However, according to recent studies, true polysomy of chromosome 17 is rare in breast cancer and focal amplifications encompassing the centromere can increase the number of CEP17 signals in ISH testing [295]. Increase in *CEP17* copies can alter the *ERBB2*/*CEP17* ratio and subsequently influence the interpretation of the final HER2 ISH result. Currently, the commonly adopted threshold for polysomy 17 is a mean of ≥3 *CEP17* signals per nucleus [290].

ERBB2 SNPs

According to the NCBI dbSNP database (accessed on 2024-01-18), there are more than 20,000 SNPs within *ERBB2*. Most of the studies on *ERBB2* SNPs have focused on two coding SNPs, rs1058808 (Pro1170Ala) and rs1136200

(Ile655Val). Some studies have found that HER2 overexpression is more common in proline carriers of rs1058808 SNP [296], while others reported it can identify a subset of HER2+ BC patients who are at increased risk of cardiotoxicity from trastuzumab therapy [297]. SNPs rs1136201and rs1058808 were also found to be significantly associated with the susceptibility of cervical cancer [298]. rs1058808 was also proposed as potential biomarker of a good response to anti-HER2 treatment in patients with early HER2+ BC [299]. rs1136200 might be a susceptibility factor that favors early-onset BC and was found to be strongly associated with BC susceptibility in the young female population [300]. In **Study II**, as an exploratory analysis the SNP rs1058808 was genotyped as part of the custom dPCR assay, but no association between rs1058808 genotype and clinical HER2 status or outcome was found [301]. In **Study V**, we quantified *ERBB2* mRNA expression levels and analyzed two alleles of the SNP rs1136201 using dPCR.

HER2 targeted therapies

The importance of HER2 status as a prognostic marker and drug target in various cancer types, in particular in BC, lead to the discovery and development of several FDA-approved therapeutic mAb such as trastuzumab (Herceptin), pertuzumab (Perjeta), and trastuzumab emtansine (Kadcyla) [302–306]. These mAbs exert therapeutic effects partly by blocking downstream signaling through targeting the HER2 receptor-binding regions. This interaction leads to several cellular responses, including G1 arrest, induction of apoptosis, and inhibition of cell growth and migration [307]. Additionally, these mAbs facilitate the internalization and degradation of the HER2 receptor. Moreover, they can also engage immune mechanisms such as antibody-dependent cellular cytotoxicity (ADCC), enhancing their efficacy against cancer cells [308]. Small molecule inhibitors of HER2 kinase such as lapatinib, neratinib, and tucatinib [309–311] also block downstream signaling by preventing phosphorylation of HER2 receptor by blocking ATP-binding regions. Another class of anti-HER2-targeted agents known as antibody-drug conjugates (ADCs) are among the fastest growing drug class that combines a humanized mAb, mainly immunoglobulin G (IgG), bound to a cytotoxic agent, called the payload, via a molecular linker [312–314].

The first HER2-targeting ADC approved for the treatment of advanced HER2-positive BC was trastuzumab emtansine (T-DM1; Kadcycla) [305,315]. The second FDA-approved ADC was trastuzumab deruxtecan (T-DXd; Enhertu) given as second line of treatment in advanced BC after at least two prior lines of HER2-targeting therapy [316–318]. Recently, T-DXd has become the first

FDA approved tumor-agnostic HER2-directed therapy for previously treated patients with unresectable or metastatic HER2-positive solid tumors [319].

HER2-low BC

Across BC, HER2 expression levels are on a continuous spectrum from negligible expression to, in the case of amplified cases, very high overexpression. Until recently, only the amplified case was viewed as a therapeutic target. However, quite recently, a new subgroup called "HER-low" has been identified. HER2-low is defined as IHC 1+ or IHC 2+ and ISH not-amplified. Approximately 50% of BC are reported to be HER2-low, according to current scoring systems [318,320,321]. In standard clinical practice, these tumors are reported as HER2-negative, either TNBC or luminal-like, depending on the absence or expression of hormone receptors [320]. The HER2-low definition has gained relevance because these tumors with low levels of HER2 expression and no detectable *ERBB2* amplification were shown to derive survival benefit from targeting HER2 with T-DXd, which improved progression-free survival (PFS) and OS compared with chemotherapy in patients with pretreated HER2-low metastatic BC in the phase III DESTINY-Breast04 trial [322]. Due to these recent therapeutic implications, which may also become relevant for primary BC in the future, the proper assessment of HER2-low status with high accuracy and sensitivity is of high importance. This issue, to categorize HER2-zero, HER2-low, and HER2-positive subgroups using absolute quantification techniques, is addressed in **Study V** of this thesis.

Trastuzumab resistance

In the neoadjuvant treatment of primary BC, the addition of trastuzumab to neoadjuvant chemotherapy is associated with a pathological complete response (pCR) of the breast and lymph nodes in 38-55% of patients [323]. However, despite initial responses, some patients after surgery eventually experience disease progression, a phenomenon known as acquired HER2 resistance. This resistance to trastuzumab can develop at any point during treatment and is a significant contributor to the failure of anti-HER2 therapies [324].

In recent years, several resistance mechanisms have been identified, such as disrupted trastuzumab-HER2 receptor binding [325] and *ERBB2* gene mutations in kinase activation (intracellular domain) region that impairs the ability of the mAb to bind to HER2 [326]. HER2 heterogeneity, currently clinically defined as the presence of any aggregate population of amplified cells comprising >10% of the tumor cells on the slide (not scattered single cells in a

mosaic pattern) also affects trastuzumab responses and survival in patients with HER2-positive primary or metastatic BC [327]. Constitutive activation of signaling pathways parallel or downstream of HER2 such as PI3K/Akt/mTOR pathways are also implicated in acquired resistance to HER2-targeted therapy [328], as well as metabolic reprogramming or reduced immune system activation [329]. PTEN loss and *PIK3CA* mutation in PI3K/Akt/mTOR signaling were shown to be the cause of de novo resistance to HER2-targeted therapy in preclinical and clinical studies [330]. Adaptation of HER2-targeting ADCs that incorporate technological advances in the antibody, linker and/or payload conception have shown promising activity in preclinical and clinical studies and some of them are now being evaluated in larger clinical trials [331]. It is crucial to reassess the mechanisms of resistance to anti-HER2 therapies, especially in the context of combination treatments. This is particularly important given reports from the recent DAISY trial that resistance to antibody-drug conjugates like T-DXd is emerging and thus, further understanding of these resistance patterns are essential for optimizing treatment strategies [332]. An ongoing challenge is how to adapt the treatment algorithms and their sequence to include all these recently approved and emerging agents and combination approaches.

1.5.4 Progesterone receptors

In addition to ER and HER2, there are additional clinically relevant biomarkers that affect cancer biology, tumor cell phenotype, patient prognosis, and treatment efficacy. This includes progesterone receptor (PR), which has two predominant isoforms, PR-A and PR-B, both of which are encoded from the same gene *PGR* located on chromosome 11 but from two different promoters. PR-B is a full-length receptor, whereas PR-A is a truncated form lacking 164 amino acids at the N-terminus of the protein [333]. Being structurally different, they also exhibit different transcriptional and biological activities as members of steroid nuclear receptor family ligand-activated TFs. Breast tissues are a main target of progesterone, where it regulates development of the branched ductal epithelium and expansion of milk-secreting alveoli during lactation. Progesterone is produced in a cyclical manner in the ovaries, but also in peripheral tissues including adrenal glands, nervous system and brain in premenopausal women [334,335]. While circulating progesterone levels decline sharply at menopause, local tissue production continues is unknown. As a key target of ERα, ligand activated PR presence indicates ER activation and interplays with ER in BC [336]. Expression of PR is an established prognostic factor associated with improved outcomes in particular in patients with ER-

positive BC [337]. It also plays a determining factor in treatment decision making in ER-positive BC patients [338]. As such, PR along with ERα and HER2, is part of the routine pathological assessment of every breast cancer.

1.5.5 Intrinsic subtypes

As first described in 2000 by Perou and colleagues, utilizing DNA microarrays representing more than 8000 genes to analyze 65 breast tumor surgical specimens from 42 patients and 17 cultured cell lines, the so-called "intrinsic subtypes" of BC were defined. Using identified gene expression patterns, tumors were classified into five intrinsic subtypes with distinct clinical outcomes, i.e., luminal A, luminal B, HER2-enriched, basal-like, and normal-like subtypes [339,340]. The rationale behind this classification is that the differences underlying the gene expression patterns among cancer subtypes reflects the fundamental differences of the tumors at the molecular level. Towards a clinical assay, in 2009 Perou and colleagues developed a simpler predictor, where the number of intrinsic genes were reduced from about 2000 to only 50, termed the PAM50 genes (utilizing a method called Prediction Analysis of Microarray) [341].

Table 1.2: Intrinsic subtypes of breast cancer, their proportions, biological profile, therapy used for their management.

Subtype	Proportion	Biological Profile	Tumor Grade	Therapy of Choice
Luminal A	60-70%	ERα+ and/or PR+, HER2−, Ki67 low	1-2	Endocrine therapy
Luminal B	10-20%	ERα+ and/or PR+, HER2+/−, Ki67 high	2-3	ET +/− chemo
HER2-enriched	15-30%	ERα+/−, PR+/−, HER2+	2-3	Anti-HER2 therapy
TNBC / Basal-like	10-20%	ERα−, PR−, HER2−, + basal markers	3	Chemotherapy
Normal-like	5-20%	[ER+|PR+] HER2−, Ki67−	1-3	

Summarized in Table 1.2, luminal A and B are mainly composed of ER-positive tumors. Luminal A tumors are characterized by the presence of ER and/or PR and the absence of HER2 and have a low expression of cell proliferation marker Ki67 (less than 20%). Clinically these tumors are low grade, slow growing, and have the best prognosis with less incidence of relapse and higher survival rate. Luminal B tumors are ER positive and can be PR negative and have a high expression of Ki67, higher grade and worse prognosis compared to luminal A (Figure 1.12). These tumors may benefit from chemotherapy in addition to hormonal therapy. Luminal B tumors are characterized by more rapid growth, elevated Ki67, and generally worse prognosis than luminal A [342]. HER2-enriched breast cancer is characterized by high HER2 expression and can occur with or without the presence of ER and PR. They are more aggressive and fast-growing than luminal tumors, however the prognosis has improved dramatically after the introduction of HER2-targeted therapies. They can be further classified as luminal HER2 (ER+, PR+, HER2+ and Ki67 15–30%) and HER2-enriched (HER2+, ER-, PR-, Ki67 >30%) [343]. Triple-negative BC (TNBC), as the name implies, do not express the three most important BC biomarkers: ER-negative, PR-negative, and HER2-negative. Constituting 10-20% of all BC, TNBC are poorly differentiated, highly proliferative, highly aggressive, have early relapse, a greater tendency to present at advanced stages, and frequently harbor aberrations in DNA repair genes and exhibit increased genomic instability. Approximately 80% of mutated BRCA1 BC belongs to this group. TNBC can be further divided into 7 subgroups [344], although the clinical utility of these subgroups has not been established. Generally, depending on the surrogate algorithm being employed, there is a 70-80% concordance between the gene expression-based molecular subtypes and IHC-based clinical subtypes [345,346].

After some time, at least six commercial genomic assays were developed for the prediction of clinical outcome for BC patients. In 2007, a 70-gene signature assay MammaPrint became the first FDA-cleared prognostic signature assay [347]. Another well-known FDA-approved assay named OncotypeDX was developed which uses qPCR to measure the expression of 21 genes (including 16 cancer-related and 5 reference genes) to predict risk of recurrence and treatment response [348]. Other tests include the PAM50-based Prosigna Risk of Recurrence (ROR) score prediction assay (NanoString Technologies) [349], EndoPredict (EP; Myriad Genetics) [350] and the Breast Cancer Index test (BCI; Biotheranostics) [351]. Within the SCAN-B study, single- and multi-gene classifiers and single-sample predictor (SSP) models were developed for clinical markers, subtypes, and ROR prediction from RNA-seq data [352,353]

1.6 Clinical Diagnostics

Breast cancer detection primarily involves self-examination, periodic clinical examination, and mammographic screening. In most Western countries, mammographic screening regimes typically recommend that women aged 50 to 74 undergo mammography every two years [354]. Some guidelines suggest starting screening as early as age 40, especially for those at higher risk, with annual screenings until the age of 55, after which biennial screening may be sufficient depending on individual risk factors [355]. While screening detects many cancers early, it also poses risks of overdiagnosis, with the possibility that some lesions never becoming invasive [356]. Current screening methods may miss tumors due to factors like the lesion's lobular phenotype or high breast density [357]. Clinically, BC often presents as a lump. Despite its imperfections, mammographic screening is currently the most effective method for early BC detection, offering the best chance for successful treatment outcomes and higher survival rates. In Sweden, 70% of all breast cancer cases are detected via screening, with guidelines recommending a diagnostic triad: clinical examination, radiological imaging, and biopsy or cytological evaluation [358].

Histopathology

Tumor biopsies and surgical samples are analyzed through histopathological techniques to inform treatment choices. These analyses classify breast tumors based on their distinct morphological and clinicopathological biomarker characteristics. Morphological assessment is crucial and remains the mainstay in diagnosing neoplastic lesions, where pathologists analyze gross and microscopic tissue structures. This analysis helps determine the disease's origin, aids in prognosis prediction, and is correlated with clinical symptoms [359]. Most of the breast malignancies are adenocarcinomas, developing from glandular epithelial cells, and constitutes more than 95% of BC. Morphological analysis of BC assesses if tumors are confined to the epithelium (in situ) or whether malignant abnormal proliferation has invaded the stroma (invasive), and their origin from the ducts or lobules.

Histopathologically, tumors are distinguished as ductal or lobular by cellular characteristics, cell quantity, secretion, immunohistochemical profile, and tissue architecture. Invasive carcinoma and carcinoma in situ are classified as ductal or lobular based on the site from which the tumor appears to have originated. In situ breast carcinomas are classified into ductal carcinoma in situ (DCIS), which makes up about 80% of cases, and lobular carcinoma in situ (LCIS), accounting for roughly 20%. DCIS, considered as a precursor to invasive

breast cancer (IBC), is the most common type of non-invasive breast malignancy and comprises ~20% of all breast neoplasms diagnosed [360]. Invasive ductal carcinoma (IDC) is the most common histologic type comprising 72-80% of all invasive BC, while invasive lobular carcinoma (ILC) is less common and accounts for 5-15% of all invasive BC [361,362]. The most frequent subgroup of IDC (40-80%) is termed "not otherwise specified" (IDC-NOS) or "of no special type" (IDC-NST) because these tumors do not exhibit sufficient characteristics to be determined as a more specific histological type. Some tumors can also be recognized as a "special type" if they present sufficient distinctive characteristics, and particular cellular and molecular behavior, e.g., mucinous adenocarcinoma and invasive micropapillary carcinoma of the breast [363,364].

Tumor grade

The Nottingham histological grade (NHG) is a well-established, simple yet strong prognostic method for evaluating all invasive breast cancer types, translating three histological features into a numerical tumor grade score to assess cancer characteristics, that is broadly used in clinical decision-making.

The grading system was modified from Bloom and Richardson [365] by Elston and Ellis and became more semi-quantitative as compared to the previous methods [366]. The features analyzed are tubule formation, nucleus pleomorphism, and mitotic count, evaluating the level of tumor differentiation to inform prognosis and treatment decisions. For each feature, a compound score is calculated, where score 1 for low, score 2 for moderate, and score 3 for high scores is assigned, then these feature scores are summed, and with increasing grade scores the resemblance to normal cells decreases and tumor aggressiveness generally increases.

Despite its straightforward approach, NHG's application can be complex due to the diversity and morphological variability of BCs, sometimes causing discrepancies among pathologists, particularly for grade 2 cases [367]. The International Collaboration on Cancer Reporting, ICCR dataset commentary, crafted and reviewed by an international panel of breast pathology experts, offers comprehensive guidance for pathologists on NHG's practical implementation [368].

Tumor staging

Tumor staging describes how advanced a cancer is, taking into account the anatomical properties and spread of the tumor, and in the TNM staging system is classified according to the categories of T (tumor size), N (status of the regional lymph nodes), and M (distant metastasis) [369]. T stage is based on the size and degree of loco-regional invasion by primary tumor and is categorized from T1 to T4. The N stage is determined by the extent of nodal involvement including axillary, internal mammary, and ipsilateral supraclavicular lymph nodes, and can be assessed clinically (by imaging or clinical examination) or pathologically (by examination of sampled tissues). Distant metastases are evaluated to determine the M stage, with positive spread, M1, indicative of stage IV disease. The updated 8th edition of American Joint Committee on Cancer (AJCC) defines nine stages (0, IA, IB, IIA, IIB, IIIA, IIIB, IIIC, and IV) based on different combinations of T, N, and M status [370].

For higher stages in advanced tumors, imaging methods such as magnetic resonance imaging (MRI) may provide a more accurate assessment than mammography or ultrasound (for example, of chest wall extension) [371], however pathological assessment is optimal. Tumor staging is critical for guiding treatment decisions and prognostic assessments in breast cancer. However, considerable inter-assessor and inter-laboratory variability exists, which can introduce uncertainty in staging accuracy. This variability often stems from differences in interpretation and measurement standards between different assessors and laboratories [372].

IHC and ISH markers

Immunohistochemistry is the gold standard technique in research and diagnostics for detecting and quantifying antigens in tissue samples and is most often performed on formalin-fixed paraffin-embedded (FFPE) tissue samples. It is preferred for its ease, dependability, and adaptability. IHC typically employs chromogenic detection via light microscopy for visualizing antigen-antibody reactions, although fluorescence detection is also done.

As previously described, IHC detection of prognostic and therapeutic markers ER, PR, HER2, and Ki67 are part of the routine work-up in BC [373,374]. For the hormone receptors, different thresholds are suggested for proportion of positive stained cells, at least 1% (or ≥10% in some countries such as Sweden) is considered a endocrine-treatable state, categorized responsive (10%), response uncertain (1%-9%), and nonresponsive (0%). According to the recent 2023 St. Gallen treatment guidelines, the panel voted 50% favoring a

1% ER-positive breast cancer to be treated with endocrine therapy, jumping to nearly 80% in favor of endocrine therapy at an ER IHC score of 9% [375,376].

According to the Swedish guidelines, ER and PR status are considered positive when ≥10% of tumor cells show ER- and PR-specific staining in tumor nuclei detected by IHC. Both ER and PR status taken into account as binary variables, either positive (≥10%) or negative (<10%) [377]. According to the same Swedish guidelines, Ki67 proliferation index is evaluated as a continuous measure defined by the percentage of positively staining tumor cell nuclei and the threshold for Ki67 intermediate is 6% to 29%, and ≥30% as Ki67-high [377]. For HER2 overexpression and *ERBB2* amplification testing, the ASCO/CAP 2018 guidelines are followed in Sweden and performed on the invasive component. IHC is performed first, with IHC 3+ tumors considered HER2 positive, and IHC 0, 1+ are considered negative. For IHC 2+, a reflex ISH assay is performed [290]. The *ERBB2* ISH detects specific DNA sequences through the use of labeled complementary nucleic acid probes, and either a fluorescent (FISH) or chromogenic (CISH or SISH) readout. For *ERBB2*, a case is ISH positive if the HER2/CEP17 ratio is ≥2.0, and the HER2 copy number signals/cell is ≥4, negative if HER2/CEP17 ratio is <2.0, and the HER2 copy number signals/cell is <4. Further HER2 groups known as groups 2, 3, and 4, are determined pending further workup, as explained in the earlier 'Clinical assessment of HER2 status' section and in Figure 1.11. Concordance between IHC and ISH results in case of HER2/*ERBB2* have been reported with varied results, although, significantly lower incongruities for both IHC−/ISH+ and IHC+/ISH− were found in the studies that were published after the 2018 guideline update [378]. IHC and/or ISH determination of ER, PR, and HER2, along with other markers Ki67 and NHG (Figure 1.12) are used as standard for clinical subtyping of breast cancers, as well as prognostic and predictive biomarkers to guide treatment [379].

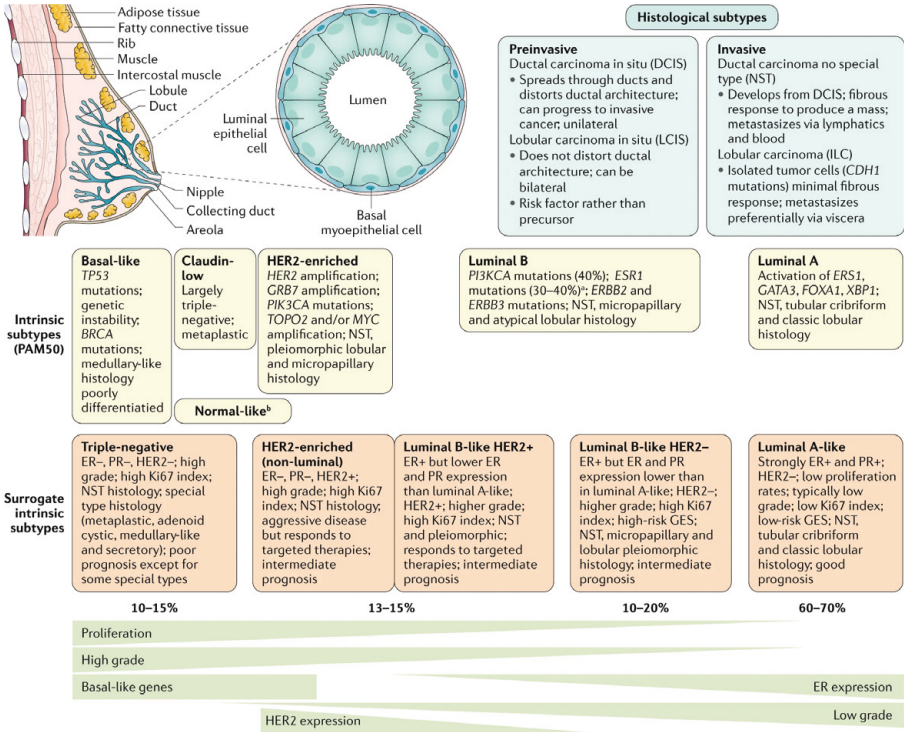

Figure 1.12: Breast Cancer histological and intrinsic subtypes. Reprinted with permission from [380].

Liquid biopsies

Tissue biopsies and surgical samples are invasive, challenging to perform on a repeat basis, and in some cases can lead to complications. A liquid biopsy is the concept of taking a liquid biological sample from a patient, typically a blood sample, in order to learn something about the disease state in the patient in a less-invasive manner. In particular, circulating tumor DNA (ctDNA) analysis is emerging as a very promising approach that has many potential clinical uses, from early detection, to detection of targetable mutations, to monitoring treatment response and identification of relapse disease [381,382]. This has spurred extensive research into liquid biopsies' applications in cancer screening, prognosis, treatment monitoring, and decision-making, leading to the recent development of FDA-approved techniques for ctDNA assessment [383].

1.7 Breast Cancer Treatments

Breast cancer treatment is guided by tumor burden (total amount of cancer in the body, assessed with radiological exams) as well as tumor subtype. Generally, hormone receptor positive (HR+) tumors receive endocrine therapy, with chemotherapy for cases with more aggressive features. HER2+ tumors are treated with anti-HER2-directed antibody therapy plus chemotherapy, and additional endocrine therapy if HR+. TNBC, by definition do not have these targets, and are treated exclusively with chemotherapy.

1.7.1 Early breast cancer

Therapeutic goals of primary, early, nonmetastatic BC are eradicating tumor from the breast and regional lymph nodes and preventing metastatic recurrence. Most patients are evaluated for potential pre-operative treatment, so called neoadjuvant therapy (as discussed below), which has the objective of reducing tumor burden to facilitate less radical surgery.

Surgery

Surgical removal of the primary breast tumor is key to curative treatment. Also known as 'local therapy', these tumors undergo surgical resection (lumpectomy/ breast conserving or mastectomy) and sampling or removal of axillary lymph nodes, with consideration of postoperative radiation. For patients with suspicious axillary lymph nodes, preoperative evaluation using ultrasound and fine needle aspiration or core biopsy helps in deciding the appropriate surgical method and assessing the need for neoadjuvant therapy. Patients with a clinically negative axillary examination undergoes sentinel lymph node biopsy at the time of surgery. De-escalation of nodal assessment has been successfully implemented in some pivotal clinical trials, that showed no overall survival difference between sentinel lymph node biopsy and axillary lymph node dissection [384].

Lumpectomy is a breast-conserving surgery where only the tumor and a small margin of surrounding tissue are removed, preserving most of the breast, often followed by radiation therapy (RT) to target any residual disease. On the other hand, mastectomy involves removal of the entire breast and may be appropriate for larger tumors or multifocal tumors. Breast-conserving surgery has largely replaced mastectomy, the historical standard [385]. The choice between immediate surgery or initial systemic neoadjuvant therapy depends on factors like tumor size, its ratio to breast size, tumor biology, patient

comorbidities, and personal preference, and generally the use of neoadjuvant treatment has increased considerably in the past decade. The surgical planning should be based on the updated size and extent of the tumor following neoadjuvant systemic therapy [386].

Radiation therapy

Radiation therapy (RT; or radiotherapy) involves using high-energy rays, X rays (photons), protons, or electron particles, to destroy cancer cells. Typically administered after surgery, it aims to eliminate any remaining cancer cells, reducing the risk of recurrence. RT generated by a linear accelerator is delivered by a machine that directs radiation to the affected area. The treatment plan, including the dose and duration, is tailored to the patient's specific condition and type of surgery (lumpectomy or mastectomy). It can also be given internally (brachytherapy), after surgery, where a radiation-delivery device is temporarily placed in the breast in the area where the tumor was located. Adaptive radiotherapy involves modifying the radiation plan in response to changes during therapy, to optimize therapy effectiveness ensuring precise and effective radiation delivery throughout the course of treatment [387]. RT is an integral part of BC treatment, where postoperative RT was shown to improve disease-free and overall survival for patients with early BC with lymph node involvement by the elimination of residual tumor cells after both breast-conserving surgery [388] as well as mastectomy [389] assessed in large scale longer (15 years) follow-up meta-analysis studies.

Systemic therapy

Systemic therapy consists of treatment with drugs that work throughout and affect the entire body, as opposed to localized treatments like surgery or RT. They can be given as an injection, infusion, or oral medication and includes chemotherapy, hormonal treatment, targeted therapy, and immunotherapy. The type of treatment depends on individual indication based on tumor biomarkers, molecular subtype, tumor burden, and risk of recurrence. Systemic therapy may be preoperative (neoadjuvant), postoperative (adjuvant), or both.

Neoadjuvant therapy rationale

Neoadjuvant (or preoperative) systemic therapies (NST) are treatments given before operation or surgery, and depending on the clinicopathological features, can include chemotherapy, hormonal therapy, and/or targeted therapy, and in

some cases, RT. The rationale behind neoadjuvant chemotherapy (NAC) is to reduce the size or extent of the tumor before surgery, making it easier to remove, which can improve the surgical outcomes. It may also allow for surgical de-escalation, allowing for less extensive surgery in the breast and axilla. It was shown in a clinical trial that around 40% of HER2+ and TNBC tumors that initially required mastectomy can be converted to breast-conserving surgery (BCS) candidates using NST [390]. However, it has also been observed in randomized clinical trials (RCT) that systemic chemotherapy given before or after surgery showed no differences in long-term outcomes among all BC tumors [391].

Further, NST may also offer prognostic information, providing an opportunity to assess the tumor's response to the therapy, identifying patients with residual disease at primary surgery who may then require additional adjuvant therapy, potentially improving long-term outcomes. In HER2+ and TNBC patients it was shown that achieving pathological complete response (pCR) after NAC is predictive of significantly better DFS and OS [392,393].

Therefore, based on evidence from multiple RCT, NAT is now preferred for most patients with stage II or III BC, in particular if TNBC or HER2+. For instance, the KATHERINE trial revealed that patients with HER2+ early BC without a pCR to neoadjuvant therapy with trastuzumab had improved DFS when switched from standard adjuvant HER2-targeted antibody therapy to T-DM1, an ADC, as opposed to those who continued with anti-HER2 antibody treatment, with DFS increasing from 77% with trastuzumab to 88% with T-DM1 [394].

Chemotherapy

Chemotherapy (CT) is the use of cytotoxic drugs to kill cancer cells throughout the body. Different chemotherapy regimens are considered in neoadjuvant and adjuvant settings, also depending on whether the patient is low-risk (less aggressive, 0-3 lymph nodes involved, HR+/HER2-, low ROR) or high-risk (with >3 positive nodes and aggressive subtypes TNBC or HER2+, high ROR). According to the 2018 NICE (National Institute of Clinical Excellence, United Kingdom) guidelines for low-risk patients, docetaxel, cyclophosphamide, and 5-fluorouracil are considered reasonable choices with less toxicity. For high-risk patients, anthracycline (e.g., adriamycin) and taxane (such as adriamycin/cyclophosphamide followed by a taxane) remains the appropriate choice to achieve the highest risk reduction, specifically, for TNBC with greater LN involvement, anthracyclines appear to be the most effective option [395].

In luminal HR+/HER2-, a gene expression assay such as MammaPrint, Oncotype DX, Prosigna, EndoPredict, or BCI, may be considered in determining the recommendation for chemotherapy in addition to endocrine therapy [396].

In HER2+ and TNBC, NAC and adjuvant CT is preferred for stage II or III tumors and dose-dense (in which the rate of delivery, rather than the overall dose, is increased) anthracycline and taxane-based CT are the standard of care. Also, for stage I TNBC, an anthracycline-free regimen might be considered [397]. Moreover, adding platinum based NAC was shown to be associated with improved pCR rates and significantly increased event free survival (EFS) in TNBC patients [398].

Endocrine therapy

Endocrine therapy (ET), which counteracts estrogen-promoted tumor growth, is standard systemic therapy for HR+ BC patients after surgery. It consists of antiestrogen medication taken daily, typically for 5 years, and options differ according to menopausal status. Tamoxifen (a selective estrogen receptor modulator, SERM, which binds to and inhibits ER) is effective in both pre- and postmenopausal women, and is a standard in premenopausal patients [399]. Aromatase inhibitors (AI) such as anastrozole, exemestane, and letrozole are effective only in postmenopausal women (including those who are postmenopausal because of medical ovarian suppression or oophorectomy), and work by inhibiting conversion of androgens to estrogen, thereby decreasing systemic estrogen levels. Five years of AI has been shown to reduce BC mortality by ~15%, reducing recurrence rate and BC death, compared with tamoxifen monotherapy [400].

Compared to HR- BC, HR+ breast tumors are characterized by later recurrences, which can occur from 5 to 20 years after surgery. Clinical trial results have shown that continuation of ET after 5 years and extended for up to 10-15 years decreased relapse rates in patients at high risk (e.g., node-positive or high genomic score) [401]. Treatment management summary of early breast cancer based on tumor burden and subtype is shown in Figure 1.13.

HER2-targeted treatments

As summarized in section 1.5.3, development of HER2/*ERBB2*-targeted therapy has been one of the greatest advances in BC treatment. Addition of 1 year of trastuzumab, a mAb targeting the HER2 extracellular domain, to standard adjuvant chemotherapy, has been shown to markedly improve DFS and OS in HER2+ BC in four randomized adjuvant trials [26,27]. Dual blocking by adding pertuzumab to trastuzumab and chemotherapy significantly improved the invasive-disease-free survival, iDFS rate in patients with LN+, HER2+, operable BC in the adjuvant setting [402]. For low-risk HER2+ stage I tumors, weekly CT for 12 weeks + trastuzumab for 1 year is generally accepted as a standard of care based on non-randomized clinical trials [403,404]. For high-risk HER2+ BC, an additional adjuvant treatment of neratinib (a TKI that targets HER1/HER2/HER4) for 1 year after completion of trastuzumab had shown to improve iDFS and decrease the risk of central nervous system recurrence in two independent phase III clinical trials [405,406].

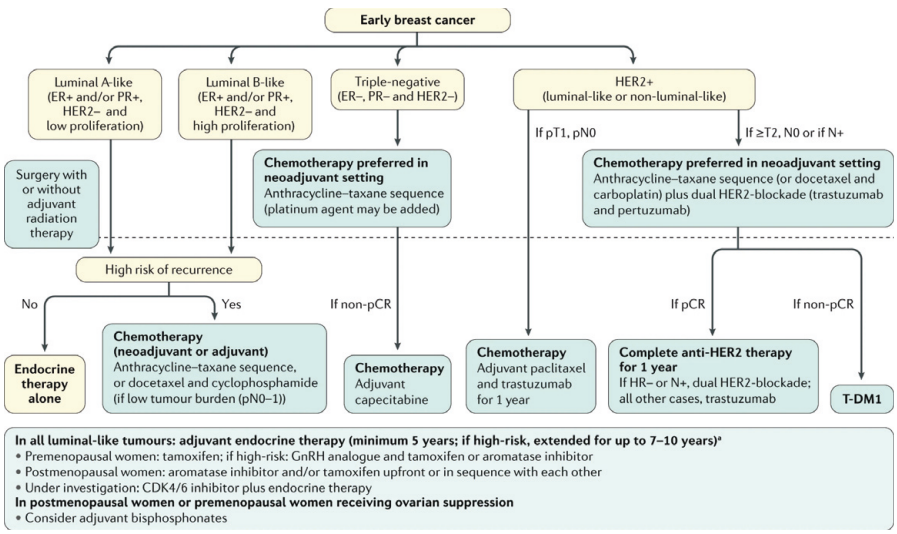

Figure 1.13: Treatment management summary for early breast cancer. The treatment algorithm is based on histological subtypes and tumor burden. Abbreviations: –, negative; +, positive; GnRH: gonadotropin-releasing hormone; HR: hormone receptor; pCR: pathological complete response; PR: progesterone receptor; N: node status; T: tumor size; T-DM1: ado-trastuzumab emtansine. Reprinted with permission from [380].

1.7.2 Advanced breast cancer

For advanced BC (ABC, or metastatic BC, MBC), prolonging life and symptom palliation are the main therapeutic goals. ABC comprises inoperable locally advanced BC, which has not spread to distant organs, together with metastatic ABC (MBC), which has common sites of spread to the bone, lungs and liver [407]. Currently, ABC management utilizes the same basic categories of systemic therapy as in neoadjuvant/adjuvant approaches in addition to some recently adapted ones, summarized here. Local therapies such as surgery and RT are typically used for palliation only in metastatic disease. Despite aggressive treatment, ABC still remains virtually an incurable disease, with metastases being the cause of death in almost all patients, with a median OS of approximately 2-3 years.

According to the BC subtype, systemic therapy in MBC includes standard regimens used early in treatment course (i.e., early lines) plus agents for consideration later in treatment course (i.e., later lines). In HR+HER2- MBC, several lines of endocrine-based therapy and their combinations (such as tamoxifen, AI, or fulvestrant) are used (along with ovarian suppression or ablation, in pre-menopausal patients) until endocrine resistance develops or unless severe organ dysfunction emerges [408]. To delay or overcome endocrine resistance, incorporating CDK4/6 inhibitors and target of rapamycin (mTOR) inhibitors, which have been shown to improve PFS and in some circumstances OS, are being introduced as standard of care in first, second or further lines in postmenopausal and premenopausal MBC patients [409–411].

For HER2+ MBC, dual HER2 blockade with trastuzumab and pertuzumab in combination with chemotherapy is currently the recommended first-line therapy [412]. For HER2+HR+ tumors, receives ET in combination with HER2-targeted therapy [318]. T-DM1, T-DXd, and pyrotinib, an irreversible pan-ErbB receptor TKI, are currently being studied as second line options. For example, T-DXd showed significant improvement in PFS versus T-DM1 in second-line treatment for HER2+ unresectable or MBC in the phase III DESTINY-Breast03 trial [413]. Another currently ongoing trial, DESTINY-Breast09, is comparing the combination of T-DXd with pertuzumab and trastuzumab plus docetaxel as first-line standard-of-care regimen [414]. After previous adjuvant exposure to trastuzumab and taxanes, the combination of capecitabine and pyrotinib, have shown to significantly improve PFS compared with lapatinib plus capecitabine, in phase III randomized PHOEBE trial in HER2+ MBC [415]. In the future, the advent of new and more specific anti-HER2 drugs should significantly bolster treatment options for HER2+ MBC, for whom continuous anti-HER2 treatment

is essential. Regardless, continuous HER2 pathway suppression is crucial, HER2-targeted therapy should be offered beyond progression.

As for other targeted therapies, in PD-L1+ metastatic TNBC, the addition of an immune checkpoint inhibitor to chemotherapy (atezolizumab or pembrolizumab + CT) as a first-line treatment, was shown to modestly improved PFS with the addition of atezolizumab in the entire study population of TNBC but indicated a significant PFS improvement in the PD-L1-positive subset of TNBC patients [416,417]. ADCs directed towards Trop-2, trophoblast cell-surface antigen 2, have also shown promising results with significantly improved both PFS and OS in TNBC without brain metastases as compared with standard therapy [418]. In PD-L1+ *BRCA1/2* germline mutated metastatic TNBC, checkpoint inhibitor-based therapy in the first and the PARPi (olaparib or talazoparib) in the latter lines are being currently tested as compared to chemotherapy [419,420]. In PD-L1-negative (without PD-L1 expression) *BRCA1/2* germline mutated metastatic TNBC, chemotherapy is currently offered as first-line treatment. Additionally, new cytotoxic agents and non-taxane microtubule dynamics inhibitor such as eribulin shown to significantly improved OS in heavily pretreated patients when compared with treatment of physician's choice [421,422].

1.8 Methodological Advancements

In the field of oncology, accurately determining gene expression levels and genetic variations is crucial for diagnosis, prognosis, and treatment decision-making. As described in section 1.6, traditional methodologies such as IHC and ISH have been extensively used in clinical settings to detect and quantify biomarkers such as HER2 in breast cancer. However, these techniques come with inherent methodological challenges that can impact their reliability and sensitivity.

Briefly, IHC method involves staining tissue sections with antibodies that are specific to the proteins of interest, and then visually inspecting the stained slides to determine the proportion of cells that exhibit positive staining. A cutoff value is used to categorize the expression level of these biomarkers into "low" and "high" categories. Despite its widespread use, IHC is susceptible to several disadvantages that can affect its reliability and reproducibility. Technical variability arises from differences in the choice of antibodies, tissue preparation and staining protocols, and equipment calibration, which can lead to inconsistent results [423]. Additionally, IHC interpretations are semi-quantitative, subjective, and can vary between pathologists or even on different occasions by the same pathologist, leading to potential

misclassification of biomarker status, impacting treatment decisions [424]. Despite efforts to standardize IHC scoring, it cannot overcome being semi-quantitative, with staining intensity and cell percentage estimates introducing further uncertainty. Moreover, in critical cases such as HER2 testing, IHC often requires supplementary genetic testing like FISH or SISH to confirm gene amplification, particularly in ambiguous cases.

ISH is another critical diagnostic tool used in breast cancer to assess gene expression directly within the tissue context. ISH techniques, including FISH/SISH/CISH, involve hybridizing a labeled DNA or RNA probe to specific nucleic acid targets within a tissue section. The probe's signal is then visualized under a microscope, providing spatial information about gene expression or copy number variations like HER2 amplification. While ISH offers valuable insights into the genetic aspects of tumor cells within their native histological context, it comes with several limitations: it is complex and considerably more costly as compared to IHC. It requires sophisticated equipment and skilled personnel, and the interpretation of results can be subjective, especially with SISH/CISH where the signal is not fluorescent. Additionally, the ISH process is time-consuming and sensitive to sample and probe quality (potential for probe signal degradation), potentially leading to delays in diagnosis and false-negative results [425].

Limitations in both IHC and ISH highlight the need for more efficient and precise methods such as digital PCR (dPCR), which offers absolute quantification of nucleic acids with high sensitivity and specificity, and without the subjective interpretation associated with traditional IHC or ISH, making it an advantageous approach for determining and assessing biomarker status in breast cancer diagnostics.

To overcome these challenges, **Study II** and **Study V** of this thesis has utilized droplet digital PCR (ddPCR), an enhancement over traditional IHC and ISH methods as well as other quantitative molecular methods such as quantitative PCR (qPCR). Although dPCR is discussed in more detail in Section 3.7, briefly, this method works by partitioning a sample into thousands of tiny droplets, each containing zero, one, or more copies of the target nucleic acid (DNA or cDNA). PCR amplification then occurs in each individual droplet, allowing for direct counting of the DNA molecules present. This allows for the precise quantification of nucleic acids, making it highly effective in detecting small variations in gene copies, mutations, and transcript levels that might be missed by IHC or ISH [426]. The key advantage of dPCR over traditional PCR methods is its ability to provide absolute quantification without the need for external standards [427]. This makes dPCR particularly useful for applications requiring high precision, such as measuring slight differences in gene expression, detecting rare genetic variants, and quantifying copy number

variations. Clinical applications where decision-making depends on precise biomarker quantification, dPCR may provide a more reproducible and reliable result by minimizing the variability associated with operator interpretation and technical execution. Its high sensitivity and accuracy make it an excellent choice for clinical diagnostics and research in oncology. Moreover, dPCR can process multiple samples simultaneously with a high degree of automation, thereby increasing the throughput and scalability, also reducing the hands-on time required for large-scale studies [428].

1.9 Precision Medicine

Precision or personalized medicine encompasses both precision diagnosis and precision therapy, each tailored to the individual patient. This approach aims to treat conditions as accurately as possible, thereby avoiding overtreatment or unnecessary interventions, enhancing the effectiveness of targeted treatment and potentially reduce side effects that could otherwise diminish a patient's quality of life. By carefully matching therapies to the individual characteristics of each patient's disease, precision medicine helps to optimize therapeutic outcomes while minimizing adverse effects. Precision medicine achieved significant milestones in oncology in 2017 with two major breakthroughs. In May, FDA granted accelerated approval to pembrolizumab, an anti-programmed cell death-1 (PD-1) mAb, for the treatment of adult and pediatric patients with unresectable or metastatic solid tumors identified as microsatellite instability-high (MSI-H) or deficient in DNA mismatch repair (dMMR) [429]. Shortly thereafter, in August, nivolumab, another anti-PD-1 mAb, also received accelerated approval for adult and pediatric patients with MSI-H or dMMR metastatic colorectal cancer and melanoma that had progressed following standard chemotherapy [430]. These approvals are particularly noteworthy as they mark the first instances of a cancer treatment being approved based on a common biomarker, regardless of the tumor's location in the body. In the same year, larotrectinib – an oral, potent, and selective inhibitor of tropomyosin receptor kinases – demonstrated unprecedented efficacy in treating unresectable or metastatic solid tumors harboring neurotrophic tropomyosin receptor kinase (NTRK) fusion proteins in both adult and pediatric patients [431]. This breakthrough highlights the potential of targeting specific genetic drivers of cancer across different tumor types.

Precision diagnosis involves using comprehensive biomarker testing to accurately diagnose each cancer patient and their unique cancer biology. This testing identifies molecular and immunological vulnerabilities in tumors which

helps stratify patients for therapy. As the foundational step in this process, precision diagnostics are critical because they enable the matching of patients with treatments specifically suited to target their cancer's unique genetic and molecular aberrations. Developing highly precise assays is essential for accurately identifying these cancer vulnerabilities in an absolute quantitative manner. Achieving this goal allows for the optimal matching of drugs to patients, ensuring the right treatment is administered at the right dose and at the right time, while minimizing potential harm to the patient. ESCAT (ESMO Scale for Clinical Actionability of Molecular Targets) system plays an important role in the execution of precision medicine for clinical cancer therapy. It ranks molecular targets in six different classes based on the clinical indication of actionability for a specific cancer type, assisting clinicians in making informed therapeutic decisions. ESCAT categorizes genomic alterations and their associated targeted therapies into tiers based on the level of evidence from clinical studies. For instance, Tier Ia represents targets evaluated in prospective, randomized trials; Tier Ib signifies evaluation in prospective, single-arm trials; and Tier Ic in basket trials. Key molecular targets such as *ERBB2*, *PIK3CA*, and germline *BRCA1/2* alterations, which have shown significant survival benefits in breast cancer when matched with appropriate therapies, have been evaluated in a number of Tier Ia studies [432,433]. In this context, the research presented in this thesis aligns with the themes of precision diagnostics emphasized by the ESCAT system. **Studies II** and **V** consists of the development of multiplex ddPCR assays to detect *ERBB2* CN and classify HER2 mRNA expression levels, including those classified as HER2-low. This advanced and sensitive diagnostic capability is crucial for accurately stratifying patients and matching them with the most effective targeted therapies.

1.10 The SCAN-B Initiative

The Sweden Cancerome Analysis Network – Breast (SCAN-B; ClinicalTrials.gov identifier NCT02306096) Initiative is a large-scale multicenter, population-based, prospective study, initiated in 2010, focusing on breast cancer [434]. The study's primary objective is to elucidate the molecular mechanisms and genetic underpinnings of breast cancer by analyzing NGS transcriptomics and genomic data from a broad cohort of patients, to develop new diagnostics tools, and to validate these and eventually clinically-implement the most promising tests. As of the time this thesis was written, nearly 21,000 patients diagnosed with breast cancer were consented and accrued into the SCAN-B study. Of these, over 20,000 have provided blood samples, over 14,000 have provided

tumor tissue samples, and more than 13,000 of these breast tumors have been processed to obtain RNA-seq data. This extensive collection allows for in-depth analysis and understanding of the diverse molecular landscapes across breast cancer subtypes. This comprehensive study seeks to improve the understanding of breast cancer biology and develop new prognostic and predictive tests, which can lead to more effective and personalized treatment strategies. The project involves collaboration among various hospitals, mostly from southern Sweden and employs advanced molecular profiling by NGS RNA-Seq for every available breast tumor [435].

Across a wide geography of Sweden, as shown in Figure 1.14, nearly all patients at the participating hospitals diagnosed with primary breast cancer are consented and included, and as part of the national treatment guidelines receive standardized therapeutic regimens. Mass enrollment of BC patients in this manner on the population level scale provides as unbiased group of cases as realistically possible with which molecular studies can be performed to deepen our understanding of cancer biology and air biomarker development.

The SCAN-B cohort is representative and generalizable for the wider population, because it reflects the disease at population level. This strategy provides comprehensive data across diverse demographic groups, reflecting real-world scenarios. This inclusivity improves understanding of the natural history of breast cancer and may also provide insights into cancer epidemiology and risk factors. Additionally, it can track long-term outcomes and treatment effectiveness across a population, aiding in identifying public health trends and guiding healthcare policies. Population-level trends can be analyzed to identify unintended consequences of screening, such as overdiagnosis, and drive efforts aimed at improving outcomes. Other cohorts such as TCGA have been found to over-represent high grade and ER negative tumors leading to difficulty in cross-comparison, which may lead to biases [436].

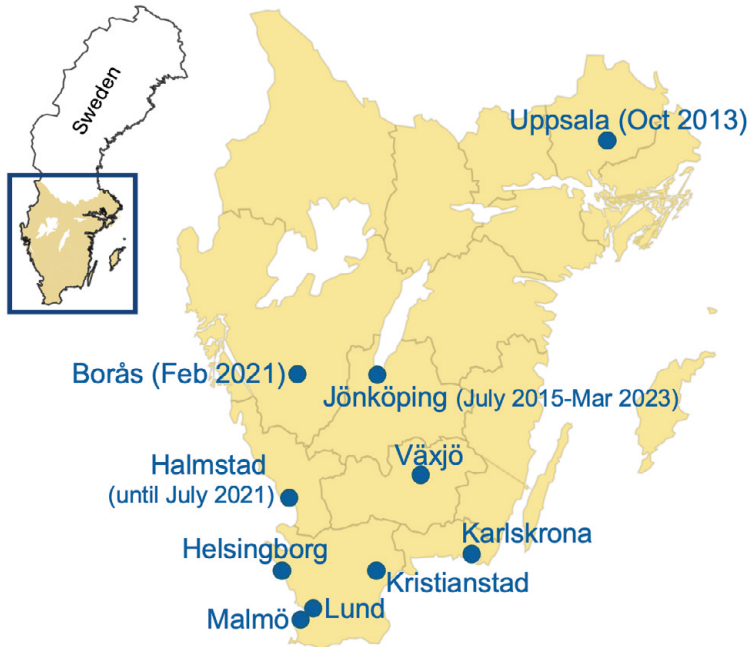

Figure 1.14: Map of Sweden indicating the hospital sites enrolling patients in SCAN-B. The month and year of sites joining after 2010 are indicated in parentheses.

Studies I, **II**, **IV**, and **V** included in this thesis uses material from SCAN-B, with **Study IV** providing a 10-year update on SCAN-B, including updates on enrolment, study demographics, baseline clinical variables, patient treatment, and long-term outcomes. Additionally, the SCAN-B cohort has facilitated a broad spectrum of research projects. These studies have explored various aspects of BC, including the characteristics of TNBC [437] and *BRCA1*-abnormal tumors [438], multi gene expression signatures benchmarking [439], dynamics of gene fusions [440], refinement of molecular and PAM50 subtyping [441–443], lymph node metastasis predictors [444], RNA-seq based classifiers for conventional histopathological biomarkers [352], SNV/indel detection from RNA-seq data [445], identification of pre-existing somatic *ESR1* mutations [446], detection of ctDNA before and after mammographic breast compression [447], and integration of single-sample predictor (SSP) models trained on RNA-seq data into clinical decision-making [353].

As previously mentioned, the SCAN-B initiative aims to i) integrate advanced -omic profiling into the clinical routine for breast cancer, ii) improve patient stratification, diagnosis, prognostication, and prediction of treatment effects,

and iii) ensure these advancements are accessible to patients by implementing them within the healthcare system via SCAN-B, and in clinical trials, and in collaboration with industry. These goals align closely with the precision medicine and diagnostics studies presented in this thesis. This thesis contributes to these objectives by exploring precision medicine and diagnostics approach in five studies, enhancing the understanding and application of biomarkers in breast cancer treatment. **Study I** investigate the role of *ESR2*, highlighting its potential as a tumor suppressor and its association with favorable clinical outcomes due to immune activation. **Study II** utilizes ddPCR for precise quantification of *ERBB2* copy number, identifying a biologically distinct group with 'ultrahigh' *ERBB2* CN, underscoring the need for tailored treatment strategies. **Study III** emphasizes CITED1 as a prognostic marker for anti-endocrine therapy in ER+/LN- breast cancer, enhancing tamoxifen responsiveness prediction. **Study IV** provides a 10-year update on how RNA-sequencing can be integrated into clinical practice to improve tumor classification, prognosis, and treatment prediction. This work highlights the utility of large-scale, population-based genomic analyses in translational research and biomarker signature development, hopefully leading to novel clinical tools. **Study V** advances the precise classification of HER2 mRNA status by developing a new multiplex ddPCR assay, with a specific focus on the HER2-low category. This assay enhances the differentiation of HER2-low from HER2-zero and HER2-positive subtypes, facilitating more targeted therapeutic strategies, particularly for the emerging subclass of HER2-low breast cancer. Together, these studies underscore the significant role of advanced molecular genomics and transcriptomics technologies that have the potential to refine the current diagnostic and treatment paradigms within the clinical framework, advancing the field towards more personalized breast cancer management.

2. Aims

The overall aim of this thesis is to enhance precision diagnostics in breast cancer by leveraging sophisticated molecular assays and refining patient stratification based on biomarker analysis through extensive transcriptomics data evaluation within SCAN-B. Focusing on pivotal biomarkers such as *ESR2*, *ERBB2*, and *CITED1*, this thesis aims to sharpen diagnostic methods using digital PCR assays, advance prognostic evaluations, and tailor treatment approaches to individual BC patient needs, in line with personalized medicine principles. Supported by the SCAN-B initiative's framework, this work demonstrates the potential of RNA-seq to improve breast cancer management by providing practical applications of advanced genomic tools in a clinical context.

Study I: The aim was to assess the prevalence and clinical significance of *ESR2* mRNA expression in invasive BC, particularly its correlation with clinicopathological parameters, survival outcomes and immune activation. The study also aimed to provide initial evidence from large-scale population-based SCAN-B dataset and enhance understanding of ERβ's role in breast tumor biology and its microenvironment.

Study II: This study aimed to validate and demonstrate the effectiveness of multiplex dPCR as a tool for precise and accurate *ERBB2* copy number analysis in invasive BC diagnostics. The objective was to establish dPCR as a potential standard diagnostic approach.

Study III: The objective was to explore the potential of *CITED1* as a prognostic biomarker for anti-endocrine therapy in ER+/LN- BC patients. This study focused on assessing the expression levels of *CITED1* at mRNA and protein level and its association with treatment outcomes, its role in ERα-mediated transcription and its impact on patient response to anti-endocrine treatments.

Study IV: The aim of this project was to provide a comprehensive 10-year update on the SCAN-B initiative, showcasing the progress made in incorporating RNA-seq into clinical decision-making, and presenting an up-to-date analysis of the clinical subtypes, treatment overviews, and survival outcomes.

Study V: This study aimed to refine the classification of HER2/*ERBB2* mRNA status in invasive BC using a newly developed multiplex dPCR assay. The goal

was to improve the precision in classifying HER2-low, HER2-ultralow, HER2-zero, and HER2-positive subgroups, thereby enhancing the selection of patients for targeted therapies like T-DXd based on a more accurate assessment of *ERBB2* mRNA expression levels.

3. Methods

3.1 Patients, Ethics, and Samples

In this thesis **Studies I**, **II**, **IV**, and **V** are based on breast cancer samples collected from patients that are part of the SCAN-B cohort, as well as a minority of cases in **Study II** from SCAN-B's precursor, the All Breast Cancer in Malmö (ABiM) cohort. **Study III** utilized the GOBO (Gene expression-based Outcome for Breast cancer Online) database of 11 publicly available breast cancer microarray datasets, comprised of 1881 cases [448], as well as the public gene expression data from 51 breast cancer cell lines [449]. Within SCAN-B and ABiM, all patients provide informed written consent. **Studies I**, **II**, **IV**, and **V** were performed under approvals from the Regional Ethical Review Board of Lund at Lund University (approval numbers 2002/613, 2007/155, 2009/658, 2009/659, 2010/383, 2012/58, 2013/459, and 2015/277), the Swedish Data Inspection group (diary number 364-2010) and were conducted in adherence to the World Medical Association's Declaration of Helsinki [434]. In **Study III**, data was retrieved from publicly available datasets and thus no ethical permissions were required: TCGA is publicly available data and information on the ethics and policies of data usage, including informed content can be found here https://www.cancer.gov/ccg/research/genome-sequencing/tcga/history/ethics-policies. Sampling of surgical tumor samples in SCAN-B and ABiM occurred after the completion of the pathological assessment, ensuring that participation does not compromise the standard clinical care of the patients. In SCAN-B, tissue samples are preserved fresh in RNAlater (Thermo Fisher Scientific) prior to storage at -80 C. In ABiM, the tumor samples were obtained were preserved flash frozen on dry ice prior to storage at -80 C. Table 3.1 summarizes the patient material and experimental setups used in the thesis works.

Table 3.1: Patient cohorts, data sources, and experimental setup used in Studies.

Study	Sources	Materials	Experimental Setup	Patient numbers
I	SCAN-B	Tumors	RNA-seq	3207
II	SCAN-B, ABiM	Tumors, cell lines	DNA ddPCR & RNA-seq	909 & 682
III	GOBO	Tumors, cell lines	Microarray	1881 & 51
IV	SCAN-B	Tumors	RNA-seq	9323
V	SCAN-B	Tumors, cell lines	RNA ddPCR & RNA-seq	1223

Study I: A total of 3207 patients were included from the SCAN-B cohort, which were a subset of previously described cases in Brueffer et al. [352] that comprised 3217 patients (reduced to 3207 samples due to additional quality controls). The median follow-up time was 6.2 years. Tumor samples are preserved in RNAlater (Thermo Fisher Scientific) at the pathology laboratory immediately after surgery, flash frozen, and processed with the AllPrep (Qiagen) method according to standardized SCAN-B protocols [434,352,435].

Study II: This study included 909 primary invasive breast tumors in the droplet digital PCR (ddPCR) part, and 682 cases in the RNA-seq validation part, with all patients diagnosed between 2006 to 2019. Among the 909 cases, 510 SCAN-B cases were comprised of three random selections of 170 BCs each from within the clinical HER2 IHC 0-1+, 2+, and 3+ groups [434], and 399 of 405 ABiM cases previously described [352]. Of the 909 patients, 177 patients were clinically HER2-positive, received no neoadjuvant treatment, and received adjuvant trastuzumab alone or in combination with chemotherapy and/or endocrine therapy according to Swedish treatment guidelines. For validation, an RNA-seq data set for 682 consecutive patients with HER2-positive BC from SCAN-B was used, wherein the patients were diagnosed between 2010 and 2018, and similarly received no neoadjuvant therapy but received adjuvant trastuzumab alone or in combination with chemotherapy and/or endocrine therapy. Tumor DNA isolated from the BC samples using the AllPrep method according to standardized SCAN-B protocols [434,352,435] was utilized.

Study III: The primary dataset for the analysis comprised gene expression profiles from 1,881 breast tumor samples and 51 breast cancer cell lines,

classified according to molecular subtypes [448]. All gene expression measurements were generated using Affymetrix U133A microarrays across several published studies. The tumor samples consisted of ER-positive, ER-negative, untreated tumors and tumors treated with tamoxifen (TAM). TAM treated ER+/luminal tumors were selected for mRNA expression and survival analysis. Additionally, TCGA data was utilized to further explore the prognostic significance of *CITED1* expression within the ER+/luminal subtype. Tissue microarray (TMA) dataset was used for *CITED1* mRNA and protein correlation analysis, consisting of 400 breast tumors with long term follow up [450]. Additionally, cell line experiments with CITED1 gene overexpression and TAM treatment were performed, western blots and IHC were performed to assess CITED1 and ER-α protein expression, and ddPCR was performed on complementary DNA (cDNA) to quantify gene expression levels of *CITED1*, *IPO8* (internal control), *AREG*, and *TGFα*.

Study IV: This manuscript summarizes the SCAN-B cohort for its first 10 complete calendar years, from August 30, 2010, to December 31, 2020, thus comprising a total of 16381 patients (16269 women and 112 men) diagnosed with primary breast cancer. Of these, 9915 patients have a tumor specimen sampled for research, of which 9323 have available RNA-seq data.

Study V: This study utilized 1,242 primary invasive ductal breast tumors from 1223 consecutive patients enrolled in SCAN-B that met the following selection criteria: diagnosed between September 2010 to February 2013 at the hospitals in Kristianstad, Halmstad, Helsingborg, Lund, or Malmö, and receiving no neoadjuvant treatment. All samples were handled as described previously for SCAN-B [434,435].

3.2 High-throughput Sequencing

The thesis studies herein utilize data generated by massively-parallel sequencing, also known as next-generation sequencing (NGS), which refers to high-throughput sequencing technologies developed in the mid-1990s and commercially available since 2005 [451]. NGS methods utilize the concept of parallel processing, enabling the simultaneous sequencing of millions to billions of short DNA fragments. Each sequencing run can generate vast quantities of data, with individual reads ranging up to 400 bases in typical NGS approaches [452]. This technological advancement has revolutionized genomic research by significantly increasing the speed and scope of DNA and RNA analysis.

Foundational methodology

The foundational innovations underpinning sequencing are the Sanger dideoxy synthesis [453,454] and Maxam-Gilbert [455] chemical cleavage methods. The Maxam-Gilbert method involves chemical modification of DNA and subsequent cleavage at modified nucleotides. Sanger sequencing, on the other hand, uses chain-terminating dideoxy nucleotides which halt DNA synthesis at specific points, with these terminators being detectable via radioactive and later fluorescent labeling. Significant advancements in Sanger sequencing were: 1) development of fluorescent dyes, 2) thermal-cycle sequencing with reduced DNA input and thermostable polymerases for efficient terminator dye incorporation, 3) lower requirement for toxic chemicals and radioisotopes compared to the Maxam-Gilbert method, and 4) software advancements for sequence interpretation and analysis [454]. Due to its advantages, the Sanger method for many years was the dominant DNA sequencing method prior to NGS. The demand for higher throughput, largely fueled by the Human Genome Project, led to automation and parallelization of Sanger-based instruments across many laboratories, and contributed to a 13 years effort to completing the first human genome sequence in 2004 [456].

3.2.1 Next-generation sequencing

The need for greater throughput sequencing of large genomes at lower cost triggered the development of many second-generation or 'next' generation methods which have become the mainstay in biomedical research and diagnostics. NGS enables sequencing of all nucleic acid types, including DNA, mRNA, and small RNAs and long non-coding RNAs as well as methylated DNA using either whole genome, whole transcriptome, or targeted approaches [457]. Second-generation sequencing or NGS methods fall into three main categories: sequencing by hybridization, sequencing by ligatation, and sequencing by synthesis (SBS). As we have utilized Illumina instrumentations, which implement SBS, for generating sequencing data in **Studies I**, **II**, **IV**, and **V**, only SBS technology will be covered here.

Sequencing by synthesis

SBS has been the dominant NGS method for the past decade. SBS is an evolution of Sanger sequencing, combining continuous synthesis cycles on a solid substrate in massively-parallel fashion, fluorescent dye-tagged reversible dideoxy terminators, and advanced imaging (Figure 3.1). Upon addition of a single dNTP tagged and terminating base at each of millions of growing

sequencing clusters, a series of stitched together images are taken of the entire surface (so called flowcell), identifying the base added at each cluster coordinate one base at a time. Via flow chemistry, the terminator is reversed, and the next dye terminator is added, and new images are taken. In such a way, each sequencing cycle adds one of the four possible sequence bases at each cluster position, and thus millions of templates can be sequenced simultaneously. This approach was pioneered by Solexa, which was later acquired by Illumina and commercialized at scale [458]. SBS methods typically are limited to relatively short reads of 50-400 bases. Although SBS may have a higher per-base error rate of individual reads, the overall sequence accuracy can be very high, now exceeding that of Sanger sequencing when sufficient coverage is obtained [459]. Depending on the setup, DNA sequencing (DNA-seq) or RNA-seq can be performed, as well as specialized variants for epigenetic profiling and identification of specific bound fragments (e.g., chromatin immunoprecipitation sequencing, CHIP-seq; assay for transposase-accessible chromatin with sequencing, ATAC-Seq; and many others). The most common DNA-seq setups include whole genome sequencing (WGS), whole exome sequencing (WES), or targeted sequencing using hybrid capture-based panels. The key distinction among these methods lies in the specific parts of the genome they target for sequencing. Although SBS methods have evolved through a variety of approaches [460], driven by the need for faster, cheaper, and more accurate sequencing methods, the Illumina SBS implementation is described here, which was used in this thesis **Studies I**, **II**, **IV**, and **V**.

Library preparation

As depicted in Figure 3.1, the first step is to prepare the nucleic acid (NA; from either DNA or RNA) library. After the NA is purified, it is fragmented into smaller pieces primarily by one of three methods: chemical cation-mediated, sonication/acoustic shearing, or enzymatic tagmentation. Chemical fragmentation involves the use of chemicals, typically divalent cations like magnesium or zinc, to induce breaks in the DNA. The process is influenced by factors such as pH and temperature, allowing for controlled cleavage that produces fragments of specific sizes [461]. Chemical fragmentation can be precisely controlled to yield fragments of desired sizes, typically ranging from several hundred to a few thousand bps, making it particularly suitable for standardized protocols in large-scale studies like SCAN-B. Sonication/acoustic DNA fragmentation is a multi-step method that uses high-frequency ultrasonic sound waves to physically shear NA randomly into smaller pieces, typically ranging from a few hundred to a few thousand bp in length. These fragments have variable strand overhanging ends, therefore enzymes such as T4 DNA

polymerase (which has both polymerase and exonuclease activity) are used to create blunt ends. In a A-tailing process, a single adenine (A) nucleotide is then added to the 3' ends of the blunt DNA fragments, preparing them for adapter ligation. Right and left adapters are ligated by T7 DNA polymerase and T4 DNA ligase, providing a priming site for sequencing, and allowing the NA fragments to attach to the sequencing flow cell.

Tagmentation is a more recent and quicker method that uses transposases to simultaneously fragment the DNA and add adapters [462]. The generated NA is then size-selected to obtain fragments within a desired range for optimal sequencing performance, followed by PCR in order to both amplify the amount of NA available for sequencing and to add additional sequences, adapters, required for binding to the sequencing flow cell and for indexing, if samples were multiplexed. Adapters contains three key segments – a sequence that binds to the flow cell, a unique barcode for sample identification (index), and a sequencing primer binding site – are attached to NA fragments. These adapters allow for multiplexing, enabling simultaneous sequencing of multiple samples (up to 96) by grouping reads with the same index during analysis. The sequencing then takes place in a flow cell, where fragmented DNA hybridizes to complementary oligonucleotides, facilitating the sequence by synthesis process.

Bridge amplification

Within the flow cell, the NA fragments undergo bridge amplification to form distinct, clonal clusters on the flowcell surface. NA fragments with left and right adapters attach to complementary oligonucleotides on the flowcell. DNA polymerase synthesizes a complementary strand for each attached NA fragment. These newly synthesized strands bend and attach to complementary oligonucleotides on the flowcell, forming a bridge-like structure. A new round of polymerization begins on these bent strands, creating double-stranded NA clusters. The double-stranded NA is denatured, allowing each single strand to independently form new bridges. The process is repeated multiple times, resulting in millions of dense clusters of identical DNA sequences. This step ensures sufficient signal for accurate sequencing in subsequent steps and quality control.

Clonal sequencing on a flowcell

During clonal amplification in Illumina sequencing, all reverse strands are removed from the flow cell, leaving only forward strands. A primer specific to the adapter's primer binding site attaches to each forward strand. DNA

polymerase adds fluorescently tagged deoxynucleotides (dNTPs) to the NA strand (Figure 3.1). Each dNTP carries a distinct fluorescent label and a reversible blocking group. The use of blocking group or reversible terminators on the fluorophore is the key process, ensuring that only one base is added per cycle. It enables the sequencer to pause and record each base addition before continuing with the next cycle, allowing for the precise identification of which base is incorporated at each physical position on the flowcell. The sequencer uses four-color chemistry, with each of the four bases (A, T, C, G) tagged with a unique fluorescent dye. After each dNTP addition, the machine takes an image to record the base added, based on the color emitted by the fluorescent tag. After the base incorporation is recorded, the fluorescent label and blocking group are chemically removed, or 'cleaved', from the NA strand. This process is repeated with the flow of new dNTPs over the flowcell, continuing the cycle of base addition, imaging, and cleavage. This meticulous process of sequencing one base at a time and recording each incorporation allows for highly accurate determination of the NA sequence in a massively-parallelized manner [463]

A notable technical issue in Illumina sequencing is the decline in base-call accuracy with longer read lengths, largely attributed to “dephasing noise.” This occurs when nucleotides are incorrectly added or omitted, or when the blocking groups are not properly removed during sequencing cycles. As these errors accumulate over multiple cycles, they result in a cluster of NA strands of varying lengths. This variation leads to reduced signal clarity and impacts the precision of identifying each base, particularly at the 3′ ends of the reads [464].

In sequencing experiments, “depth of coverage” or “read depth” is a crucial parameter, reflecting the average number of times each base in the target region is sequenced. For DNA-seq, a common target depth might be 30X, implying that, on average, each base in the targeted genomic region is covered by 30 independent sequencing reads. This uniform distribution of reads across the genome, however, can be affected by various biases. For example, GC-rich regions may exhibit skewed coverage due to library preparation and sequencing biases. RNA-seq, on the other hand, fundamentally differs from DNA-seq in terms of read distribution. In RNA-seq, the distribution of reads is not uniform but is instead roughly proportional to the expression levels of the RNA in the sample. This means that highly expressed genes will have more reads compared to those that are less expressed. Due to this non-uniform distribution, the average sequencing depth in a RNA-seq dataset is not a particularly informative measure. Instead, the total number of reads generated is a more relevant metric for RNA-seq, as it indicates the extent to which the sample's RNA content has been sequenced, or “how deep” the sequencing overall has been performed for a sample. This total read count is critical in ensuring sufficient data to accurately quantify and compare the expression

levels of different genes or transcripts within the sample [465,466]. For SCAN-B, Illumina HiSeq 2000 or NextSeq 500 sequencers were utilized, aiming for approximately 25-30 million read-pairs per sample.

The different sets of RNA-seq data from SCAN-B that were included in **Study I**, **II**, **IV**, and **V** were prepared using a customized dUTP, TruSeq, or NeoPrep protocol, and sequenced in paired-end mode [434]. Briefly, starting from 1μg total RNA, mRNA is purified using poly-DT DynaBeads (Thermo Fisher Scientific). Following isolation, the mRNA is subjected to zinc-mediated fragmentation (Ambion), resulting in approximately 240bp fragments which are isolated using Zymo spin columns (Zymo Research). Using this fragmented mRNA as input, first-strand cDNA synthesis is performed by adding random hexamer primers, reverse transcriptase, and dNTPs. During cDNA synthesis, the second strand is synthesized by adding polymerase and dNTPs and dUTP instead of dTTP (i.e. uracil is incorporated into the second strand). The rest of library preparation continues as usual, with adapter ligation and amplification. However, before PCR amplification, an enzyme called uracil-DNA glycosylase (UDG) is used to degrade the second strand (the one containing uracil). During the PCR amplification stage, the uracil-containing strand is not amplified because UDG has removed the uracil bases causing this DNA strand to fragment and no longer able to serve as a template (in a process known as selective amplification). As a result, only the original, first-strand cDNA is efficiently amplified, which corresponds to the mRNA's coding strand. Hence the final library is “stranded” because it retains the information about which strand of DNA the mRNA is transcribed from. In SCAN-B, the customized dUTP method is favored because it provides high strand specificity and preserving this information can be crucial for alignment and correct gene annotation, especially in areas where genes are overlapping or closely adjacent on opposite strands, and for understanding gene regulation [463,465].

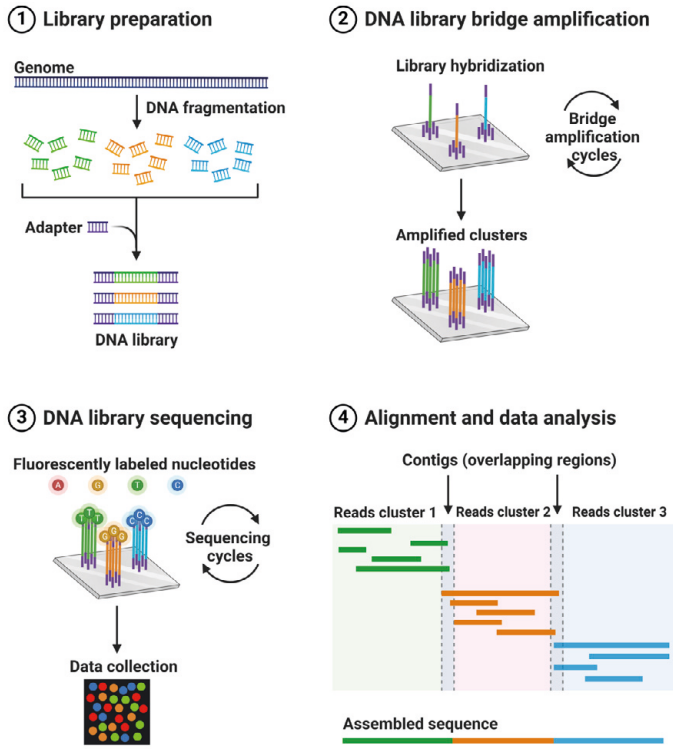

Figure 3.1: Illumina sequencing workflow. 1. DNA library is prepared by fragmenting DNA (or RNA), then ligating adapters to the fragments. 2. Library is hybridized to the solid flowcell, which contains anchored baits to the adapter sequence, and clusters are formed by bridge amplification. 3. The millions of discrete clusters are sequenced through a series of flowed chemical cycles which add fluorescent-labeled reversible terminating nucleotides for imaging, building the complementary sequence to the template base by base. 4. Obtained images are analyzed and encoded into sequence information, which are then aligned and analyzed. Created with BioRender.com.

3.2.2 DNA microarrays

DNA microarrays, introduced in the mid-1990s, were once the primary method for analyzing RNA expression, SNPs, or methylation genetic markers, enabling the simultaneous measurement of thousands of markers in a single hybridization [464,457]. These chips carry hundreds of thousands of short oligonucleotide probes on a solid surface, each designed to match a specific DNA segment or "feature." Briefly, for gene expression studies, mRNA is converted into fluorescently labeled cDNA, which binds to complementary probes on the chip and unbound cDNA is washed off. The intensity of

fluorescence, measured by a scanner, reflects the amount of hybridized cDNA, which is proportional to gene expression.

Study III utilized Affymetrix U133A DNA microarray that has been widely used in gene expression studies [448]. This microarray platform allows for the comprehensive analysis of gene expression levels across approximately 22,000 cDNA probes simultaneously [465]. Although the U133A microarray is a powerful tool for gene expression analysis, it has certain limitations compared to RNA-Seq technology as discussed below.

DNA microarrays require prior knowledge of the genomic sequences to design specific probes, limiting their ability to detect novel transcripts, gene fusions, or other genomic elements that are not represented on the array. RNA-seq, on the other hand, can sequence and quantify all RNA molecules present in a sample without prior knowledge, offering a more comprehensive view of the transcriptome, including the detection of novel transcripts, splice variants, and gene fusions. RNA-seq has a broader dynamic range of detection, as it can identify both very lowly and highly expressed genes with greater precision. Microarrays, in contrast, have limitations in detecting lowly expressed genes and distinguishing between high expression levels due to the saturation of signal [466]. RNA-seq offers more accurate quantification of gene expression levels than microarrays [434,467,468]. The quantification with RNA-seq is based on the number of reads aligning to a gene or transcript, providing a relatively direct measure of its abundance, while microarray signals can be affected by cross-hybridization, the quality of the probe, and background noise. Since it is a sequencing technology, RNA-seq can be used to identify SNVs, indels, and other sequence variations within the transcribed regions, offering insights into genetic variability and its impact on gene function [469–472,445]. Microarrays do not typically provide information on sequence variability [460,473].

3.2.3 RNA-seq bioinformatics

Data processing and analysis are crucial components in a RNA-seq workflow, significantly impacting the reliability and interpretation of results. The RNA-seq bioinformatic data analysis for **Studies I**, **II**, **IV**, and **V** generally followed the standard SCAN-B analysis pipeline [474]. The pipeline is initially integrated within the BASE laboratory information management system [475,476], which was further enhanced by the Reggie extension package [474], providing a structured approach to data handling. The principal steps specific for the RNA-seq analysis is described below.

Quality control and data cleaning: FASTQ conversion

During sequencing, the raw sequencing data output by Illumina's instruments is in binary base call or BCL format. As depicted in Figure 3.2. this data is converted into the sample specific FASTQ format, a standardized ASCII text-based format for storing both the nucleotide sequences and their corresponding quality scores, which is then used in subsequent analysis steps.

Quality control and data cleaning: demultiplexing and read trimming

In high-throughput sequencing, the sequencing library generated from a sample is often tagged with a unique barcode sequence (indices) and then many libraries are pooled together in a single sequencing run – a process known as multiplexing. Demultiplexing is the process of sorting the sequencing reads back into separate samples based on these barcode indices, to distinguish reads from different samples in a pooled sequencing run. IlluminaBasecallsToFastq from the Picard suite was used to demultiplex SCAN-B sequencing runs [477]. Sequencing reads may also contain adapter sequences, low-quality bases, or other non-biological sequences that can interfere with accurate alignment and analysis. Trimmomatic was used in SCAN-B for this task as it can handle paired-end data [478].

Quality control and data cleaning: quality filtering

Quality filtering involves removing reads that are too short or of overall low quality or unwanted reads such as PhiX lambda phage DNA that is spiked-in as a control and ribosomal RNA (rRNA) sequences were removed to ensure that only high-quality, reliable data specific to mRNA is used in downstream analysis, which is crucial for accurate mapping and interpretation. In SCAN-B, reads that aligned to the PhiX phage genome and rRNA were removed using the University of California Santa Cruz (UCSC) RepeatMasker track [479] and Bowtie2 [480].

The quality score in a FASTQ file, called Phred quality score, which reflects the probability of a base call error, is used to guide this filtering. The Phred score is calculated using the formula: $Q = -10 \log_{10}(P)$, where Q is the quality score and P is the probability of an error ($P = 10^{(Q/-10)}$. So, a Phred score of 20 translates to a 1% chance of an error (1 incorrect base call in 100), and a score of 30 means a 0.1% chance of an error. After trimming and filtering, the FastQC [481] tool was used in SCAN-B to assess the quality of the processed reads.

Read mapping or alignment

Next, filtered high quality reads are aligned or mapped to a reference genome, or in the case of RNA, to a reference transcriptome. For RNA-seq, this process is challenging due to the complex nature of RNA, including splicing, reads spanning exon-exon junctions and RNA editing. Thus, aligners designed for RNA-seq are "splice-aware" and can align reads across exon-exon junctions, a crucial feature for accurate RNA-seq alignment. HISAT2 was used for this purpose in SCAN-B [482] and the reads were aligned to the human genome reference GRCh38/hg38 together with transcript annotations from the UCSC knownGenes track [483]. The output of RNA-seq alignment is typically a file in SAM/BAM format, containing the alignment information for each read. Assembling the aligned reads into transcripts was done by StringTie2 [484] in SCAN-B (Figure 3.2).

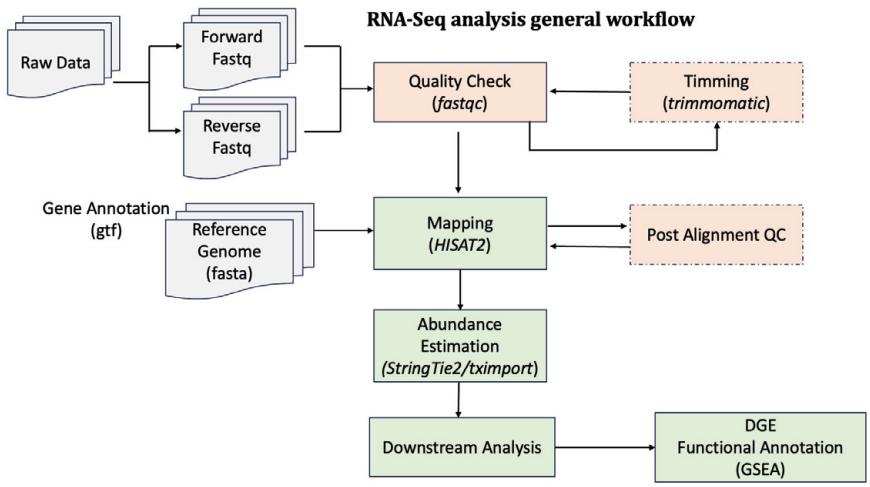

Figure 3.2: General schema of RNA-seq computational analyses utilized in this thesis. Starting from quality control and data cleaning, read mapping or alignment, quantification of gene expression and downstream data analysis. Abbreviations: DGE: Differential Gene Expression; GSEA: Gene Set Enrichment Analysis.

Quantification of gene expression and normalization

The primary goal of RNA-seq is to estimate gene and transcript expression levels, which involves various computational strategies and normalization methods.

Count-based methods

This straightforward approach involves counting the number of reads that align to each gene or transcript using a gene transfer format (GTF) file, which provides the coordinates of genes and exons in the genome. However, for most applications, simply counting raw reads for expression comparison across samples is inadequate due to influences like varying transcript lengths, total number of reads (library size), and potential biases from fragmentation, PCR amplification, and other technical issues.

Normalization methods

RPKM (Reads Per Kilobase of exon model per Million mapped reads) and is paired-end variant, FPKM (Fragments Per Kilobase of exon model per Million mapped reads), are normalization methods that account for gene length and total read count. They enable comparisons of expression levels within a sample but are not ideal for between-sample comparisons. For single-end reads, RPKM and FPKM are equivalent. While both RPKM and FPKM normalize for gene length and sequencing depth, TPM (transcripts per million) adjusts the normalized read counts against the total transcriptome size, ensuring that the sum of all TPM values in a sample is constant at one million. This makes TPM particularly useful for comparing expression levels between samples because it provides a consistent scale. The distinction between comparing gene expression within a single sample and across different samples has sometimes led to confusion. While adjusting for gene length is less critical when comparing the same gene across samples with minimal batch variability, it is crucial for accurately comparing different genes within a sample, as longer genes naturally attract more reads. The key difference between FPKM and TPM lies in how the total reads are incorporated into the normalization process. For both FPKM and TPM, each gene's read count is initially normalized by the gene length. However, FPKM then normalizes this result by the total reads in the sample, effectively adjusting for sequencing depth. In contrast, TPM further normalizes the length-adjusted read counts by the sum of all these adjusted counts across the entire sample before scaling up to a million. This approach makes TPM values generally more comparable across different samples because it adjusts for the varying composition of the transcriptome, not just the raw number of reads. This ensures that the sum of all TPM values in a sample is precisely one million, facilitating more accurate comparisons of transcript abundance across samples.

Despite normalization, these measures can still suffer from biases introduced by factors like library preparation methods, varying transcriptome sizes, low abundance transcripts, high biological variability, and nonlinear relationships.

TPM calculation are as follows:

1. Calculate RPK (Reads Per Kilobase) for each gene:
   RPK = Number of reads mapped to the gene / Length of the gene in kilobases
2. Scaling Factor = (Sum of all RPK values) / 1e6
3. Calculate TPM for each gene:
   TPM = (RPK for the gene) / (Scaling Factor)

TPM is generally more comparable between different samples but is not entirely free from biases [485]. For instance, while TPM is suitable for comparing the same gene across samples, it's less suitable for comparing different genes within the same sample because TPM normalization, by design, makes the sum of all TPMs in a sample equal to a constant (1e6), potentially obscuring significant differences in expression levels between genes. For studies I, II, III and V, TPM was used to reduce skewing of the data and ease fold-change calculations and comparisons. Briefly, in **Study I**, tximport [486] was used to summarize transcript-level abundance estimates from StringTie [484,487] v1.3.3b into gene-level TPMs from .ctab files. TPM was calculated using formula:

$$TPM = \frac{read\ count}{transcript\ length\ in\ kb}\ X \frac{1{,}000{,}000}{sum\ of\ all\ RPK}$$

Here, the read count is derived from the coverage value and the transcript length, and the RPK is calculated. Then, each RPK is scaled by the sum of all RPKs across the dataset, normalized to one million. After calculating TPM for each transcript, tximport aggregates these values to the gene level by summing the TPMs of all transcripts that belong to the same gene. The output is a matrix of TPM values for each gene across samples, which can be used for downstream differential expression analysis (DGE) [220]. Since StringTie is continuously evolving and improving, one of these enhancements includes the ability to directly output TPM values in a .tsv format. In **Study II** and **V**, TPM matrix was generated directly from .tsv files [301]. The TPM values were transformed to $\log_2$ i.e. (TPM+ 0.1). The addition of 0.1 was done to avoid infinite values, since $\log_2(0)$ is undefined.

Batch correction

Batch correction is applied to eliminate variations in data that arise from differences in experimental conditions, rather than from the biological variables of interest. It's crucial for ensuring that the results reflect true biological effects and not artifacts introduced by differing sample processing methods or protocols used. For **Study II** and **V** batch correction was performed

by normalizing *ERBB2* gene expression values between SCAN-B library preparation protocols (dUTP, TruSeq and NeoPrep) by mean centering as previously described in Staaf et al. [353]. In brief, dUTP was used as a reference protocol, *ERBB2* gene expression values were normalized to dUTP by mean centering $\log_2$ transformed TPM to derive gene specific mean differences between library protocols so that TruSeq and NeoPrep values were converted to dUTP like as baseline, dUTP remained unchanged. The calculated protocol differences for *ERBB2* were used as correction factor for converting gene expression between protocols. For example, protocol difference (for *ERBB2* gene) for dUTP and TruSeq, i.e., dutp.truseq.diff, was calculated using:

$$\text{dutp.truseq.diff} = \text{mean}(\log_2(\text{dUTP} + 0.1)) - \text{mean}(\log_2(\text{TruSeq} + 0.1))$$

TruSeq can be transformed to dUTP-like using:

$$\log_2(\text{dUTP} + 0.1) = \log_2(\text{TruSeq} + 0.1) + \text{dutp.truseq.diff}$$

Similarly, protocol differences were also calculated to be able to transform *ERBB2* gene expression between dUTP and NeoPrep library protocols.

Differential expression and functional analysis

DGE analysis, that is identifying genes or transcripts that show statistically significant differences in expression across different conditions, treatments, or sample categories, is a common analytical goal of gene expression profiling. Further, assigning biological meaning to the differentially expressed genes, such as using gene set enrichment analysis (GSEA) is also performed that could provide insights into the underlying biological processes or pathways that are active or suppressed in different condition by focusing on groups of functionally related gene sets.

In **Study I**, DGE was performed using the limma-voom package [488] in R. A linear model was fit to log TPM values using lmFit and empirical Bayes smoothing was applied using eBayes with the trend = TRUE option, without voom transformation, as suggested by Gordon Smyth and colleagues [489]. Voom was not used, because it expects count data and not normalized data like TPM. For estimation of relative library sizes, arrayWeights() was used to improve the robustness of the analysis [490]. The results of DGE analysis comprise of log-fold change ($\log_2$FC), P-values, and adjusted P-values to control for false discovery rate using methods like Benjamini-Hochberg.

GSEA was performed using the fgsea package [491] in R as well as WebGestalt [492]. Both fgsea and WebGestalt assume that the gene sets are predefined and that the ranking of genes reflects biological relevance. Gene list was by the

negative log of adjusted P-value multiplied by the sign of the log_2FC. The sign of the log_2FC (positive for up-regulated, negative for down-regulated) was preserved. This ranking method prioritizes genes that are both highly significant and have large changes. Figure 6 in **Study I** was generated using WebGestalt, chosen for its enhanced visual clarity compared to fgsea.

3.3 Polymerase Chain Reaction

Studies II and **V** utilize droplet digital PCR polymerase chain reaction (ddPCR). To understand ddPCR, first PCR must be understood. PCR is an efficient technique for amplifying a specific DNA fragment from a diverse DNA mixture. Invented by Dr. Kary Mullis, PCR is a powerful tool that exponentially increases a targeted DNA sequence, making it a highly sensitive method for DNA analysis [493]. It requires minimal DNA, which can be obtained from various sources like blood, skin, hair, saliva, or microbes, which is then exponentially amplified by thermal cycling through a repeating series of temperature changes to denature DNA, anneal primers, and allow for polymerase extension [494]. Due to its sensitivity and versatility, PCR is widely used in genetic research, forensics, and medical diagnostics.

Essential components in almost all PCR reactions include:

Template: The nucleic acid sample that contains the target region to be amplified. The source of the nucleic acid can be DNA, RNA, or cDNA.

DNA polymerase: a thermostable DNA polymerase I enzyme originally isolated from *Thermus aquaticus* (hence named Taq polymerase) by Chien et al. in 1976 [495]. It is used for synthesizing new DNA strands because it remains active during the high temperatures used in PCR.

Primers: two short single stranded DNA sequences, typically of 18-24 bases long, that are complementary to the ends of the target DNA region, marking the starting point of DNA synthesis. Primers are crucial because DNA polymerase adds nucleotides to a double-stranded DNA in the 5' to 3' direction.

Deoxynucleoside triphosphates (dNTPs): These are the building blocks of DNA. During PCR, dNTPs (A, T, G, C) are incorporated into the new DNA strand by the DNA polymerase.

Buffer solution and cations: This provides the optimal chemical environment for the DNA polymerase to function effectively. Bivalent

cations usually magnesium (Mg) or manganese (Mn) ions and monovalent cations, potassium (K) ions, which are also part of the reaction mixture.

Widespread use and standardization of PCR components have led to the commercial availability of pre-mixed solutions known as a "super mix" or "master mix." These mixes simplify the PCR setup process and ensure that the concentrations of each component are ideal and maintained consistent across experiments. This uniformity is crucial for reproducibility in PCR results, and also in reducing the preparation time significantly. The machine used for carrying out PCR is known as a thermal cycler. This device plays a critical role in the PCR process by precisely controlling the temperature cycles that are necessary for DNA amplification.

Setup: The mixture of PCR components, which includes the DNA template, primers, DNA polymerase, dNTPs, buffer solution, and necessary ions, is prepared in a PCR tube or a well of a 96-well or larger well plate. This setup allows for multiple samples to be processed simultaneously, which is particularly useful for high-throughput experiments.

Thermal cycling: The test tube or plate is placed in the thermal cycler, which then undergoes a series of temperature changes, each of which facilitates a different step of the PCR process, as shown in Figure 3.3:

- *Denaturation*: The DNA is heated to a high temperature (typically around 95°C) to separate the double-stranded DNA into single strands.
- *Annealing*: The temperature is lowered (usually to between 50°C and 65°C) to allow the primers to bind (anneal) to their complementary sequences on the DNA strands.
- *Extension/Elongation*: The temperature is raised slightly (typically to 72°C) to enable the DNA polymerase to synthesize new DNA strands by adding dNTPs to the annealed primers.

These three steps are repeated typically for 5 to 40 cycles, depending on the application and desired amount of DNA amplification. Each cycle in theory doubles the amount of target DNA, leading to an exponential increase in the number of DNA copies ($\sim 2^n$ copies). After the cycles are completed, the PCR product, which is the amplified DNA, can be used for various applications downstream such as DNA sequencing, cloning, or analysis. The thermal cycler is essential for PCR because it provides the precise and rapid temperature changes required for the different stages of the reaction. Modern thermal cyclers are highly automated and can be programmed for specific PCR protocols, making the process efficient and reproducible.

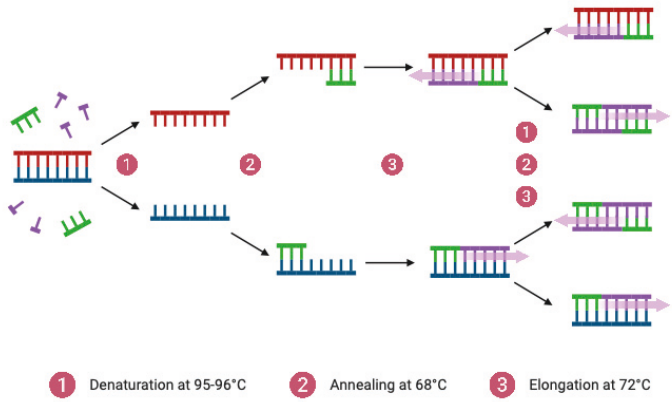

Figure 3.3: Basic schema of PCR technique. DNA polymerase is used to synthesize millions of new DNA copies via a template DNA strand (red and blue). 1. Denaturation separates the double-stranded DNA into single strands 2. Annealing allows the primers (green) to bind to their complementary sequences on the DNA strands 3. Elongation where DNA polymerase synthesizes new DNA strands by adding dNTPs (purple) to the annealed primers. Created with BioRender.com.

3.3.1 Digital PCR

Digital PCR (dPCR) is a third-generation advanced form of PCR that provides a highly accurate and sensitive method for measuring the quantity of specific DNA or RNA sequences in a sample. The capability of dPCR to accurately detect and absolute quantification across a wide dynamic range of target abundances has led to its fast-growing applications in diagnostics. The key concept of digital PCR is to randomly partition the target molecules into reaction compartments, and then to dichotomously score each compartment as either positive or negative (0 or 1, hence digital) based on whether the compartment contains the target or not. A compartment is dichotomized based on presence of a fluorescent label, whose intensity increases as the PCR product abundance increases. Once above a threshold value (above noise), the compartment is scored as positive. In droplet digital PCR (ddPCR), the compartments are comprised of microdroplets of water:oil emulsions or micelles.

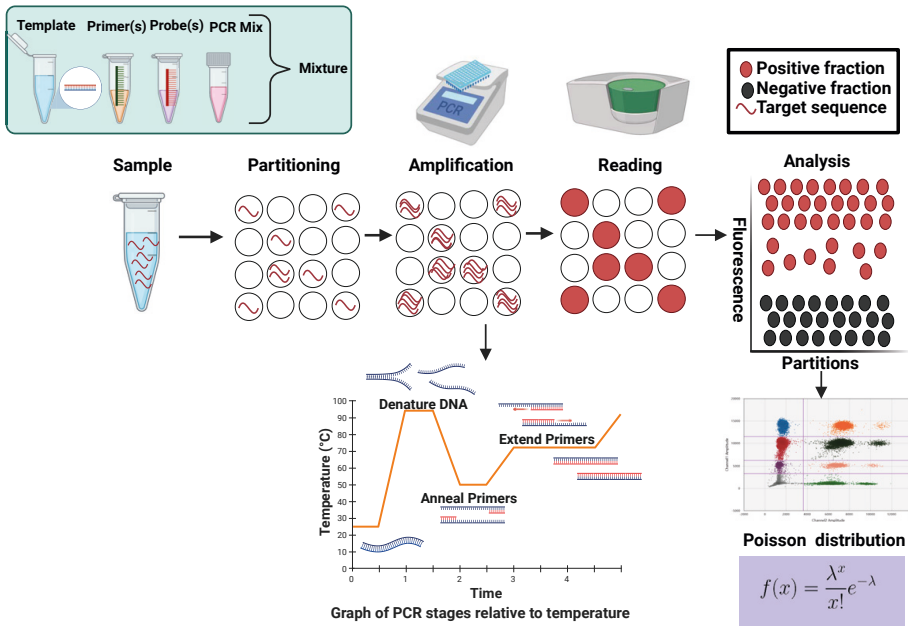

Figure 3.4: Schematic representation of digital PCR workflow. 1. Partitioning of sample 2. Amplification of target sequences inside the nano droplets. 3. Reading of each droplet for fluorescence 4. Analysis by counting number of fluorescent droplets and Poisson approximation (formula) is used to precisely quantify DNA or RNA targets within a sample. Created with BioRender.com.

Partitioning of sample: In ddPCR, the sample containing the target DNA or RNA is divided into thousands (~20,000) tiny droplets, each acting as an individual micro-reactor (Figure 3.4). This partitioning is achieved using a specialized device that creates a water:oil emulsion, with each droplet containing a fraction of the sample. Importantly, the target molecules are randomly partitioned into the droplets.

PCR inside droplets: PCR is then carried out in each of these droplets independently and in parallel. Like standard PCR, ddPCR involves the cycles of denaturation, annealing, and extension. However, due to the partitioning, each droplet will either contain amplified target sequences (positive or 1) or not (negative or 0), depending on whether the target sequence was present in that droplet.

Fluorescence detection: Each droplet is analyzed for fluorescence after the PCR. Droplets containing the amplified product will fluoresce due to the cleavage of sequence-specific fluorescent probe linkers, so called hydrolysis probes, via

the 5' nuclease activity of the polymerase. The intensity and number of droplets is measured using a flow cytometer-like system.

Unlike qPCR, which provides relative quantification compared to standards or reference genes, ddPCR allows for absolute quantification of nucleic acids without the need for external standards. It does this by counting the number of positive and negative droplets (thus total droplets), knowing the average volume of each droplet and the volume of input sample, and applying Poisson statistical analysis to determine the precise concentration of target molecules that were present in the original sample to yield the measured ratio of positive, negative, and total droplets. ddPCR is known for its high sensitivity and precision, especially in detecting low abundance targets, rare genetic variants, and in samples with complex backgrounds.

Poisson approximation

The Poisson approximation in dPCR is a technique used to precisely quantify DNA or RNA targets within a sample. The method involves dividing the sample into numerous partitions, then uses statistical analyses to determine the likelihood original concentration of the target that would yield a particular distribution of positive partitions given a total number of partitions generated. By tallying the partitions that test positive, negative, and thus total number of partitions, and utilizing the Poisson distribution, dPCR accurately measures the concentration of target DNA or RNA. This process is underpinned by a binomial distribution scenario, where 'm' represents the number of target molecules and 'n' the number of partitions. Molecules are distributed across partitions in a manner analogous to placing balls into boxes, resulting in partitions that are either positive (contain at least one molecule) or negative (contain none). The probability 'p' of a partition being positive is calculated using the formula $p = 1 - (1 - 1/n)^m$, indicating the odds of a partition not being empty. The Poisson approximation, particularly relevant when 'n' is large and the ratio of 1/n is very small, predicts the number of partitions containing specific quantities of target sequences using the probability $p \approx 1e^{(-\lambda)}$, where $\lambda = (m/n)$ is the average number of molecules per partition. This principle is central to the operation of ddPCR technologies [496,497].

Digital PCR has been extensively used to measure genetic imbalances, or CNV, that result from the deletion or amplification of genomic regions or locus. In CNV analysis, the copy number of a locus relative to another is the relevant information. In **Study II**, dPCR absolute quantification is used to determine the copy number of *ERBB2* gene (with alleles of SNP rs1058808) relative to a copy number stable region on chromosome 2p13, and also CEP17, located on

chromosome 17. In **Study V**, a novel multiplex ddPCR assay was developed that simultaneously detected and quantified HER2 expression levels and two alleles of the SNP rs1136201 within *ERBB2*, alongside two reference controls, *GNB1* and *PUM1*.

ddPCR vs. other PCR technologies

dPCR offers distinctive advantages in the replication, amplification, and accurate quantification of NA, providing highly accurate and precise quantification which is particularly useful in applications such as detecting low-abundance mutations, precise gene quantification, and applications where absolute quantification is necessary such as in **Study II** and **V**. Most significant benefits of dPCR are i) *absolute quantification*: dPCR does not require external standards or reference curves for quantification. It provides absolute counts of target NA molecules without the need for calibration against standards or references, which is often necessary in qPCR [498]. ii) *High precision and sensitivity*: in ddPCR, the partitioning of the sample into around twenty thousand droplets per sample of 20 µl allows precise measurement of low-abundance targets with better reproducibility and less susceptibility to PCR efficiency variations. iii) *Reduced impact of inhibitors*: ddPCR's partitioning approach can dilute PCR inhibitors present in the sample, potentially allowing for more accurate results with complex matrices than qPCR. iv) *High resistance to contamination*: The droplet format minimizes the risk of contamination because the reactions are isolated in tiny droplets. v) *Detection of small fold changes*: ddPCR is extremely sensitive to small fold changes in gene copy numbers, making it ideal for detecting slight variations in gene expression or small increases in mutation load. vi) *Simpler interpretation*: Since ddPCR provides direct absolute quantification, it avoids the need for complex calculations required in qPCR to determine relative gene expression or copy number variation [499].

Summary of ddPCR protocol to detect ERBB2 amplification

In **Study II**, DNA was isolated from fresh tumor samples collected at primary surgery. The genomic DNA isolated from the *ERBB2*-amplified SK-BR-3 cell line served as a positive control, and NS12911, a reference DNA sample, was utilized as a normal copy number sample.

The ddPCR assay was designed for multiplexed analysis, allowing for the simultaneous quantification of *ERBB2* alleles along with a control region *CEP17* and a copy number stable region near cytoband 2p13.1 (CNS-2p13.1). Each 20 µl PCR reaction, prepared with 10 ng of DNA and 4X ddPCR supermix according

to the manufacturer's instructions (Bio-Rad), was processed in an Automated Droplet Generator (Bio-Rad) to form emulsified droplets. These were then thermocycled in a 96-well plate with a protocol involving an initial denaturation at 95 °C for 10 minutes, followed by 40 cycles of 94 °C for 30 seconds and 60 °C for 60 seconds, a final extension at 98 °C for 10 minutes, and a hold at 4 °C. The plates were subsequently read using a Bio-Rad QX200 Droplet Reader, with data analyzed using QuantaSoft software. In total, nine separate runs were performed, each containing control samples, and an average of 19,738 droplets were analyzed per reaction (with SD of 3492 and lower quartile, median, and upper quartile of yielded droplets was 19923, 20869, and 21606, respectively).

Summary of ddPCR protocol to measure ERBB2 mRNA expression

In **Study V**, a multiplex ddPCR assay was developed to quantify *ERBB2* mRNA expression levels in breast tumors. RNA was isolated from fresh tumor samples and cell lines, then converted to cDNA by reverse transcription using the Bio-Rad iScript Synthesis Kit employing oligo(dT) and random hexamer primers. A novel multiplex ddPCR assay was designed for quantifying *ERBB2* and two alleles of SNP rs1136201, along with reference genes *GNB1* and *PUM1*, using designed primers that target long introns. *PUM1* was selected as a reference gene as it was demonstrated as the most stable gene among 32 frequently used reference genes [500]. *GNB1* was selected as a stable gene with low coefficient of variation calculated using FPKM of each gene among 9206 SCAN-B samples. The assay was validated on 19 tumor and 2 non-tumor breast cell line samples, and measurements were made for 1242 invasive breast cancer samples that received no neoadjuvant treatment from within the SCAN-B cohort. For ddPCR, a 20 µl reaction with 5 ng of cDNA was prepared, droplets were generated, and PCR was conducted with 95 °C for 10 min, 40 cycles of 94 °C for 30 s and 59 °C for 60 s, 98 °C for 10 min, and hold at 4 °C. The temperature ramp rate was 2°C/s for all steps. Digital PCR was performed with negative control (DNA from cell line MCF12A) and a no template control (containing elution buffer from Qiagen) within each run. Analysis was performed using Bio-Rad's QuantaSoft software. Dilution series of HCC1954 cell line RNA were tested to demonstrate linearity of quantification, and 21 cell lines were analyzed to show concordance between independent The Human Protein Atlas RNA-seq TPM and *ERBB2* expression. *ERBB2* expression was calculated as *ERBB2* copies/sqrt(*GNB1* copies**PUM1* copies).

3.4 Statistical Analysis

Statistical testing allows one to draw conclusions whether a specific null hypothesis can be explained or rejected by the data at hand. This process yields a P-value, representing the probability of obtaining the observed results, or more extreme ones, under the assumption that the null hypothesis is true. Although debatable, in practice, a P-value of less than 0.05 is commonly interpreted as indicating a significant difference [501]. However, it's crucial to understand that P-values do not measure the size of an effect but rather the probability under the assumption of no effect (null hypothesis) of obtaining a result equal or more extreme than what was observed, and they are significantly influenced by the size of the sample. In other words, minimal effect sizes observed in very large sample sizes can still have small P-values.

The choice of statistical tests, whether parametric or non-parametric, depends largely on the data's structure and distribution. Parametric tests, suitable for data that approximates normal distribution, include methods like Student's t-test or the Chi-squared test. On the other hand, non-parametric tests like the Mann-Whitney U-test or Fisher's exact test are used for data not following a normal distribution [502]. These tests evaluate the probability that any observed difference between groups could occur by chance. Although a P-value of 0.05 is typically used as a threshold for statistical significance, its interpretation should be contextual; a P-value of less than 0.05 is not an absolute indicator for rejecting the null hypothesis. Additionally, the likelihood of encountering a "significant" P-value increases with the number of tests performed, necessitating corrections for multiple hypothesis testing.

The confidence interval (CI) is another critical statistical measure, providing a range within which the variable of interest likely falls, with a specified level of certainty (e.g., 95%). This interval helps address the risk of sampling error, offering a perspective analogous to that provided by a P-value of 0.05.

In **Studies I-V**, continuous variables were compared using Student's t-test when data was normally distributed and for non-normally distributed data, Kruskal–Wallis non-parametric test (for significant difference between groups) as well as Wilcoxon rank sum test (for multiple pairwise comparisons between groups) were used, and categorical variables were compared using Chi-square test or Fisher's exact test. Spearman rank correlation was used to determine correlations between expression levels of genes in **Study I** and **III**. To evaluate significant differences in the clinicopathological variables between two groups, Mann Whitney U test (for continuous variables) and Fisher's exact test (for categorical data) were used in **Study I**, **IV**, and **V**. All statistical analyses were performed in R.

Survival Analysis

Survival analysis is a biostatistical method used in cancer research to study the time from a significant starting point (such as cancer diagnosis in this thesis) to an important endpoint, which could be e.g., death, disease progression, or relapse. The Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) initiative has provided guidelines for defining these endpoints in cancer research [503]. Commonly used endpoints include overall survival (OS), relapse-free survival (RFS), and relapse-free interval (RFI). OS refers to the time until death from any cause, while RFS is the time until a patient experiences a tumor relapse or death. RFI, on the other hand, is specifically concerned with the time until a locoregional or distant relapse.

The Kaplan-Meier (KM) method is a fundamental tool in survival analysis [504]. It estimates the survival function for a group of patients, for instance, those stratified by the status of a biomarker as in **Study I**, **II**, and **IV**. The KM method computes the fraction of patients still alive (or without an event such as relapse) at any given time after a starting event, such as cancer diagnosis. A KM plot visualizes this information, showing the proportion of patients without an event over time. Patients are 'censored' from the plot if they are lost to follow-up or if the event hasn't occurred by the end of the study. Diverging curves on a KM plot can indicate that a certain factor, like biomarker status, has prognostic or predictive significance. The statistical relevance of differences in survival curves is commonly assessed using the log-rank test [505].

To quantify the effect of (multiple) variables on survival, Cox proportional hazards models may be used [506]. These models can be univariable, examining the effect of one variable as seen in KM plots, or multivariable, adjusting for multiple potential confounders. The hazard ratio (HR) derived from Cox models quantifies relative risk. For example, a HR of 1 indicates no difference in risk between groups, a HR of 1.5 signifies a 50% increased risk, and a HR of 0.5 implies a 50% reduced risk. The interpretation of HR depends on the proportional hazards assumption, which posits that the factors affecting the groups have a constant impact over time [507]. Non-proportionality, suggested by crossing survival curves, can affect the model's validity. To ensure the appropriateness of the Cox model, it's crucial to test for non-proportionality using Schoenfeld residuals, as was used in **Study I-V**.

Survival analysis was performed by KM and Cox regression survival analyses using the package survival and survminer in R. OS and RFS as endpoints in **Study I**, **II**, **IV**, and **V**, distant metastasis-free survival (DMFS) was used as an endpoint in **Study III** and **V** along with OS. Proportional hazards assumptions were checked graphically by Schoenfeld residual plots.

4. Results and Discussions

Study I: Clinical associations of *ESR2* (estrogen receptor beta) expression across thousands of primary breast tumors.

Unlike ERα, which has been well-characterized and extensively studied for decades and is the target of standard anti-endocrine therapies in HR+ BC, the role of ERβ (encoded by *ESR2*) is less clear, has been controversial, and appears to be context-dependent [253,508]. While some studies suggests a link between ERβ presence and improved breast cancer outcomes, potentially by moderating ERα's tumor-promoting effects [509,510], inconsistencies and challenges in accurately detecting ERβ at protein level have plagued earlier studies, largely owing to the unreliability of available antibodies [511,512]. For example, discrepancies between mRNA and protein levels of ERβ have been noted, with events like post-transcriptional regulatory mechanisms likely influencing ERβ expression [513,514]. This underscores the complexity in assessing ERβ's role due to the challenges in detecting the protein reliably, even in cell lines once thought to express it.

To address these challenges and to better understand the potential role of ERβ across a large cohort of primary breast cancers, our study aimed to quantify *ESR2* mRNA expression and explore its correlation with clinical features and patient outcomes. We utilized RNA-seq to analyze gene expression across a substantial cohort of 3,207 primary invasive breast tumors, drawing from a broad, population-based sample. This approach bypasses the issues of antibody reliability, providing a robust assessment of *ESR2*'s expression levels in breast tumors.

ESR2 mRNA expression and its association with prognosis

In this study, *ESR2* mRNA expression (TPM values) across both SCAN-B (n=3207) and TCGA (n=1089) cohorts were generally low and there was a slight inverse correlation between *ESR1* and *ESR2* expression, indicating differing expression patterns based on clinical ERα status (Figure 1 in Paper I). By categorizing the samples into tertiles for *ESR2* expression to capture its non-linear distribution across the sample population, we identified that *ESR2*-high was most prevalent in the basal subtype and ERα-negative tumors in general (with ~47% of ERα-negative and ~31% of ERα-positive cases classified into

the *ESR2*-high category, Table 1 in Paper I). In the SCAN-B dataset, *ESR2* expression was highest in normal-like tumors, decreasing through basal-like, HER2-enriched, luminal A, to lowest in luminal B. Conversely, *ESR1* showed the opposite pattern, peaking in luminal B and lowest in basal-like tumors. Age analysis revealed *ESR1* expression increased with patient age, whereas *ESR2* remained stable across all ages (Figure 2 in Paper I).

High *ESR2* expression correlated with improved OS in patients, particularly those receiving endocrine therapy (logrank test p=0.03; Figure 3D) and in TNBC (logrank test p=0.01; Figure 4F). No significant association was found between *ESR2* expression and relapse-free interval (RFI). Multivariable Cox regression analysis, adjusting for potential confounding factors such as age, tumor size, lymph node status, and grade, confirmed high *ESR2* expression as a significant positive prognostic factor for OS with a hazard ratio (HR) of 1.34 (95% CI 1.06–1.32; p=0.01) in the full cohort (Figure 5),. In the endocrine therapy group, low *ESR2* expression was not significantly associated with OS (HR 1.24; p=0.1). However, in TNBC, low *ESR2* expression significantly increased the risk of poor outcomes, with an HR of 2.0 (95% CI 1.25–3.23; p=0.004). A significant association of *ESR2* expression was not found with RFI as an endpoint for this analysis.

Our findings from the SCAN-B cohort indicating high *ESR2* expression's association with improved OS were validated in the TCGA breast cancer dataset. Despite TCGA's bias towards larger, higher-grade tumors, *ESR2* mRNA showed low expression similar to what was observed in the SCAN-B cohort. *ESR1* and *ESR2* exhibited a weak inverse correlation in both datasets. The distribution of *ESR2* expression across molecular subtypes in TCGA mirrored SCAN-B, with higher expression in normal-like, basal-like, and HER2-enriched tumors compared to luminal subtypes. However, No significant differences in *ESR2* expression across molecular subtypes were found in TCGA, contrasting with SCAN-B. Analysis of OS and RFI in TCGA, stratified by treatment received and clinical subgroups, did not show a consistent association with *ESR2* expression, except in the ERα-negative (p=0.02, logrank test) and HER2-positive subgroup (p=0.03, logrank test), where high *ESR2* was linked to significantly improved outcomes (Figure S2 in Paper I).

Differential gene expression (DGE) and GSEA analysis

DGE analysis between tumors with high versus low *ESR2* expression, while controlling for ERα's influence (by performing the analysis separately within the ERα-positive and -negative subgroups), revealed distinct upregulation of genes associated with immune responses in both ERα-positive and -negative

groups. This analysis, applying stringent criteria for identifying differentially expressed genes ($\log_2$FC ≥ 1.5 for up-regulated genes and $\log_2$FC ≤ −1.5 for down-regulated genes), uncovered a significant upregulation of genes involved in B-cell activation, immune response signaling, and other processes linked to immune system modulation in *ESR2*-high tumors. These findings suggest a more pronounced immune landscape in *ESR2*-high tumors, potentially contributing to their improved prognosis.

Next, gene set enrichment analysis (GSEA) further reinforced the notion of an activated immune environment in *ESR2*-high tumors, with enriched biological processes related to immune system activities, including B and T cell activation and leukocyte cell-cell adhesion (Figure 6 in Paper I). The presence of genes involved in these processes suggests a robust immune-mediated tumor surveillance in *ESR2*-high cases, possibly underpinning the observed survival benefits. This immune-centric view posits that high *ESR2* expression may foster a tumor microenvironment more conducive to immune infiltration and activation, thereby enhancing immunogenicity and potentially improving patient outcomes through more effective immune-mediated tumor suppression. Alternatively, the measured *ESR2* expression may primarily be coming from the immune compartment of the analyzed tissue, and thus be a proxy for immune infiltration.

According to Reese et al., ERβ is implicated in upregulating cystatins, a family of secreted proteins that are known to interfere with canonical TGFβ signaling pathways, which are crucial for tumor progression and metastasis in TNBC [515]. TGFβ signaling typically plays a role in immune evasion by suppressing immune cell activity within the tumor microenvironment [516]. By downregulating this pathway, cystatins indirectly promote a more active and vigilant immune surveillance against tumor cells. This mechanism highlights a potential pathway through which high *ESR2* expression could contribute to a more immunogenic tumor microenvironment, enabling better recognition and destruction of cancer cells by the immune system. The suppression of TGFβ signaling by cystatins could mitigate the metastatic phenotype of TNBC, thereby enhancing the efficacy of immune-mediated tumor suppression in *ESR2*-high tumors. This insight provides a possible molecular basis for the observed correlation between high *ESR2* expression and improved clinical outcomes, underlining the potential of *ESR2* in being a modulator of tumor-immune interactions in our study.

In summary, our comprehensive analysis revealed that while *ESR2* expression in primary breast tumors is generally low, elevated levels are linked to enhanced survival outcomes. Despite the technical challenges associated with mRNA-based quantification methods, such as distinguishing between tumor

and stromal/immune expression and navigating issues with template amplification and RNA editing, our findings underscore the potential of mRNA quantification to provide insights into *ESR2*'s role in breast cancer. This approach offers a complementary perspective to antibody-based assays, which have produced inconsistent results, thereby enriching our understanding of ERβ's involvement in breast cancer pathology and its impact on patient prognosis.

Study II: Quantitative digital PCR measurement of *ERBB2* copy number is predictive of outcome in early breast cancer patients treated with adjuvant trastuzumab.

The *ERBB2* oncogene, known as HER2 and located on chromosome 17q12, plays a crucial role in primary breast cancer progression through its function in cell growth and proliferation, and is amplified in approximately 15-20% of cases [266,517]. Recognized for promoting aggressive tumor growth and conferring poor prognosis, HER2 amplification has directed the development of targeted therapies such as trastuzumab, which has become a symbol of personalized therapy and precision medicine, and has significantly improved the clinical outcomes for patients with HER2-positive breast cancer [518,519]. However, despite these advancements, a subset of HER2-positive patients still face relapse [520], highlighting the ongoing need for refined diagnostic tools to better predict treatment responses.

ERBB2 DNA copy number assay development

Our study developed a novel multiplex ddPCR assay for precise quantification of *ERBB2* gene amplification, utilizing reference controls CEP17 and CNS-2p13.1, to assess *ERBB2* status in 909 primary breast cancer samples (Figure 1 in Paper II). This assay's validation revealed its high reproducibility and accuracy, correlating well with clinical HER2 assessments for IHC and ISH, thereby offering a reliable diagnostic tool for stratifying HER2 status. The assay showed high reproducibility and measurements in line with expectations for both the Coriell NS12911 normal human DNA, with mean copy numbers closely aligning with the expected normal copy numbers, and the *ERBB2*-amplified SK-BR-3 cell line, matching well with literature-reported estimates (with mean CEP17 CN ~3, mean *ERBB2* CN ~20 and mean *ERBB2*/CEP17 ratio of 6.5) (Supplementary Figure 1 in Paper II). This underscores the assay's accuracy in quantifying *ERBB2* gene copies, demonstrating its potential as a reliable tool for HER2 status assessment in breast cancer diagnostics.

ERBB2 ddPCR measurements and clinical HER2 evaluations

To determine thresholds for *ERBB2* CN and *ERBB2*/CEP17 ratio, we first divided the 909 breast tumors into balanced 70% training and 30% validation groups. Initially, we conducted a ROC analysis exclusively on the training set to establish the most effective thresholds for discerning *ERBB2* ddPCR status, focusing on achieving high sensitivity and specificity. This approach allowed us to refine our criteria for *ERBB2* CN and *ERBB2*/CEP17 ratio, ensuring robust predictive accuracy. Satisfied with the thresholds determined from the training set, we applied these criteria to the validation group (Figure 2 and Table 2 in Paper II). This sequential analysis reinforced the validity of our initial findings, demonstrating the assay's capacity to accurately classify HER2 status across different subsets of breast cancer samples.

Combined categorization of 909 cases according to ddPCR metrics aligned well with 2018 ASCO/CAP guidelines, identifying classic HER2 amplified and non-amplified groups, alongside monosomy and co-amplification subsets. This categorization revealed a high concordance between ddPCR results and clinical HER2 status, highlighting the assay's predictive value (Table 3 in Paper II). Notably, a few cases showed discrepancies between ddPCR classification and clinical HER2 status, possibly hinting at clinical false-negatives (FN) or intratumoral heterogeneity. For example, 6 cases were defined as ddPCR group 1 amplified but were clinically HER2-negative and therefore did not receive anti-HER2 therapy (Figure 3 and Table 3 in Paper II). 3 out 6 ddPCR+ patients had relapse during follow up. These discrepant cases may represent clinical HER2 FN due to technical or biological factors where the subclone analyzed at diagnosis being HER2-negative but the specimen analyzed by ddPCR in this study containing a subclone with *ERBB2* CN gain. These findings underscore the ddPCR assay's potential in enhancing HER2/*ERBB2* status assessment in breast cancer, offering insights into cases that might benefit from targeted therapies despite conventional diagnostic challenges.

Our study also included analysis of the *ERBB2* SNP rs1058808 as a genetic marker due to previous suggestions of association of this SNP with HER2 protein expression, cancer risk, and trastuzumab-induced cardiotoxicity [297,298]. Despite these prior publications, our findings revealed no correlation between rs1058808 genotype and any clinical or pathological variables, including HER2 status, *ERBB2* CN, or the *ERBB2*/CEP17 ratio. However, notably our analysis utilizing this SNP confirmed, for informative heterozygous cases, that amplifications of *ERBB2* were always monoallelic, involving either the G or C allele but not both (Supplementary Figure 2 in Paper II).

ERBB2 CN survival analysis in ddPCR cohort

We further explored the impact of *ERBB2* ddPCR CN on survival outcomes in patients with HER2+ disease treated with adjuvant trastuzumab , either as a standalone treatment or in combination with chemotherapy and/or endocrine therapy. Our dataset comprised 177 patients, within which we identified an "ultrahigh" CN group with a *ERBB2* CN above 19.7 (24 patients, 13.6%). Over a median follow-up of 5.4 years, KM survival analysis indicated significantly inferior RFS (p=0.039, log rank test) and OS (p=0.040, log rank test) among patients with ultrahigh *ERBB2* CN (Figure 4a and 4c in Paper II). Further, multivariate Cox regression analysis, accounting for all clinically significant variables from univariate analyses, identified *ERBB2* CN as an independent predictor of poor prognosis in trastuzumab-treated patients, with hazard ratios of 3.3 (95% CI 1.1-9.6; p=0.031) for RFS and 3.6 (95% CI 1.1-12.6; p=0.041) for OS (Figure 4b and 4d in Paper II). Notably, all patients in the ultrahigh *ERBB2* CN group were also found to have high Ki67 levels (p=0.016), suggesting a link between *ERBB2* CN and cell proliferation, although no significant correlations were found with other clinical or pathological features.

ERBB2 CN survival analysis in surrogate RNA-seq SCAN-B cohort

To validate our ddPCR survival analysis findings, we aimed to include a larger cohort treated with in a uniform way with trastuzumab. Therefore, we utilized RNA-seq data from the SCAN-B study, incorporating 682 consecutive patients with HER2+ disease and treated with adjuvant trastuzumab, diagnosed between 2010 and 2019, with available RNA-seq data. The median follow-up for this group was 6.8 years. By selecting the top 13.6% of cases based on highest *ERBB2* mRNA expression, paralleling the "ultrahigh" *ERBB2* CN category from ddPCR analysis, we observed similar outcomes: notably worse OS (p = 0.044), albeit without significant differences in RFS (Figure 5a and 5b). Given that *ERBB2* mRNA levels serve as a proxy for DNA CN, we further refined the "ultrahigh" group by adjusting the threshold to encompass 17.5% of cases based on mRNA expression. This adjustment underscored the initial findings, with the ultrahigh *ERBB2* mRNA group showing a trend towards reduced RFS (p = 0.083) and significantly poorer OS (p = 0.021) (Figure 5c and 5d). Multivariable analyses affirmed that having four or more positive lymph nodes was a critical factor for RFS (p = 0.001) and both ≥ 4 positive lymph nodes and ultrahigh *ERBB2* status significantly impacted OS (p = 0.001 and p = 0.039, respectively), aligning with ddPCR-based classifications.

Poorer survival outcomes observed in the *ERBB2* ultrahigh group prompted speculation on the underlying biological contributions. One plausible

explanation is the incomplete engagement of HER2 receptors by therapeutic antibodies in cases of extreme HER2 overexpression, leaving residual signaling still in an active state. Additionally, the ultrahigh HER2 expression may be a biomarker for the presence of other aggressive tumor characteristics, such as immune evasion capabilities or a heightened mutation rate, facilitating rapid development of resistance to trastuzumab. It is noteworthy that the temporal pattern of survival curves, diverging only after 2-3 years post-surgery (a period during which the patient is receiving trastuzumab), implies initial responsiveness to therapy in these ultrahigh *ERBB2* patients, followed by eventual relapse perhaps due to tumor aggressiveness or activation of compensatory signaling pathways downstream of HER2. While these hypotheses shed light on the complex dynamics between HER2 expression levels and treatment response, they underscore the need for further investigation into the molecular underpinnings of HER2-driven tumorigenesis and resistance mechanisms.

Our findings on the prognostic significance of 'ultrahigh' *ERBB2* CN, as determined by ddPCR, were recently corroborated by an independent study. This validation confirms that patients with the highest levels of *ERBB2* gene amplification or *ERBB2* mRNA have a significantly worse prognosis (RFS) compared to those with intermediate levels (RFS with HR = 2.7, p = 0.003 for both ERBB2 CNA and *ERBB2* mRNA) [521]. This parallel finding not only reinforced our results but also underscores the need for further clinical investigations to refine HER2-targeted therapies for this group.

In summary, we have successfully designed and validated a robust, cost-effective ddPCR assay for the precise quantification of *ERBB2* CN in breast cancer, demonstrating high accuracy, specificity, and sensitivity. This assay offers a significant advancement in the stratification of HER2+ BC patients, identifying those with "ultrahigh" HER2 amplification who exhibit worse outcomes despite adjuvant trastuzumab treatment. Our findings highlight the complexity of HER2-driven BC and suggest the need for tailored therapeutic approaches for patients with extreme levels of *ERBB2* overexpression. This work paves the way for further research into the mechanisms underlying differential responses to HER2-targeted therapies, with the potential to improve personalized treatment strategies for breast cancer patients.

Study III: CITED1 as a marker of favourable outcome in anti-endocrine treated, estrogen-receptor positive, lymph-node negative breast cancer.

This study aimed to explore the role of CITED1, a transcription coactivator, as a biomarker for predicting the response to anti-endocrine therapy and recurrence in breast cancer. This investigation builds on previous research that identified CITED1's involvement in estrogen-dependent transcription and its influence on mammary gland development [522,523].

To achieve these objectives, *CITED1* mRNA and protein levels were analyzed in human breast cancer cell lines and several independent tumor datasets in GOBO database [448] in a characterized tissue microarray (TMA) as well as in the TCGA breast cancer dataset. Additionally, MCF7 cell lines were engineered with stable overexpression of CITED1 to investigate its impact on downstream transcriptional programs. In the GOBO dataset, *CITED1* mRNA was found to be selectively expressed in ER+ luminal-molecular subtype cell lines and tumors (Figure 1a and 1b in Paper III). Higher levels of *CITED1* were correlated with improved outcomes in patients treated with tamoxifen, especially notable in ER+, lymph-node-negative (ER+/LN−) patients. This correlation became evident after five years, highlighting *CITED1*'s potential role in predicting long-term anti-estrogen treatment response (Figure 1e and 1f in Paper III). IHC analysis of tissue microarrays further supported CITED1 protein's association with favorable outcomes in tamoxifen-treated, ER+ patients (Figure 2a and 2b in Paper III). Interestingly, while a general favorable response to anti-endocrine treatment was observed in a larger TCGA dataset (Figure 2c and 2d in Paper III), the tamoxifen-specific effect noted in GOBO dataset was not replicated (Figure S1d). In vitro, MCF7 cells overexpressing CITED1 showed an increased expression of AREG but not TGFα, suggesting a specific role of ERα-CITED1 mediated transcription in the response to anti-endocrine therapy (Figure 3 in Paper III).

In summary, CITED1 expression was found to be associated with favorable prognosis in ER+, tamoxifen-treated breast cancer patients, particularly in the ER+/LN− subset, suggesting its utility as a prognostic biomarker for anti-endocrine therapy response. The differential expression of AREG in CITED1-overexpressing cells underlines the importance of specific ERα-CITED1 mediated pathways in sustaining long-term response to anti-endocrine treatment. While this study confirms CITED1's potential role in breast cancer prognosis and treatment response, further research, including larger datasets and prospective studies, are needed to fully establish its diagnostic and prognostic value. The findings also highlight the broader therapeutic potential of targeting coregulators like CITED1 in disease management.

Study IV: The Sweden Cancerome Analysis Network – Breast Initiative: a 10-year interim update of the ongoing multicenter multiomic study to improve personalization of breast cancer care.

Breast cancer, with its significant heterogeneity, presents challenges in treatment due to the current limitations in predicting outcomes and responses to therapy. The SCAN-B initiative, launched in 2010, aims to integrate advanced NGS-based diagnostics into clinical routines to address these challenges, with the goal to improve personalized care through enhanced patient stratification [434,435]. Despite advances in treatment, breast cancer's complexity demands more precise biomarkers for better management strategies. SCAN-B's approach leverages RNA-sequencing for real-time clinical implementation, aiming to refine prognostic and therapeutic decisions. The study adhered to ethical standards and involved comprehensive patient enrollment across Southern Healthcare Region, with sample collection and processing protocols, from across seven hospital centers-namely Malmö, Lund, Helsingborg, Kristianstad, Halmstad, Växjö, and Karlskrona. In this Paper, we have summarized the SCAN-B cohort for its first 10 full calendar years: from initiation in 2009 through the end of 2020, SCAN-B enrolled 16381 eligible breast cancer patients, collecting a vast array of samples, including tumor samples and baseline and follow-up blood samples (Figure 2 in Paper IV). RNA-seq was performed on a large cohort of 9323 out of 9915 tumor samples (94 %) employing a high-throughput protocol to analyze transcriptomes and identify transcriptomic alterations relevant to breast cancer [474–476]. Initial analyses have demonstrated the feasibility of molecular subtyping and mutation screening from RNA-seq data, offering insights into the genomic landscape of breast cancer in a population-based setting [352,445]. Although the analysis is still in progress at the time of writing this manuscript, the statistics provided herein represent an initial summary of all accrued breast cancer patients (n=13940) in SCAN-B.

Real world treatment profile analysis

Treatment profiles from patients who were accrued and have tumor specimen available delineated the pre-neoadjuvant and post-adjuvant therapeutic interventions, including chemotherapy, immunotherapy, endocrine therapy, and radiotherapy (Figure 4 in Paper IV). This categorization provides as overview of the treatment sequence (from pre to post transitions) as well as the timing of treatments in breast cancer management in clinics in real time. For instance, neo-adjuvant chemotherapy and HER2 targeted therapy reflects the use of therapies intended to reduce tumor size or address metastasis before surgical intervention. Adjuvant chemotherapy and HER2 targeted

therapy suggests use of these interventions following initial surgical or primary treatments in the aggressive tumor types.

Survival analyses

Survival analysis using OS, RFS as endpoints was performed across five key clinical subgroups: HER2 positive with estrogen receptor positive (HER2pERp), estrogen receptor positive and HER2 negative with lymph node positive (ERpHER2nLNp), estrogen receptor positive and HER2 negative with lymph node negative (ERpHER2nLNn), HER2 positive with estrogen receptor negative (HER2pERn), and triple-negative breast cancer (TNBC) (Figure 5). This analysis demonstrated the distinct prognostic outcomes associated with each clinical subgroup. HER2 positive subgroup showed more favorable survival outcome compared to TNBC, reflecting the impact of targeted HER2 therapies. In contrast, patients with TNBC exhibit a more challenging prognosis, underscored by lower OS and RFS outcomes ($p < 0.0001$).

OS analysis in ER+ HER- LN+ patients that received endocrine therapies, including tamoxifen and aromatase inhibitors (AI), chemotherapies including anthracyclines and taxanes, showed variable outcomes as compared to the patients who did not receive any endocrine therapy. ER+ HER2– LN+ tumors with either no therapy or chemotherapy only, where estrogenic signaling was not targeted, had poor RFS and OS ($p < 0.0001$; Figure 6), whereas for LN– cases this was not as apparent: patients receiving no therapy had slightly better OS than patients receiving endocrine-only, which is likely because most of these patients receiving no therapy were clinical stage 1, whereas those receiving endocrine-only were generally of higher stages. HER2+ patients who have been treated with targeted anti-HER2 therapies (including trastuzumab and pertuzumab) distinctly showed a significant improvement in both OS and RFS ($p < 0.0001$ for both) compared to those who did not (Figure 7A-B). This data reinforces the importance of HER2 targeted treatments in the clinical management of HER2 + tumors. TNBC patients who received chemotherapy, including anthracycline and docetaxel showed improved OS and RFS ($p < 0.0001$; Figure 7C-D) as compared to the untreated group reflects the effectiveness of chemotherapy in not only prolonging life but also in sustaining periods free from cancer recurrence, a critical outcome for TNBC patients given the disease's propensity for early recurrence and absence of targeted treatments.

In summary, SCAN-B initiative demonstrates the feasibility of large-scale, population-based cohort and RNA-sequencing analysis for breast cancer patients, making this project the largest study of its kind. This approach has

and continues to expedite the discovery, validation, and clinical application of innovative biomarkers and molecular tests that have the potential to transform into clinics in future [301].

Study V: Characterization of HER2-low and -ultralow using digital PCR for *ERBB2* expression profiling of breast cancer in breast cancer.

Breast cancer treatment has significantly advanced and is continuously evolving with HER2-targeting therapies and other antibody-drug conjugates (ADCs), especially for cancers expressing HER2. Approximately 60% of traditionally defined HER2-negative breast cancers express low levels of HER2, a new subclass now known as "HER2-low" [524]. HER2-low cancers are defined as i) HER2 non-amplified but ii) having lowered *ERBB2* expression levels, and iii) responding to new antibody-drug conjugates such as trastuzumab deruxtecan (T-DXd) [525]. Since HER2-low tumors in the metastatic treatment setting have been shown as targetable in DESTINY-Breast04 trial, this change in the clinical landscape poses significant challenges related to HER2 assessment. Traditional HER2 clinical classification is by IHC (with scores of 0, 1+, 2+ or 3+) can be subjective and varies in precision, underscoring the need for a more objective approach [526]. This highlights the need for more precise and reliable methods to define "HER2-low" status [318].

Hence, this study aimed to refine HER2 classification at the gene expression level using ddPCR to measure *ERBB2* mRNA expression across a large consecutive series (n=1242) of breast tumor samples from 1223 BC patients in SCAN-B. In this study we developed and validated a novel single-reaction multiplex ddPCR assay for quantification of *ERBB2* expression levels, including two alleles of SNP rs1136201 within *ERBB2*, simultaneously with two reference control regions, *GNB1*, and *PUM1* (Figure 1 and Primers for *ERBB2, GNB1 and* PUM1 are listed in Table 2 in Paper V). This approach aimed to offer a more objective and quantitative method for assessing *ERBB2* mRNA expression, hopefully allowing us to define more clearly HER2-low and -ultralow breast cancer. The ddPCR results with IHC scores and consensus among pathologists were also compared.

ddPCR assay validation

ERBB2 mRNA ddPCR assay performance was validated using a twofold RNA input serial dilution series using the HER2-amplified breast cancer cell line HCC1954, beginning with 4 ng RNA input down to 0.0156 ng RNA input. Multiplex mRNA measurements for *ERBB2*, *GNB1*, and *PUM1* showed high

linearity across the dilution series (Figure 2 in Paper V). The quotient of *ERBB2* compared to each housekeeping gene individually as well as to their geometric mean was calculated, showing high consistency of the gene expression ratio across the entire dilution series.

Additionally, the multiplex *ERBB2* mRNA gene expression ddPCR assay's performance was validated and analyzed across 21 cell lines, including 19 breast cancer cell lines and 2 normal mammary epithelial cell lines (Table 1 in Paper V). The ddPCR quantifications were compared to publicly available RNA-seq measurements for the same cell lines and showed very high reproducibility and consistency in the quantification between laboratories and measurement methods. After categorizing the cell lines into three distinct groups based on their HER2 status, cell lines with known HER2 amplification, cell lines with normal HER2 copy number, and the normal epithelial cell lines with a normal HER2 copy number, the distribution of *ERBB2*-to-reference gene ratios within each group showed the expected patterns of ERBB2 expression ratios within each cell line group, with some variability within the HER2 amplification group (Figure 3 in Paper V). This distinction not only reaffirms the ddPCR assay's precision in quantifying *ERBB2* expression but also highlights its capacity to differentiate between varying levels of *ERBB2* expression.

Although in this manuscript some analyses are still ongoing, such as survival analyses and final delineation of cutoffs between HER2 mRNA expression groups, in summary, this ddPCR assay accurately quantifies *ERBB2* mRNA levels and is easily scalable for analysis of many tumor samples. This assay may enable a more objective assessment compared to current IHC methods. This study therefore may contribute to significant progress in the field of breast cancer diagnostics and treatment, with the potential to elucidate further the molecular characteristics of HER2-low breast cancer.

5. Conclusions

Study I

We have comprehensively assessed *ESR2* expression across the most extensive population-based breast cancer cohort SCAN-B, and validated it in TCGA, detailing its correlation with various clinicopathological features and patient outcomes. Our analysis revealed that while *ESR2* mRNA is generally expressed at low levels in primary breast cancer, it is notably more prevalent within ERα-negative breast cancer subtypes. Importantly, elevated *ESR2* expression is significantly associated with improved survival outcomes, especially among patients receiving endocrine therapy and those diagnosed with TNBC. The upregulation of genes associated with immune activation and surveillance in *ESR2*-high tumors was also observed, indicating an association to immune-mediated gene expression patterns. This study adds critical insights to the ongoing discourse on ERβ/*ESR2*, emphasizing its potential impact on prognosis and inviting further investigation into this relatively underexplored receptor.

Study II

Our study validates the effectiveness of ddPCR for *ERBB2* CN analysis in breast cancer, demonstrating high concordance with clinical HER2 evaluation by IHC and ISH. Our ddPCR assay achieved an accuracy of 93.7% and 94.1% in training and validation cohorts, respectively, with positive and negative predictive values of 97.2% and 94.8% for identifying HER2 amplified cases. Importantly, we identified an "ultrahigh" *ERBB2* CN subgroup associated with significantly poorer survival in patients with HER2+ disease treated with trastuzumab (HR of 3.3 for RFS and 3.6 for OS). These findings suggest that ddPCR could serve as a rapid, cost-effective initial screening tool in conjunction with IHC and before ISH, particularly for complex cases. The presence of the ultrahigh *ERBB2* CN group highlights the need for additional confirmatory studies as well as clinical trials to explore tailored treatments that could improve outcomes for these higher-risk HER2+ patients.

Study III

This study highlights CITED1's potential as a prognostic biomarker for enhancing the prediction of anti-endocrine treatment outcomes and recurrence in breast cancer, particularly within the ER+/LN- patient subgroup. We found that elevated *CITED1* mRNA levels are significantly associated with improved survival outcomes in patients treated with tamoxifen, notably in luminal subtype tumors. This association underscores CITED1's role in modulating ERα-dependent transcription, which was further evidenced by the selective amplification of AREG in CITED1-overexpressing MCF7 cells, highlighting its specific involvement in sustaining anti-endocrine therapy response. While broader dataset analyses from TCGA confirmed the positive response to anti-endocrine treatments, the specific enhancement of tamoxifen response was notably prominent in the ER+ tamoxifen-treated cohort. These findings suggest CITED1 as a biomarker could help tailor anti-endocrine therapies, warranting further research.

Study IV

This study shows that over a decade, SCAN-B initiative has successfully demonstrated the feasibility of conducting large-scale, population-based RNA-sequencing analysis for breast cancer. During the first 10 full calendar years of SCAN-B, from late 2010 through the end of 2020, this initiative has accrued nearly 14,000 patients, achieving comprehensive molecular profiling for more than 9300 breast tumors that is representative of the broader patient population. This extensive dataset has enabled not only the deep exploration of breast cancer biology but also the development and clinical integration of novel biomarker assays. Implementation of molecular diagnostics, via the integration of RNA-seq based SSP models into clinical practice in Region Skåne in 2021 exemplifies how SCAN-B's findings are being translated into actionable clinical tools that enhance personalized treatment strategies. This model serves as a benchmark for comprehensive cancer treatment centers and even nation-states worldwide, highlighting the potential benefits of continued expansion and collaboration where translational research is closely integrated and symbiotic to clinical implementation and patient care.

Study V

Our study underscores the potential utility of a novel ddPCR assay for refined HER2/*ERBB2* mRNA classification in invasive breast cancer, especially distinguishing the HER2-low subgroup. This assay was validated using 21 normal and neoplastic breast epithelial cell lines and 1,242 breast tumor RNA

samples from within the SCAN-B project have been analyzed, demonstrating high concordance with HER2 IHC results. The assay's capability for multiplex detection of *ERBB2*, *GNB1*, and *PUM1* genes allows for precise and accurate HER2/*ERBB2* expression quantification. It showed robust performance across a diverse range of RNA inputs and maintained consistent expression ratios critical for accurate gene expression studies. Moreover, the assay's reliable quantitative measurement in diverse cell lines highlights its sensitivity and specificity, and in the future potentially could be used in clinical decision-making for guiding use of targeted therapies such as the ADC T-DXd. This *ERBB2* mRNA ddPCR assay is expected to enrich our understanding of HER2 expression dynamics and support application of new anti-HER2 therapies in breast cancer treatment.

6. Future Perspectives

Acknowledging the rapid advancements and shifting paradigms in BC research and treatment driven by precision diagnostics and large-scale initiatives like SCAN-B, it is crucial to explore further enhancements and applications in this field. This thesis represents an effort to enhance diagnostic precision by exploring the capabilities of dPCR in precisely quantifying the key biomarker *ERBB2* and refining the classification of HER2 expression levels. It also investigates the roles of *ESR2* and *CITED1* in influencing patient outcomes. The insights gained contributes towards the potential of precise molecular diagnostics to personalize treatment strategies. Moving forward, it is important to expand upon these findings by integrating multi-omics data, incorporating liquid biopsies, enhancing machine learning (ML) applications for data analysis, and developing robust clinical trials to validate and refine these diagnostic assays [527].

SCAN-B

Due to the complexities in the formation and development of BC, the study of mechanisms underlying cancer has gone beyond just one field of the omics arena. As cancer involves complex genomic, epigenomics, transcriptomics, proteomic, and metabolic alterations, multi-omics approaches are crucial for comprehensive insights. As the field progresses, sequencing techniques are expanding to include emerging areas such as the microbiome, exposome, and immunome, further broadening our view of cancer's complex behavior. Each of these layers contributes to our understanding of BC's multifaceted nature and have the potential to provide an in-depth investigation of the tumor for a thorough characterization of tumor evolution, plasticity, heterogeneity, microenvironment, immune evasion and drug resistance [528].

Future research in the SCAN-B initiative should aim to generate, integrate, analyze, and connect different resultant data emerging from these different -omic layers that has the potential to pave the way facilitating the discovery of novel prognostic, diagnostic, and therapeutic approaches, as conceptualized in Figure 6.1. RNA-seq using Illumina SBS is currently the method of choice for molecular profiling in SCAN-B. This technique is cost-effective, accurate, supported by a wide range of analysis tools and pipelines, and invaluable for

its high-throughput capabilities and precision in sequencing short reads, it does encounters limitations. These limitations include difficulties in accurately assembling highly repetitive or complex regions of the genome, and an inability to read through long segments of DNA or RNA without interruption. This can lead to gaps or ambiguities in genomic reconstructions and an incomplete understanding of structural variations, gene fusions and detecting allele-specific expression. NGS depends on PCR amplification as a crucial step for DNA amplification. However, PCR amplification can introduce biases and errors, such as preferential amplification of certain sequences over others, base misincorporations and skewing allelic frequencies, which can skew the results and complicate data interpretation [529]. This is particularly problematic in quantifying gene expression levels or in detecting rare genetic variants. Additionally, because of errors that arise during cluster amplification, cycle sequencing, and image analysis, ~1% of bases are incorrectly identified across various sequencing platforms, including Illumina [530,531]. Therefore, high-coverage assembly is required to compensate for this high error rate, that often results in very low allele frequency mutations being lost in the inherent sequencing noise.

Third-generation sequencing (TGS) technologies, like those offered by Pacific Biosciences' single-molecule-real-time (SMRT) and Oxford Nanopore's platforms, can mitigate some of these challenges by enabling direct sequencing of single DNA molecules without the need for PCR amplification. These approaches reduce amplification biases, better resolves complex genomic regions, and can provide more accurate characterization of structural variations, chimeric transcripts, and transcript isoform identification. The ability to read through long segments of DNA or RNA in a single, continuous read in real time spanning large genomic regions, fully capturing long transcripts in transcriptome analysis, and identifying complex rearrangements and variants that are otherwise missed by short-read sequencing [532]. Furthermore, TGS can directly detect RNA and DNA base modifications during the sequencing process, which is pivotal for advanced epigenetic profiling [533]. Recently, owing to its capability in generating real-time sequencing data, nanopore sequencing has been adopted for clinical diagnostics, including the rapid variant detection of SARS-CoV-2 [534]. By integrating TGS with single-cell sequencing (sc-seq) technologies, understanding of individual cancer cells' heterogeneity, clonal and sub clonal evolution of tumors, their reactions to therapeutic interventions, and their contributions to disease progression can be explored with greater depth and precision.

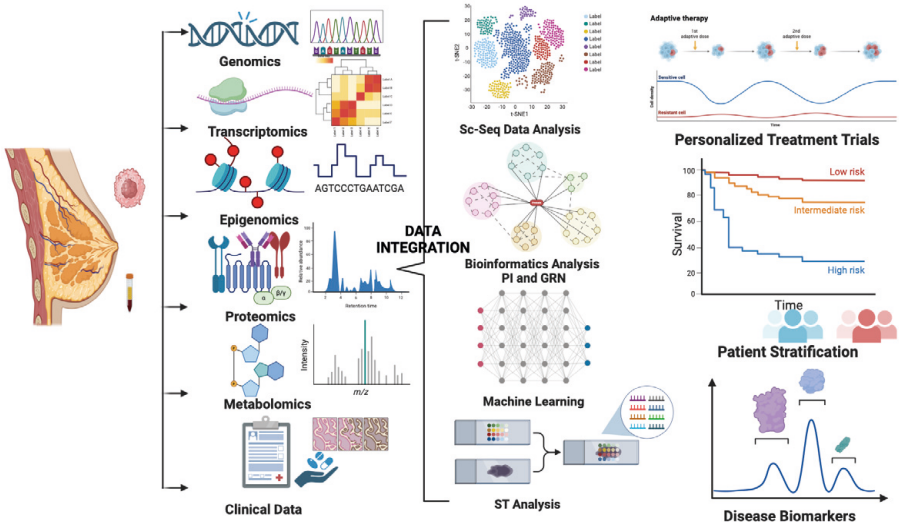

Figure 6.1: Multi-omics data integration and analysis framework. Starting from tumor tissue or liquid biopsy samples, different layers of omics data such as genomics, transcriptomics, epigenomics, proteomics, metabolomics and clinical data can be integrated and analyzed to do bioinformatics and ML analysis, adding spatial information (ST) to it to generate outcomes such as improved patient treatment monitoring, patient stratification, and identification of BC biomarkers. Abbreviations: sc-seq: single-cell sequencing; ST: spatial transcriptomics; PI: protein interaction; GRN: gene regulatory networks. Created with BioRender.com.

Although TGS technologies offer improvements for research and clinical diagnostics, they also present significant challenges. A primary concern is the relatively high error rate observed during sequencing runs, approximately 15%, making TGS less suitable for accurate detection of SNPs and point mutations, thus NGS still remains the best technology for mutational analysis [535]. Efforts are ongoing to refine the sequencing chemistry of TGS platforms to reduce these error rates and improve overall accuracy [536]. Furthermore, being relatively new, TGS technologies necessitate the development of specialized additional bioinformatics tools, pipelines and algorithms for downstream data analysis [536].

While TGS and sc-seq provides detailed insights into individual cells' transcriptomes, it loses the spatial context during tissue processing. On the other hand, spatially resolved techniques like IHC and ISH are limited by the need to pre-select targets, which constrains their utility for high-throughput exploratory analyses. Tumor-associated cell states are complex and cannot always be fully characterized by a limited set of markers or surface receptors [537]. Thus, spatial transcriptomics (ST), pioneered by Ståhl and Salmén et al.,

offers a powerful solution by providing comprehensive, transcriptome-wide expression data within its spatial context [538]. This technology is particularly effective for analyzing cellular interactions and the spatial dynamics of gene expression in the tumor stroma, bridging the gap between high-throughput RNA-seq and the necessity to maintain spatial resolution. By mapping the gene expression patterns directly to the locations of cells within a tumor, ST helps identify different tumor regions and their potential response or resistance to treatments. In BC research ST has been extensively used to define high resolution map of cellular interactions in HER2+ BC [539], analyze the transcriptional states of TNBC [540], define cancer–immune interactions and niches as determinants of Immune checkpoint blockade (ICB) benefit in TNBC [541], characterization of immune cell subsets and TILs, predictive of patient outcome and response to systemic therapy in TNBC [542], probe ITH of tumor and their TME and TME's influence on drug responses in breast cancer [543], detection of rare high-plasticity (mesenchymal transdifferentiation) BC subtypes [544], build a detailed transcriptional atlas of BC and stratified them into nine clusters ('ecotypes') with unique cellular compositions and clinical outcomes [545]. A significant challenge in treating BC is the tumor heterogeneity and development of resistance to therapies – which is also crucial for SCAN-B – can be addressed by utilization of ST. Improvements in all these technologies are enhancing the accuracy of tumor profiling, leading to better patient stratification. The integration of ST and TGS into clinical practice continues to evolve, with ongoing research focused on making these technologies more accessible and interpretable. This progression is expected to further refine treatment strategies and improve the precision of interventions in BC care, ensuring that each patient receives the most appropriate therapy based on a comprehensive understanding of their tumor biology.

As data complexity increases with the integration of various omics layers, there is a critical need to develop and refine bioinformatics pipelines and ML algorithms that can effectively analyze, predict and interpret this data. Integrating ML in the bioinformatics workflow has been shown enhance the accuracy of tumor classification [546,547], prognosis [548–550], biomarker identification [551,552] and prediction of treatment responses [553] in BC.

Precision diagnostics and clinical implications

There is a need to continuously improve diagnostic assays such as dPCR performed in **Study II** and **V**, for even better sensitivity and specificity, on a larger patient population. The future should see the development of newer, even more precise diagnostic assays that can detect low-abundance, clinically

significant variations in biomarkers and provide faster results to accelerate clinical decision-making. Clinical translation of these assays requires robust validation in FFPE tissues and liquid biopsy (ctDNA) samples, different patient populations and finally through clinical trials. For example, in **Study II**, the separation of KM survival curves emerges after about 2-3 years post-surgery between ultrahigh and the moderate-high groups, meaning the tumors appear to be initially responsive to treatment or at least kept under control during the treatment. Therefore, assuming that ultrahigh BC relapse due to aggressiveness and resistance or downstream HER2 signaling activation, these tumors likely need to be treated differently. Keeping these speculations in mind, a clinical trial can be envisioned with different arms to try new, improved, more aggressive combination therapies including at least one HER2-targeting agent for HER2+ ultrahigh group and another arm for prolonged treatment duration or varied dosage of HER2 treatment. After establishing the results from the clinical trial, improved treatment strategies may change the way ultrahigh HER2+ group is treated, potentially improving HER+ BC patient outcomes.

Moreover, to integrate this assay into clinical routines effectively, it is possible that dPCR could supplant conventional ISH techniques and be used as a screening test earlier in the testing workup algorithm, due to its lower cost and relative ease of use compared to ISH (~$10-30/run vs. a minimum of $300 for FISH) [554–556]. Also, implementing dPCR could reduce the total number of ISH tests required, thereby decreasing false positive and false negative rates and expediting the diagnostic process without significant delays. It would be pragmatic to implement dPCR alongside the current standard of IHC. For all BC cases, both IHC and dPCR should be conducted in parallel to assess HER2 status. This dual-testing approach would leverage the sensitivity of dPCR for gene copy number alterations and the visual confirmation of protein expression provided by IHC. If our results are validated in a larger patient cohort, these hypotheses could then be tested in clinical trials, as stated above. Integrating dPCR into clinical practice represents a significant advancement in precision diagnostics, offering a faster, more cost-effective, and potentially more accurate assessment of *ERBB2* amplification (**Study II**) and HER2 expression status (**Study V**). By adopting a strategic approach that includes parallel testing, resolving discrepancies, economically feasible for widespread clinical adoption and optimizing adjuvant therapy durations and combinations based on precise genetic insights, dPCR can profoundly impact the management of BC patients.

Liquid biopsy

Liquid biopsy technologies, particularly those utilizing ctDNA, are increasingly recognized for their potential to revolutionize cancer management. These techniques offer promising avenues for early cancer detection, monitoring minimal residual disease, and assessing treatment response in solid tumors including BC [557–559]. The implementation of liquid biopsies into clinical practice as companion diagnostics depends on the success of well-designed prospective clinical studies that validate their effectiveness in enhancing patient treatment outcomes and survival rates [560].

For SCAN-B initiative, liquid biopsy projects such as the NeoCircle study and SCAN-B-rec, both of which I have been involved in, are paving the way for these advancements. The NeoCircle study focuses on using a novel ultra-sensitive tumor-informed approach based on structural variant analysis combining whole-genome sequencing and multiplex dPCR (*unpublished*). This study aims to track ctDNA dynamics during the neoadjuvant treatment of early BC, helping to identify patients with high-risk disease. Persistent detection of ctDNA after tumor resection (molecular residual disease, MRD) signals the presence of occult metastatic disease and a higher risk of disease relapse. Similarly, the SCAN-B-rec study is currently gathering plasma samples from BC patients experiencing recurrence or metastasis. This project aims to provide a deeper understanding of the molecular underpinnings of cancer recurrence through ctDNA analysis. Both initiatives underscore the growing importance of integrating liquid biopsy techniques into large-scale research frameworks like SCAN-B, enhancing the capacity to tailor treatments to individual molecular profiles and improve overall patient management in breast cancer.

Acknowledgments

Reflecting on my PhD journey, I attribute a significant part of my academic and personal growth to the guidance and support of my supervisors, colleagues, and friends.

Dr. Lao Saal, as my main supervisor, was the one who opened the door to my PhD studies. Under his guidance, I have flourished academically and scientifically. His unwavering support and keen guidance have been instrumental throughout my research. I still remember the first day he picked me up from the train station, I was full of doubts. My journey towards the completion of my PhD has been profoundly influenced by his dedication, exceptional guidance and wisdom. His steadfast mentorship and the intellectual freedom he granted me have been the key factors in my academic and personal growth. Dr. Saal's commitment transcended the bounds of typical supervision; he was a mentor in the truest sense. His ability to provide timely advice and thoughtful feedback, despite his many commitments, has been remarkable. He created an environment that fostered independent thinking and encouraged me to venture beyond my comfort zone, all while ensuring I had the support needed to succeed. His kindness and empathy, especially during the most challenging periods marked by my personal struggles and the pandemic, have been a source of great strength. He assisted with my PhD extension, promptly providing reference letters, advising on postdoctoral positions, and offering me space and understanding during difficult times. Such support was more than I could have ever expected and has left an indelible impact on my life. The dedication I have witnessed in Dr. Saal has not only inspired me but has also instilled in me a deep respect for scientific pursuits. It is his model of leadership and compassion that I aspire to emulate in my future endeavors. He is truly a gem of a person. Therefore, it is with immense respect and profound gratitude that I dedicate my work to Dr. Lao Saal. His mentorship has been one of the greatest gifts of my PhD journey, and I am honored to have had the opportunity to work under his guidance. His influence extends beyond the pages of my thesis and into the foundational fabric of my career and character.

My co-supervisor, *Dr. Niklas Loman*, has been an invaluable part of my academic journey. His contributions have been greatly appreciated, complementing the collective guidance I received throughout my PhD. Dr.

Loman supported my future aspirations by providing crucial references for my postdoctoral applications.

Christian, my senior and frequent co-author, extended his generosity beyond what could be expected, assisting me with all aspects of bioinformatics analysis. His willingness to help, regardless of how trivial my questions seemed, has been a testament to the collaborative environment fostered within our team. *Sergii*, your expertise as a skilled bioinformatician has been immensely beneficial. I've learned a lot about bioinformatics, Snakemake, and GitHub from you. *Pei*, a colleague and friend has been instrumental in teaching me dPCR and wet lab techniques. I will not forget our spicy food sessions. Thanks to - *Malin*, for our great teamwork on ERβ and CITED1 projects and for providing guidance throughout. *Alan* (*Yilun*) and *Robert* for your readiness to assist with my questions and providing me with a good headstart in the lab. *Anna-Sophia*, for great teamwork on HER2-low project.

A big thanks to all my friends at MV - *Mats*, *Carina*, *Kamila*, *Pontus*, *Jacob*, as well as my Indian group including *Om, Kumkum*, *Megha*, *Varsha,* and *Kreema* for memorable times and for fun potlucks.

Additionally, I would like to thank my co-authors - *Jill Howlin, Sofia Gruvberger-Saal, Anthony (Tony), Miguel, Anna Ehinger*, and my collaborators on the DYPD project - *Kristín* and *Ana Bosch.*

I am deeply grateful for the financial support received from the CanFaster program. My heartfelt gratitude to *Sara Ek* and *Jana,* your dedication and hard work in orchestrating such a supportive environment for researchers is immensely appreciated. To my CanFaster colleagues, thank you for being part of this enriching journey. Also, to all the funding bodies that supported my PhD research.

Susanne André and *Anne Vähäniemi* for support in all things administrative.

Stefan Hansson for being a great support and approving my extension application. I also express my sincere gratitude to my M.Phil. supervisor *Dr. Simon Tomlinson*, for his support and guidance before I came to Lund.

I also want to thank *Kevin Blighe*, *Michael Love* and all contributors of Biostars and Stackoverflow - for providing bioinformatics support.

Pedro, my heartfelt thanks to you for being there to guide and support me whenever I needed it. Your unwavering support and profound insights have been my guiding light throughout this journey.

To my son, *Yatharth*, though you are young now and may read this when you are older, you have been my greatest inspiration throughout this journey. The sacrifices made, especially the times spent away from you, have been some of

the most challenging moments of my life. I hope one day you will understand the depth of my love and commitment to our future, which fueled my efforts. You are my heart and my motivation.

My parents for their love and support.

I extend my deepest gratitude to the thousands of patients who participated in the SCAN-B and ABiM studies, generously donating their biological materials. Their invaluable contributions have not only laid the foundation for this thesis but also continue to support numerous ongoing and future projects through the SCAN-B biobank, potentially aiding patients around the globe.

I owe a debt of gratitude to those who have listened, advised, and encouraged me throughout this process. Without your invaluable support, this would not have been possible. You have each in your own way, contributed to the completion of this work. For all those who walked this path with me but remain unnamed here, know that your contributions are deeply appreciated and cherished.

॥ अंत अस्ति प्रारंभ ॥

End is the new beginning

Heena Dalal
Lund, May 2024

References

1. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell. 2000;100: 57–70. doi:10.1016/S0092-8674(00)81683-9
2. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011;144: 646–674. doi:10.1016/j.cell.2011.02.013
3. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discovery. 2022;12: 31–46. doi:10.1158/2159-8290.CD-21-1059
4. Hippocrates. The Genuine Works of Hippocrates. Sydenham Society; 1849.
5. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018;68: 394–424. doi:10.3322/caac.21492
6. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71: 209–249. doi:10.3322/caac.21660
7. Socialstyrelsen. Socialstyrelsen Cancer Statistics Database. Statistics on Cancer Incidence 2020. 2020. Visited: 2024-04-12
8. Thun M, Linet MS, Cerhan JR, Haiman CA, Schottenfeld D, Thun M, et al., editors. Cancer Epidemiology and Prevention. Fourth Edition, Fourth Edition. Oxford, New York: Oxford University Press; 2018.
9. Vineis P, Wild CP. Global cancer patterns: causes and prevention. Lancet. 2014;383: 549–557. doi:10.1016/S0140-6736(13)62224-2
10. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291: 1304–1351. doi:10.1126/science.1058040
11. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458: 719–724. doi:10.1038/nature07943
12. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. Science. 2013;339: 1546–1558. doi:10.1126/science.1235122
13. Ledford H. Cancer treatment: The killer within. Nature. 2014;508: 24–26. doi:10.1038/508024a
14. Kumar. Robbins Basic Pathology - 10th Edition. Available: https://shop.elsevier.com/books/robbins-basic-pathology/kumar/978-0-323-35317-5
15. Five decades of advances in cancer research. Cell. 2024;187: 1567–1568. doi:10.1016/j.cell.2024.02.044
16. Knudson AG. Mutation and Cancer: Statistical Study of Retinoblastoma. Proc Natl Acad Sci U S A. 1971;68: 820–823.
17. Oren M, Prives C. p53: A tale of complexity and context. Cell. 2024;187: 1569–1573. doi:10.1016/j.cell.2024.02.043
18. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, et al. PTEN, a Putative Protein Tyrosine Phosphatase Gene Mutated in Human Brain, Breast, and Prostate Cancer. Science. 1997;275: 1943–1947. doi:10.1126/science.275.5308.1943
19. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. The Molecular Basis of Cancer-Cell Behavior. Molecular Biology of the Cell 4th edition. Garland Science; 2002. Available: https://www.ncbi.nlm.nih.gov/books/NBK26902/
20. Datta N, Chakraborty S, Basu M, Ghosh MK. Tumor Suppressors Having Oncogenic Functions: The Double Agents. Cells. 2020;10: 46. doi:10.3390/cells10010046
21. Alexandrova EM, Mirza SA, Xu S, Schulz-Heddergott R, Marchenko ND, Moll UM. p53 loss-of-heterozygosity is a necessary prerequisite for mutant p53 stabilization and gain-of-function in vivo. Cell Death Dis. 2017;8: e2661. doi:10.1038/cddis.2017.80
22. Pitolli C, Wang Y, Mancini M, Shi Y, Melino G, Amelio I. Do Mutations Turn p53 into an Oncogene? Int J Mol Sci. 2019;20: 6241. doi:10.3390/ijms20246241

23. Grabher C, von Boehmer H, Look AT. Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. Nat Rev Cancer. 2006;6: 347–359. doi:10.1038/nrc1880
24. Dotto GP. Crosstalk of Notch with p53 and p63 in cancer growth control. Nat Rev Cancer. 2009;9: 587–595. doi:10.1038/nrc2675
25. Swanton C, Bernard E, Abbosh C, André F, Auwerx J, Balmain A, et al. Embracing cancer complexity: Hallmarks of systemic disease. Cell. 2024;187: 1589–1616. doi:10.1016/j.cell.2024.02.009
26. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med. 2005;353: 1673–1684. doi:10.1056/NEJMoa052122
27. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, et al. Adjuvant trastuzumab in HER2-positive breast cancer. N Engl J Med. 2011;365: 1273–1283. doi:10.1056/NEJMoa0910383
28. Bradley R, Braybrooke J, Gray R, Hills R, Liu Z, Peto R, et al. Trastuzumab for early-stage, HER2-positive breast cancer: a meta-analysis of 13 864 women in seven randomised trials. The Lancet Oncology. 2021;22: 1139–1150. doi:10.1016/S1470-2045(21)00288-6
29. Antoniou A, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet. 2003;72: 1117–1130. doi:10.1086/375033
30. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol. 2007;25: 1329–1333. doi:10.1200/JCO.2006.09.1066
31. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA. 2017;317: 2402–2416. doi:10.1001/jama.2017.7112
32. Li FP, Fraumeni JF, Mulvihill JJ, Blattner WA, Dreyfus MG, Tucker MA, et al. A cancer family syndrome in twenty-four kindreds. Cancer Res. 1988;48: 5358–5362.
33. Guha T, Malkin D. Inherited TP53 Mutations and the Li–Fraumeni Syndrome. Cold Spring Harb Perspect Med. 2017;7: a026187. doi:10.1101/cshperspect.a026187
34. Blondeaux E, Arecco L, Punie K, Graffeo R, Toss A, De Angelis C, et al. Germline TP53 pathogenic variants and breast cancer: A narrative review. Cancer Treat Rev. 2023;114: 102522. doi:10.1016/j.ctrv.2023.102522
35. Olafson LR, Gunawardena M, Nixdorf S, McDonald KL, Rapkins RW. The role of TP53 gain-of-function mutation in multifocal glioblastoma. J Neurooncol. 2020;147: 37–47. doi:10.1007/s11060-019-03318-5
36. Ognjanovic S, Olivier M, Bergemann TL, Hainaut P. Sarcomas in TP53 germline mutation carriers: a review of the IARC TP53 database. Cancer. 2012;118: 1387–1396. doi:10.1002/cncr.26390
37. Raymond VM, Else T, Everett JN, Long JM, Gruber SB, Hammer GD. Prevalence of germline TP53 mutations in a prospective series of unselected patients with adrenocortical carcinoma. J Clin Endocrinol Metab. 2013;98: E119-125. doi:10.1210/jc.2012-2198
38. Cybulski C, Wokołorczyk D, Kładny J, Kurzwaski G, Suchy J, Grabowska E, et al. Germline CHEK2 mutations and colorectal cancer risk: different effects of a missense and truncating mutations? Eur J Hum Genet. 2007;15: 237–241. doi:10.1038/sj.ejhg.5201734
39. Wang Y, Dai B, Ye D. CHEK2 mutation and risk of prostate cancer: a systematic review and meta-analysis. Int J Clin Exp Med. 2015;8: 15708–15715.
40. Rainville I, Hatcher S, Rosenthal E, Larson K, Bernhisel R, Meek S, et al. High risk of breast cancer in women with biallelic pathogenic variants in CHEK2. Breast Cancer Res Treat. 2020;180: 503–509. doi:10.1007/s10549-020-05543-3
41. Oropeza E, Seker S, Carrel S, Mazumder A, Lozano D, Jimenez A, et al. Molecular portraits of cell cycle checkpoint kinases in cancer evolution, progression, and treatment responsiveness. Science Advances. 2023;9: eadf2860. doi:10.1126/sciadv.adf2860
42. Nelen MR, Kremer H, Konings IB, Schoute F, van Essen AJ, Koch R, et al. Novel PTEN mutations in patients with Cowden disease: absence of clear genotype-phenotype correlations. Eur J Hum Genet. 1999;7: 267–273. doi:10.1038/sj.ejhg.5200289
43. Xing Y, Lin NU, Maurer MA, Chen H, Mahvash A, Sahin A, et al. Phase II trial of AKT inhibitor MK-2206 in patients with advanced breast cancer who have tumors with PIK3CA or AKT mutations, and/or PTEN loss/PTEN mutation. Breast Cancer Res. 2019;21: 78. doi:10.1186/s13058-019-1154-8
44. Chen J, Sun J, Wang Q, Du Y, Cheng J, Yi J, et al. Systemic Deficiency of PTEN Accelerates Breast Cancer Growth and Metastasis. Front Oncol. 2022;12. doi:10.3389/fonc.2022.825484

45. Colombo C, Pogliaghi G, Tosi D, Muzza M, Bulfamante G, Persani L, et al. Thyroid cancer harboring PTEN and TP53 mutations: A peculiar molecular and clinical case report. Front Oncol. 2022;12. doi:10.3389/fonc.2022.949098
46. Guo C, Song W, Sun P, Jin L, Dai H. LncRNA-GAS5 induces PTEN expression through inhibiting miR-103 in endometrial cancer cells. J Biomed Sci. 2015;22: 100. doi:10.1186/s12929-015-0213-4
47. Pon JR, Marra MA. Driver and Passenger Mutations in Cancer. Annual Review of Pathology: Mechanisms of Disease. 2015;10: 25–50. doi:10.1146/annurev-pathol-012414-040312
48. Klein MI, Cannataro VL, Townsend JP, Newman S, Stern DF, Zhao H. Identifying Modules of Cooperating Cancer Drivers. bioRxiv; 2020. p. 2020.06.29.168229. doi:10.1101/2020.06.29.168229
49. O'Leary C, Gasper H, Sahin KB, Tang M, Kulasinghe A, Adams MN, et al. Epidermal Growth Factor Receptor (EGFR)-Mutated Non-Small-Cell Lung Cancer (NSCLC). Pharmaceuticals (Basel). 2020;13: 273. doi:10.3390/ph13100273
50. Tyc KM, Kazi A, Ranjan A, Wang R, Sebti SM. Novel mutant KRAS addiction signature predicts response to the combination of ERBB and MEK inhibitors in lung and pancreatic cancers. iScience. 2023;26: 106082. doi:10.1016/j.isci.2023.106082
51. Huang L, Guo Z, Wang F, Fu L. KRAS mutation: from undruggable to druggable in cancer. Signal Transduct Target Ther. 2021;6: 386. doi:10.1038/s41392-021-00780-4
52. Bahar ME, Kim HJ, Kim DR. Targeting the RAS/RAF/MAPK pathway for cancer therapy: from mechanism to clinical studies. Sig Transduct Target Ther. 2023;8: 1–38. doi:10.1038/s41392-023-01705-z
53. Mutgan AC, Besikcioglu HE, Wang S, Friess H, Ceyhan GO, Demir IE. Insulin/IGF-driven cancer cell-stroma crosstalk as a novel therapeutic target in pancreatic cancer. Molecular Cancer. 2018;17: 66. doi:10.1186/s12943-018-0806-0
54. Du C, da Silva A, Morales-Oyarvide V, Costa AD, Kozak MM, Dunne RF, et al. Insulin-like growth factor-1 receptor expression and disease recurrence and survival in patients with resected pancreatic ductal adenocarcinoma. Cancer Epidemiol Biomarkers Prev. 2020;29: 1586–1595. doi:10.1158/1055-9965.EPI-19-1315
55. Vander Velde R, Yoon N, Marusyk V, Durmaz A, Dhawan A, Miroshnychenko D, et al. Resistance to targeted therapies as a multifactorial, gradual adaptation to inhibitor specific selective pressures. Nat Commun. 2020;11: 2393. doi:10.1038/s41467-020-16212-w
56. Hahm JY, Park J, Jang E-S, Chi SW. 8-Oxoguanine: from oxidative damage to epigenetic and epitranscriptional modification. Exp Mol Med. 2022;54: 1626–1642. doi:10.1038/s12276-022-00822-z
57. Kim Y-J, Wilson DM. Overview of Base Excision Repair Biochemistry. Curr Mol Pharmacol. 2012;5: 3–13.
58. Budden T, Bowden NA. The Role of Altered Nucleotide Excision Repair and UVB-Induced DNA Damage in Melanomagenesis. Int J Mol Sci. 2013;14: 1132–1151. doi:10.3390/ijms14011132
59. van de Kamp G, Heemskerk T, Kanaar R, Essers J. DNA Double Strand Break Repair Pathways in Response to Different Types of Ionizing Radiation. Front Genet. 2021;12. doi:10.3389/fgene.2021.738230
60. Tiwari V, Wilson DM. DNA Damage and Associated DNA Repair Defects in Disease and Premature Aging. The American Journal of Human Genetics. 2019;105: 237–257. doi:10.1016/j.ajhg.2019.06.005
61. Chatterjee N, Walker GC. Mechanisms of DNA damage, repair and mutagenesis. Environ Mol Mutagen. 2017;58: 235–263. doi:10.1002/em.22087
62. Rahimian E, Amini A, Alikarami F, Pezeshki SMS, Saki N, Safa M. DNA repair pathways as guardians of the genome: Therapeutic potential and possible prognostic role in hematologic neoplasms. DNA Repair. 2020;96: 102951. doi:10.1016/j.dnarep.2020.102951
63. Albertson TM, Ogawa M, Bugni JM, Hays LE, Chen Y, Wang Y, et al. DNA polymerase ε and δ proofreading suppress discrete mutator and cancer phenotypes in mice. Proc Natl Acad Sci U S A. 2009;106: 17101–17104. doi:10.1073/pnas.0907147106
64. Dexheimer TS. DNA Repair Pathways and Mechanisms. In: Mathews LA, Cabarcas SM, Hurt EM, editors. DNA Repair of Cancer Stem Cells. Dordrecht: Springer Netherlands; 2013. pp. 19–32. doi:10.1007/978-94-007-4590-2_2
65. Ramsoekh D, Wagner A, van Leerdam ME, Dooijes D, Tops CM, Steyerberg EW, et al. Cancer risk in MLH1, MSH2 and MSH6 mutation carriers; different risk profiles may influence clinical management. Hered Cancer Clin Pract. 2009;7: 17. doi:10.1186/1897-4287-7-17

66. Nolano A, Medugno A, Trombetti S, Liccardo R, De Rosa M, Izzo P, et al. Hereditary Colorectal Cancer: State of the Art in Lynch Syndrome. Cancers (Basel). 2022;15: 75. doi:10.3390/cancers15010075
67. Li K, Luo H, Huang L, Luo H, Zhu X. Microsatellite instability: a review of what the oncologist should know. Cancer Cell Int. 2020;20: 16. doi:10.1186/s12935-019-1091-8
68. Pećina-Šlaus N, Kafka A, Salamon I, Bukovac A. Mismatch Repair Pathway, Genome Stability and Cancer. Front Mol Biosci. 2020;7. doi:10.3389/fmolb.2020.00122
69. Krokan HE, Bjørås M. Base Excision Repair. Cold Spring Harb Perspect Biol. 2013;5: a012583. doi:10.1101/cshperspect.a012583
70. Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding nucleotide excision repair and its roles in cancer and ageing. Nat Rev Mol Cell Biol. 2014;15: 465–481. doi:10.1038/nrm3822
71. Duan M, Ulibarri J, Liu KJ, Mao P. Role of Nucleotide Excision Repair in Cisplatin Resistance. Int J Mol Sci. 2020;21: 9248. doi:10.3390/ijms21239248
72. Aiello MM, Vigneri PG, Bruzzi P, Verderame F, Paratore S, Restuccia N, et al. Excision repair cross complementation group 1 (ERCC-1) gene polymorphisms and response to nivolumab in advanced non-small cell lung cancer (NSCLC). JCO. 2017;35: 3032–3032. doi:10.1200/JCO.2017.35.15_suppl.3032
73. Vítor AC, Huertas P, Legube G, de Almeida SF. Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox. Frontiers in Molecular Biosciences. 2020;7. Available: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00024
74. Elbakry A, Löbrich M. Homologous Recombination Subpathways: A Tangle to Resolve. Front Genet. 2021;12: 723847. doi:10.3389/fgene.2021.723847
75. Ali RMM, McIntosh SA, Savage KI. Homologous recombination deficiency in breast cancer: Implications for risk, cancer development, and therapy. Genes Chromosomes Cancer. 2021;60: 358–372. doi:10.1002/gcc.22921
76. Yi C, He C. DNA Repair by Reversal of DNA Damage. Cold Spring Harb Perspect Biol. 2013;5: a012575. doi:10.1101/cshperspect.a012575
77. Gillyard T, Davis J. DNA double-strand breaks repair in cancer: A path to achieving precision medicine. Int Rev Cell Mol Biol. 2021;364: 111–137. doi:10.1016/bs.ircmb.2021.06.003
78. Li B, Yu L, Gao L. Cancer classification based on multiple dimensions: SNV patterns. Computers in Biology and Medicine. 2022;151: 106270. doi:10.1016/j.compbiomed.2022.106270
79. Chu D, Wei L. Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. BMC Cancer. 2019;19: 359. doi:10.1186/s12885-019-5572-x
80. Castro-Mondragon JA, Aure MR, Lingjærde OC, Langerød A, Martens JWM, Børresen-Dale A-L, et al. Cis-regulatory mutations associate with transcriptional and post-transcriptional deregulation of gene regulatory programs in cancers. Nucleic Acids Res. 2022;50: 12131–12148. doi:10.1093/nar/gkac1143
81. Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous mutations frequently act as driver mutations in human cancers. Cell. 2014;156: 1324–1335. doi:10.1016/j.cell.2014.01.051
82. Sharma Y, Miladi M, Dukare S, Boulay K, Caudron-Herger M, Groß M, et al. A pan-cancer analysis of synonymous mutations. Nat Commun. 2019;10: 2569. doi:10.1038/s41467-019-10489-2
83. Zaidi SH, Harrison TA, Phipps AI, Steinfelder R, Trinh QM, Qu C, et al. Landscape of somatic single nucleotide variants and indels in colorectal cancer and impact on survival. Nat Commun. 2020;11: 3644. doi:10.1038/s41467-020-17386-z
84. Chen J, Guo J. Structural and functional analysis of somatic coding and UTR indels in breast and lung cancer genomes. Sci Rep. 2021;11: 21178. doi:10.1038/s41598-021-00583-1
85. Bennett EP, Petersen BL, Johansen IE, Niu Y, Yang Z, Chamberlain CA, et al. INDEL detection, the 'Achilles heel' of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. Nucleic Acids Research. 2020;48: 11958–11981. doi:10.1093/nar/gkaa975
86. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Human Molecular Genetics. 2010;19: R131–R136. doi:10.1093/hmg/ddq400
87. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. Nature. 2011;470: 59–65. doi:10.1038/nature09708
88. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575: 210–216. doi:10.1038/s41586-019-1689-y

89. Bolkestein M, Wong JKL, Thewes V, Körber V, Hlevnjak M, Elgaafary S, et al. Chromothripsis in Human Breast Cancer. Cancer Res. 2020;80: 4918–4931. doi:10.1158/0008-5472.CAN-20-1920
90. Shen MM. Chromoplexy: a new category of complex rearrangements in the cancer genome. Cancer Cell. 2013;23: 567–569. doi:10.1016/j.ccr.2013.04.025
91. Fernandes A, Shanmuganathan N, Branford S. Genomic Mechanisms Influencing Outcome in Chronic Myeloid Leukemia. Cancers (Basel). 2022;14: 620. doi:10.3390/cancers14030620
92. Aaltonen LA, Abascal F, Abeshouse A, Aburatani H, Adams DJ, Agrawal N, et al. Pan-cancer analysis of whole genomes. Nature. 2020;578: 82–93. doi:10.1038/s41586-020-1969-6
93. Cosenza MR, Rodriguez-Martin B, Korbel JO. Structural Variation in Cancer: Role, Prevalence, and Mechanisms. Annual Review of Genomics and Human Genetics. 2022;23: 123–152. doi:10.1146/annurev-genom-120121-101149
94. Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of somatic structural variation in human cancer genomes. Nature. 2020;578: 112–121. doi:10.1038/s41586-019-1913-9
95. Pös O, Radvanszky J, Buglyó G, Pös Z, Rusnakova D, Nagy B, et al. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. Biomed J. 2021;44: 548–559. doi:10.1016/j.bj.2021.02.003
96. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. Pathogenetics. 2008;1: 4. doi:10.1186/1755-8417-1-4
97. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet. 2009;10: 551–564. doi:10.1038/nrg2593
98. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578: 94–101. doi:10.1038/s41586-020-1943-3
99. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Research. 2017;45: D777–D783. doi:10.1093/nar/gkw1121
100. Brady SW, Gout AM, Zhang J. Therapeutic and prognostic insights from the analysis of cancer mutational signatures. Trends Genet. 2022;38: 194–208. doi:10.1016/j.tig.2021.08.007
101. Ramón y Cajal S, Sesé M, Capdevila C, Aasen T, De Mattos-Arruda L, Diaz-Cano SJ, et al. Clinical implications of intratumor heterogeneity: challenges and opportunities. J Mol Med (Berl). 2020;98: 161–177. doi:10.1007/s00109-020-01874-2
102. Fumagalli C, Barberis M. Breast Cancer Heterogeneity. Diagnostics (Basel). 2021;11: 1555. doi:10.3390/diagnostics11091555
103. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. Cancer Cell. 2020;37: 471–484. doi:10.1016/j.ccell.2020.03.007
104. Lopez JI, Cortes JM. A divide-and-conquer strategy in tumor sampling enhances detection of intratumor heterogeneity in routine pathology: A modeling approach in clear cell renal cell carcinoma. F1000Res. 2016;5: 385. doi:10.12688/f1000research.8196.2
105. Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. Cell. 2021;184: 2239-2254.e39. doi:10.1016/j.cell.2021.03.009
106. Lüönd F, Tiede S, Christofori G. Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. Br J Cancer. 2021;125: 164–175. doi:10.1038/s41416-021-01328-7
107. Cieślik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet. 2018;19: 93–109. doi:10.1038/nrg.2017.96
108. Tsimberidou AM, Fountzilas E, Bleris L, Kurzrock R. Transcriptomics and solid tumors: The next frontier in precision cancer medicine. Seminars in Cancer Biology. 2022;84: 50–59. doi:10.1016/j.semcancer.2020.09.007
109. Hu W, Wu Y, Shi Q, Wu J, Kong D, Wu X, et al. Systematic characterization of cancer transcriptome at transcript resolution. Nat Commun. 2022;13: 6803. doi:10.1038/s41467-022-34568-z
110. Bautista-Sánchez D, Arriaga-Canon C, Pedroza-Torres A, De La Rosa-Velázquez IA, González-Barrios R, Contreras-Espinosa L, et al. The Promising Role of miR-21 as a Cancer Biomarker and Its Importance in RNA-Based Therapeutics. Mol Ther Nucleic Acids. 2020;20: 409–420. doi:10.1016/j.omtn.2020.03.003

111. Sen R, Ghosal S, Das S, Balti S, Chakrabarti J. Competing Endogenous RNA: The Key to Posttranscriptional Regulation. The Scientific World Journal. 2014;2014. doi:10.1155/2014/896206
112. Rodon J, Soria J-C, Berger R, Miller WH, Rubin E, Kugel A, et al. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. Nat Med. 2019;25: 751–758. doi:10.1038/s41591-019-0424-4
113. Weidenbusch B, Richter GHS, Kesper MS, Guggemoos M, Gall K, Prexler C, et al. Transcriptome based individualized therapy of refractory pediatric sarcomas: feasibility, tolerability and efficacy. Oncotarget. 2018;9: 20747–20760. doi:10.18632/oncotarget.25087
114. Worst BC, van Tilburg CM, Balasubramanian GP, Fiesel P, Witt R, Freitag A, et al. Next-generation personalised medicine for high-risk paediatric cancer patients – The INFORM pilot study. European Journal of Cancer. 2016;65: 91–101. doi:10.1016/j.ejca.2016.06.009
115. Oberg JA, Glade Bender JL, Sulis ML, Pendrick D, Sireci AN, Hsiao SJ, et al. Implementation of next generation sequencing into pediatric hematology-oncology practice: moving beyond actionable alterations. Genome Medicine. 2016;8: 133. doi:10.1186/s13073-016-0389-6
116. Oftedal O. The mammary gland and its origin during synapsid evolution. Journal of mammary gland biology and neoplasia. 2002;7. doi:10.1023/a:1022896515287
117. Macias H, Hinck L. Mammary Gland Development. Wiley Interdiscip Rev Dev Biol. 2012;1: 533–557. doi:10.1002/wdev.35
118. Ghosh K, Vachon CM, Pankratz VS, Vierkant RA, Anderson SS, Brandt KR, et al. Independent Association of Lobular Involution and Mammographic Breast Density With Breast Cancer Risk. JNCI: Journal of the National Cancer Institute. 2010;102: 1716–1723. doi:10.1093/jnci/djq414
119. Arendt LM, Kuperwasser C. Form and function: how estrogen and progesterone regulate the mammary epithelial hierarchy. J Mammary Gland Biol Neoplasia. 2015;20: 9–25. doi:10.1007/s10911-015-9337-0
120. Sung H, J F, Rl S, M L, I S, A J, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a cancer journal for clinicians. 2021;71. doi:10.3322/caac.21660
121. Joko-Fru W, E J-A, A K, O O, Cp D, E C, et al. The evolving epidemic of breast cancer in sub-Saharan Africa: Results from the African Cancer Registry Network. International journal of cancer. 2020;147. doi:10.1002/ijc.33014
122. Reports UNDPHD. Human Developmental Index UN. Human Development Report 2021-22. Human Development Reports. United Nations; 2022 Sep. Available: https://hdr.undp.org/content/human-development-report-2021-22
123. Khattab A, Kashyap S, Monga DK. Male Breast Cancer. StatPearls. Treasure Island (FL): StatPearls Publishing; 2023. Available: http://www.ncbi.nlm.nih.gov/books/NBK526036/
124. Socialstyrelsen. Socialstyrelsen Cancer Statistics Database . Statistics on Cancer Incidence 2021. 2021. Available at: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/statistik/2022-12-8309.pdf. Visited: 2024-05-12
125. Yedjou CG, Sims JN, Miele L, Noubissi F, Lowe L, Fonseca DD, et al. Health and Racial Disparity in Breast Cancer. Adv Exp Med Biol. 2019;1152: 31–49. doi:10.1007/978-3-030-20301-6_3
126. Arzanova E, Mayrovitz HN. The Epidemiology of Breast Cancer. Breast Cancer [Internet]. Exon Publications; 2022. doi:10.36255/exon-publications-breast-cancer-epidemiology
127. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71: 209–249. doi:10.3322/caac.21660
128. American Cancer Society. Cancer Facts and Figured. 2022. Available at: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2022/2022-cancer-facts-and-figures.pdf. Visited: 2024-05-15.
129. DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A. International Variation in Female Breast Cancer Incidence and Mortality Rates. Cancer Epidemiol Biomarkers Prev. 2015;24: 1495–1506. doi:10.1158/1055-9965.EPI-15-0535
130. Wilkinson L, Gathani T. Understanding breast cancer as a global health concern. Br J Radiol. 2022;95: 20211033. doi:10.1259/bjr.20211033
131. Pan Hongchao, Gray Richard, Braybrooke Jeremy, Davies Christina, Taylor Carolyn, McGale Paul, et al. 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. New England Journal of Medicine. 2017;377: 1836–1846. doi:10.1056/NEJMoa1701830

132. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al. Breast Cancer Statistics, 2022. CA: A Cancer Journal for Clinicians. 2022;72: 524–541. doi:10.3322/caac.21754
133. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. Cancers (Basel). 2021;13: 4287. doi:10.3390/cancers13174287
134. Barrios CH. Global challenges in breast cancer detection and treatment. The Breast. 2022;62: S3–S6. doi:10.1016/j.breast.2022.02.003
135. McGuire A, Brown JAL, Malone C, McLaughlin R, Kerin MJ. Effects of age on the detection and management of breast cancer. Cancers (Basel). 2015;7: 908–929. doi:10.3390/cancers7020815
136. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. Cancers (Basel). 2021;13: 4287. doi:10.3390/cancers13174287
137. Endogenous Hormones and Breast Cancer Collaborative Group, Key TJ, Appleby PN, Reeves GK, Travis RC, Alberg AJ, et al. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. Lancet Oncol. 2013;14: 1009–1019. doi:10.1016/S1470-2045(13)70301-2
138. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet. 2001;358: 1389–1399. doi:10.1016/S0140-6736(01)06524-2
139. Bethea TN, Ochs-Balcom HM, Bandera EV, Beeghly-Fadiel A, Camancho F, Chyn D, et al. First- and second-degree family history of ovarian and breast cancer in relation to risk of invasive ovarian cancer in African American and White women. Int J Cancer. 2021;148: 2964–2973. doi:10.1002/ijc.33493
140. Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. Ann Oncol. 2015;26: 1291–1299. doi:10.1093/annonc/mdv022
141. Mehrgou A, Akouchekian M. The importance of BRCA1 and BRCA2 genes mutations in breast cancer development. Med J Islam Repub Iran. 2016;30: 369.
142. Kotsopoulos J. BRCA Mutations and Breast Cancer Prevention. Cancers (Basel). 2018;10: 524. doi:10.3390/cancers10120524
143. Shahbandi A, Nguyen HD, Jackson JG. TP53 Mutations and Outcomes in Breast Cancer: Reading beyond the Headlines. Trends Cancer. 2020;6: 98–110. doi:10.1016/j.trecan.2020.01.007
144. Corso G, Veronesi P, Sacchini V, Galimberti V. Prognosis and outcome in CDH1-mutant lobular breast cancer. Eur J Cancer Prev. 2018;27: 237–238. doi:10.1097/CEJ.0000000000000405
145. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science. 1997;275: 1943–1947. doi:10.1126/science.275.5308.1943
146. Steck PA, Pershouse MA, Jasser SA, Yung WK, Lin H, Ligon AH, et al. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet. 1997;15: 356–362. doi:10.1038/ng0497-356
147. Chen J, Lindblom A. Germline mutation screening of the STK11/LKB1 gene in familial breast cancer with LOH on 19p. Clin Genet. 2000;57: 394–397. doi:10.1034/j.1399-0004.2000.570511.x
148. Mao X, Omeogu C, Karanth S, Joshi A, Meernik C, Wilson L, et al. Association of reproductive risk factors and breast cancer molecular subtypes: a systematic review and meta-analysis. BMC Cancer. 2023;23: 644. doi:10.1186/s12885-023-11049-0
149. Migliavacca Zucchetti B, Peccatori FA, Codacci-Pisanelli G. Pregnancy and Lactation: Risk or Protective Factors for Breast Cancer? Adv Exp Med Biol. 2020;1252: 195–197. doi:10.1007/978-3-030-41596-9_27
150. Kim EY, Chang Y, Ahn J, Yun J-S, Park YL, Park CH, et al. Mammographic breast density, its changes, and breast cancer risk in premenopausal and postmenopausal women. Cancer. 2020;126: 4687–4696. doi:10.1002/cncr.33138
151. Dyrstad SW, Yan Y, Fowler AM, Colditz GA. Breast cancer risk associated with benign breast disease: systematic review and meta-analysis. Breast Cancer Res Treat. 2015;149: 569–575. doi:10.1007/s10549-014-3254-6
152. Thune I, Brenn T, Lund E, Gaard M. Physical activity and the risk of breast cancer. N Engl J Med. 1997;336: 1269–1275. doi:10.1056/NEJM199705013361801
153. Hoffman-Goetz L. Influence of physical activity and exercise on innate immunity. Nutr Rev. 1998;56: S126-130. doi:10.1111/j.1753-4887.1998.tb01629.x

154. Pati S, Irfan W, Jameel A, Ahmed S, Shahid RK. Obesity and Cancer: A Current Overview of Epidemiology, Pathogenesis, Outcomes, and Management. Cancers (Basel). 2023;15: 485. doi:10.3390/cancers15020485
155. Protani M, Coory M, Martin JH. Effect of obesity on survival of women with breast cancer: systematic review and meta-analysis. Breast Cancer Res Treat. 2010;123: 627–635. doi:10.1007/s10549-010-0990-0
156. Sun L, Zhu Y, Qian Q, Tang L. Body mass index and prognosis of breast cancer: An analysis by menstruation status when breast cancer diagnosis. Medicine (Baltimore). 2018;97: e11220. doi:10.1097/MD.0000000000011220
157. Zeinomar N, Knight JA, Genkinger JM, Phillips K-A, Daly MB, Milne RL, et al. Alcohol consumption, cigarette smoking, and familial breast cancer risk: findings from the Prospective Family Study Cohort (ProF-SC). Breast Cancer Res. 2019;21: 128. doi:10.1186/s13058-019-1213-1
158. Coronado GD, Beasley J, Livaudais J. Alcohol consumption and the risk of breast cancer. Salud Publica Mex. 2011;53: 440–447.
159. Erol A, Ho AM-C, Winham SJ, Karpyak VM. Sex hormones in alcohol consumption: a systematic review of evidence. Addict Biol. 2019;24: 157–169. doi:10.1111/adb.12589
160. Terry PD, Rohan TE. Cigarette smoking and the risk of breast cancer in women: a review of the literature. Cancer Epidemiol Biomarkers Prev. 2002;11: 953–971.
161. Huss L, Butt ST, Borgquist S, Elebro K, Sandsveden M, Rosendahl A, et al. Vitamin D receptor expression in invasive breast tumors and breast cancer survival. Breast Cancer Res. 2019;21: 84. doi:10.1186/s13058-019-1169-1
162. Dandamudi A, Tommie J, Nommsen-Rivers L, Couch S. Dietary Patterns and Breast Cancer Risk: A Systematic Review. Anticancer Res. 2018;38: 3209–3222. doi:10.21873/anticanres.12586
163. Casey SC, Vaccari M, Al-Mulla F, Al-Temaimi R, Amedei A, Barcellos-Hoff MH, et al. The effect of environmental chemicals on the tumor microenvironment. Carcinogenesis. 2015;36 Suppl 1: S160-183. doi:10.1093/carcin/bgv035
164. Leso V, Ercolano ML, Cioffi DL, Iavicoli I. Occupational Chemical Exposure and Breast Cancer Risk According to Hormone Receptor Status: A Systematic Review. Cancers (Basel). 2019;11: 1882. doi:10.3390/cancers11121882
165. Narod SA. Hormone replacement therapy and the risk of breast cancer. Nat Rev Clin Oncol. 2011;8: 669–676. doi:10.1038/nrclinonc.2011.110
166. Vinogradova Y, Coupland C, Hippisley-Cox J. Use of hormone replacement therapy and risk of breast cancer: nested case-control studies using the QResearch and CPRD databases. BMJ. 2020;371: m3873. doi:10.1136/bmj.m3873
167. Wernli KJ, Hampton JM, Trentham-Dietz A, Newcomb PA. Antidepressant medication use and breast cancer risk. Pharmacoepidemiol Drug Saf. 2009;18: 284–290. doi:10.1002/pds.1719
168. Friedman GD, Oestreicher N, Chan J, Quesenberry CP, Udaltsova N, Habel LA. Antibiotics and risk of breast cancer: up to 9 years of follow-up of 2.1 million women. Cancer Epidemiol Biomarkers Prev. 2006;15: 2102–2106. doi:10.1158/1055-9965.EPI-06-0401
169. Hewitt SC, Korach KS. Estrogen Receptors: New Directions in the New Millennium. Endocr Rev. 2018;39: 664–675. doi:10.1210/er.2018-00087
170. Jensen EV, Jacobson HI. Fate of Steroid Estrogens in Target Tissues. Biological Activities of Steroids in Relation to Cancer. 1960; 161–178. doi:10.1016/b978-1-4832-2866-2.50015-5
171. Jensen EV, Suzuki T, Kawashima T, Stumpf WE, Jungblut PW, DeSombre ER. A two-step mechanism for the interaction of estradiol with rat uterus. Proc Natl Acad Sci U S A. 1968;59: 632–638. doi:10.1073/pnas.59.2.632
172. Toft D, Gorski J. A receptor molecule for estrogens: isolation from the rat uterus and preliminary characterization. Proc Natl Acad Sci U S A. 1966;55: 1574–1581. doi:10.1073/pnas.55.6.1574
173. Green S, Walter P, Kumar V, Krust A, Bornert JM, Argos P, et al. Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature. 1986;320: 134–139. doi:10.1038/320134a0
174. Kumar V, Green S, Staub A, Chambon P. Localisation of the oestradiol-binding and putative DNA-binding domains of the human oestrogen receptor. EMBO J. 1986;5: 2231–2236. doi:10.1002/j.1460-2075.1986.tb04489.x
175. Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA. Cloning of a novel receptor expressed in rat prostate and ovary. Proc Natl Acad Sci U S A. 1996;93: 5925–5930. doi:10.1073/pnas.93.12.5925

176. Ogawa S, Inoue S, Watanabe T, Hiroi H, Orimo A, Hosoi T, et al. The Complete Primary Structure of Human Estrogen Receptor β (hERβ) and Its Heterodimerization with ER αin Vivoandin Vitro. Biochemical and Biophysical Research Communications. 1998;243: 122–126. doi:10.1006/bbrc.1997.7893
177. Nilsson S, Mäkelä S, Treuter E, Tujague M, Thomsen J, Andersson G, et al. Mechanisms of Estrogen Action. Physiological Reviews. 2001;81: 1535–1565. doi:10.1152/physrev.2001.81.4.1535
178. Gruvberger S. Estrogen receptor alpha and beta in breast cancer - gene expression profiles and clinical implications. Doctoral Thesis (compilation), Department of Oncology, Clinical Sciences, Lund University. 2005.
179. Carmeci C, Thompson DA, Ring HZ, Francke U, Weigel RJ. Identification of a gene (GPR30) with homology to the G-protein-coupled receptor superfamily associated with estrogen receptor expression in breast cancer. Genomics. 1997;45: 607–617. doi:10.1006/geno.1997.4972
180. Filardo EJ, Quinn JA, Bland KI, Frackelton AR Jr. Estrogen-Induced Activation of Erk-1 and Erk-2 Requires the G Protein-Coupled Receptor Homolog, GPR30, and Occurs via Trans-Activation of the Epidermal Growth Factor Receptor through Release of HB-EGF. Molecular Endocrinology. 2000;14: 1649–1660. doi:10.1210/mend.14.10.0532
181. Pepermans RA, Sharma G, Prossnitz ER. G Protein-Coupled Estrogen Receptor in Cancer and Stromal Cells: Functions and Novel Therapeutic Perspectives. Cells. 2021;10: 672. doi:10.3390/cells10030672
182. Hall JM, McDonnell DP, Korach KS. Allosteric regulation of estrogen receptor structure, function, and coactivator recruitment by different estrogen response elements. Mol Endocrinol. 2002;16: 469–486. doi:10.1210/mend.16.3.0814
183. Enmark E, Pelto-Huikko M, Grandien K, Lagercrantz S, Lagercrantz J, Fried G, et al. Human estrogen receptor beta-gene structure, chromosomal localization, and expression pattern. J Clin Endocrinol Metab. 1997;82: 4258–4265. doi:10.1210/jcem.82.12.4470
184. Kong EH, Pike ACW, Hubbard RE. Structure and mechanism of the oestrogen receptor. Biochemical Society Transactions. 2003;31: 56–59. doi:10.1042/bst0310056
185. Bai Z, Gust R. Breast Cancer, Estrogen Receptor and Ligands. Archiv der Pharmazie. 2009;342: 133–149. doi:10.1002/ardp.200800174
186. Pratt WB, Toft DO. Steroid receptor interactions with heat shock protein and immunophilin chaperones. Endocr Rev. 1997;18: 306–360. doi:10.1210/edrv.18.3.0303
187. Joab I, Radanyi C, Renoir M, Buchou T, Catelli MG, Binart N, et al. Common non-hormone binding component in non-transformed chick oviduct receptors of four steroid hormones. Nature. 1984;308: 850–853. doi:10.1038/308850a0
188. Hall JM, Couse JF, Korach KS. The multifaceted mechanisms of estradiol and estrogen receptor signaling. J Biol Chem. 2001;276: 36869–36872. doi:10.1074/jbc.R100029200
189. Delaunay F, Pettersson K, Tujague M, Gustafsson J-Å. Functional Differences between the Amino-Terminal Domains of Estrogen Receptors α and β. Mol Pharmacol. 2000;58: 584–590. doi:10.1124/mol.58.3.584
190. Devin-Leclerc J, Meng X, Delahaye F, Leclerc P, Baulieu EE, Catelli MG. Interaction and dissociation by ligands of estrogen receptor and Hsp90: the antiestrogen RU 58668 induces a protein synthesis-dependent clustering of the receptor in the cytoplasm. Mol Endocrinol. 1998;12: 842–854. doi:10.1210/mend.12.6.0121
191. Barkhem T, Carlsson B, Nilsson Y, Enmark E, Gustafsson J, Nilsson S. Differential response of estrogen receptor alpha and estrogen receptor beta to partial estrogen agonists/antagonists. Mol Pharmacol. 1998;54: 105–112. doi:10.1124/mol.54.1.105
192. An J, Tzagarakis-Foster C, Scharschmidt TC, Lomri N, Leitman DC. Estrogen Receptor β-Selective Transcriptional Activity and Recruitment of Coregulators by Phytoestrogens*. Journal of Biological Chemistry. 2001;276: 17808–17814. doi:10.1074/jbc.M100953200
193. Chrzan BG, Bradford PG. Phytoestrogens activate estrogen receptor β1 and estrogenic responses in human breast and bone cancer cell lines. Molecular Nutrition & Food Research. 2007;51: 171–177. doi:10.1002/mnfr.200600091
194. Fuentes N, Silveyra P. Estrogen receptor signaling mechanisms. Adv Protein Chem Struct Biol. 2019;116: 135–170. doi:10.1016/bs.apcsb.2019.01.001
195. Yaşar P, Ayaz G, User SD, Güpür G, Muyan M. Molecular mechanism of estrogen–estrogen receptor signaling. Reproductive Medicine and Biology. 2017;16: 4. doi:10.1002/rmb2.12006
196. Katzenellenbogen BS, Katzenellenbogen JA. Estrogen receptor transcription and transactivation: Estrogen receptor alpha and estrogen receptor beta: regulation by selective estrogen receptor

modulators and importance in breast cancer. Breast Cancer Res. 2000;2: 335–344. doi:10.1186/bcr78
197. Zhao C, Gao H, Liu Y, Papoutsi Z, Jaffrey S, Gustafsson J-A, et al. Genome-wide mapping of estrogen receptor-beta-binding regions reveals extensive cross-talk with transcription factor activator protein-1. Cancer Res. 2010;70: 5174–5183. doi:10.1158/0008-5472.CAN-09-4407
198. Acconcia F, Totta P, Ogawa S, Cardillo I, Inoue S, Leone S, et al. Survival versus apoptotic 17β-estradiol effect: Role of ERα and ERβ activated non-genomic signaling. Journal of Cellular Physiology. 2005;203: 193–201. doi:10.1002/jcp.20219
199. Helguero LA, Faulds MH, Gustafsson J-Å, Haldosén L-A. Estrogen receptors alfa (ERα) and beta (ERβ) differentially regulate proliferation and apoptosis of the normal murine mammary epithelial cell line HC11. Oncogene. 2005;24: 6605–6616. doi:10.1038/sj.onc.1208807
200. Bardin A, Boulle N, Lazennec G, Vignon F, Pujol P. Loss of ERbeta expression as a common step in estrogen-dependent tumor progression. Endocr Relat Cancer. 2004;11: 537–551. doi:10.1677/erc.1.00800
201. Alferez DG, Simões BM, Howell SJ, Clarke RB. The Role of Steroid Hormones in Breast and Effects on Cancer Stem Cells. Curr Stem Cell Rep. 2018;4: 81–94. doi:10.1007/s40778-018-0114-z
202. Igarashi H, Kouro T, Yokota T, Comp PC, Kincade PW. Age and stage dependency of estrogen receptor expression by lymphocyte precursors. Proc Natl Acad Sci U S A. 2001;98: 15131–15136. doi:10.1073/pnas.011513098
203. Seto K, Hoang M, Santos T, Bandyopadhyay M, Kindy MS, Dasgupta S. Non-genomic oestrogen receptor signal in B lymphocytes: An approach towards therapeutic interventions for infection, autoimmunity and cancer. Int J Biochem Cell Biol. 2016;76: 115–118. doi:10.1016/j.biocel.2016.04.018
204. Hedges VL, Chen G, Yu L, Krentzel AA, Starrett JR, Zhu J-N, et al. Local Estrogen Synthesis Regulates Parallel Fiber-Purkinje Cell Neurotransmission Within the Cerebellar Cortex. Endocrinology. 2018;159: 1328–1338. doi:10.1210/en.2018-00039
205. Allred DC, Brown P, Medina D. The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer. Breast Cancer Res. 2004;6: 240–245. doi:10.1186/bcr938
206. Bertucci F, Finetti P, Cervera N, Charafe-Jauffret E, Buttarelli M, Jacquemier J, et al. How different are luminal A and basal breast cancers? Int J Cancer. 2009;124: 1338–1348. doi:10.1002/ijc.24055
207. Petrossian K, Kanaya N, Lo C, Hsu P-Y, Nguyen D, Yang L, et al. ERα-mediated cell cycle progression is an important requisite for CDK4/6 inhibitor response in HR+ breast cancer. Oncotarget. 2018;9: 27736. doi:10.18632/oncotarget.25552
208. Hua H, Zhang H, Kong Q, Jiang Y. Mechanisms for estrogen receptor expression in human cancer. Experimental Hematology & Oncology. 2018;7: 24. doi:10.1186/s40164-018-0116-7
209. Johnston SR, Saccani-Jotti G, Smith IE, Salter J, Newby J, Coppen M, et al. Changes in estrogen receptor, progesterone receptor, and pS2 expression in tamoxifen-resistant human breast cancer. Cancer Res. 1995;55: 3331–3338.
210. Yan M, Rayoo M, Takano EA, kConFab Investigators, Fox SB. Nuclear and cytoplasmic expressions of ERβ1 and ERβ2 are predictive of response to therapy and alters prognosis in familial breast cancers. Breast Cancer Res Treat. 2011;126: 395–405. doi:10.1007/s10549-010-0941-9
211. Elebro K, Borgquist S, Rosendahl AH, Markkula A, Simonsson M, Jirström K, et al. High estrogen receptor β expression Is prognostic among adjuvant chemotherapy-treated patients-results from a population-based breast cancer cohort. Clin Cancer Res. 2017;23: 766–777. doi:10.1158/1078-0432.CCR-16-1095
212. Chang J, Liu J, Li H, Li J, Mu Y, Feng B. Expression of ERβ gene in breast carcinoma and the relevance in neoadjuvant therapy. Oncol Lett. 2017;13: 1641–1646. doi:10.3892/ol.2017.5659
213. Hopp TA, Weiss HL, Parra IS, Cui Y, Osborne CK, Fuqua SAW. Low levels of estrogen receptor beta protein predict resistance to tamoxifen therapy in breast cancer. Clin Cancer Res. 2004;10: 7490–7499. doi:10.1158/1078-0432.CCR-04-1114
214. Gruvberger-Saal SK, Bendahl P-O, Saal LH, Laakso M, Hegardt C, Edén P, et al. Estrogen Receptor β Expression Is Associated with Tamoxifen Response in ERα-Negative Breast Carcinoma. Clinical Cancer Research. 2007;13: 1987–1994. doi:10.1158/1078-0432.CCR-06-1823
215. Hamilton N, Márquez-Garbán D, Mah V, Fernando G, Elshimali Y, Garbán H, et al. Biologic roles of estrogen receptor-β and insulin-like growth factor-2 in triple-negative breast cancer. Biomed Res Int. 2015;2015: 925703. doi:10.1155/2015/925703

216. Guo L, Zhang Y, Zhang W, Yilamu D. Correlation between estrogen receptor β expression and the curative effect of endocrine therapy in breast cancer patients. Exp Ther Med. 2014;7: 1568–1572. doi:10.3892/etm.2014.1634
217. Guo L, Zhang YU, Yilamu D, Liu S, Guo C. ERβ overexpression results in endocrine therapy resistance and poor prognosis in postmenopausal ERα-positive breast cancer patients. Oncol Lett. 2016;11: 1531–1536. doi:10.3892/ol.2016.4095
218. Austin D, Hamilton N, Elshimali Y, Pietras R, Wu Y, Vadgama J. Estrogen receptor-beta is a potential target for triple negative breast cancer treatment. Oncotarget. 2018;9: 33912–33930. doi:10.18632/oncotarget.26089
219. Stefanou D, Batistatou A, Briasoulis E, Arkoumani E, Agnantis NJ. Estrogen receptor beta (ERbeta) expression in breast carcinomas is not correlated with estrogen receptor alpha (ERalpha) and prognosis: the Greek experience. Eur J Gynaecol Oncol. 2004;25: 457–461.
220. Dalal H, Dahlgren M, Gladchuk S, Brueffer C, Gruvberger-Saal SK, Saal LH. Clinical associations of ESR2 (estrogen receptor beta) expression across thousands of primary breast tumors. Sci Rep. 2022;12: 4696. doi:10.1038/s41598-022-08210-3
221. Dey P, Wang A, Ziegler Y, Kumar S, Yan S, Kim SH, et al. Estrogen Receptor Beta 1: A Potential Therapeutic Target for Female Triple Negative Breast Cancer. Endocrinology. 2022;163: bqac172. doi:10.1210/endocr/bqac172
222. Zhang X, Leung Y-K, Ho S-M. AP-2 regulates the transcription of estrogen receptor (ER)-beta by acting through a methylation hotspot of the 0N promoter in prostate cancer cells. Oncogene. 2007;26: 7346–7354. doi:10.1038/sj.onc.1210537
223. Zhou Y, Zeng C, Li X, Wu P-L, Yin L, Yu X-L, et al. IGF-I stimulates ERβ and aromatase expression via IGF1R/PI3K/AKT-mediated transcriptional activation in endometriosis. J Mol Med (Berl). 2016;94: 887–897. doi:10.1007/s00109-016-1396-1
224. Cai W, Rambaud J, Teboul M, Masse I, Benoit G, Gustafsson J-A, et al. Expression levels of estrogen receptor beta are modulated by components of the molecular clock. Mol Cell Biol. 2008;28: 784–793. doi:10.1128/MCB.00233-07
225. Hirata S, Shoda T, Kato J, Hoshi K. The multiple untranslated first exons system of the human estrogen receptor beta (ER beta) gene. J Steroid Biochem Mol Biol. 2001;78: 33–40. doi:10.1016/s0960-0760(01)00071-1
226. Leung Y-K, Mak P, Hassan S, Ho S-M. Estrogen receptor (ER)-beta isoforms: a key to understanding ER-beta signaling. Proc Natl Acad Sci U S A. 2006;103: 13162–13167. doi:10.1073/pnas.0605676103
227. Moore JT, McKee DD, Slentz-Kesler K, Moore LB, Jones SA, Horne EL, et al. Cloning and characterization of human estrogen receptor beta isoforms. Biochem Biophys Res Commun. 1998;247: 75–78. doi:10.1006/bbrc.1998.8738
228. Hawse JR, Carter JM, Aspros KGM, Bruinsma ES, Koepplin JW, Negron V, et al. Optimized immunohistochemical detection of estrogen receptor beta using two validated monoclonal antibodies confirms its expression in normal and malignant breast tissues. Breast Cancer Res Treat. 2020;179: 241–249. doi:10.1007/s10549-019-05441-3
229. Marotti JD, Collins LC, Hu R, Tamimi RM. Estrogen receptor-beta expression in invasive breast cancer in relation to molecular phenotype: results from the Nurses' Health Study. Mod Pathol. 2010;23: 197–204. doi:10.1038/modpathol.2009.158
230. Elebro K, Borgquist S, Rosendahl AH, Markkula A, Simonsson M, Jirström K, et al. High Estrogen Receptor β Expression Is Prognostic among Adjuvant Chemotherapy-Treated Patients-Results from a Population-Based Breast Cancer Cohort. Clin Cancer Res. 2017;23: 766–777. doi:10.1158/1078-0432.CCR-16-1095
231. Hall JM, McDonnell DP. The estrogen receptor beta-isoform (ERbeta) of the human estrogen receptor modulates ERalpha transcriptional activity and is a key regulator of the cellular response to estrogens and antiestrogens. Endocrinology. 1999;140: 5566–5578. doi:10.1210/endo.140.12.7179
232. Reese JM, Suman VJ, Subramaniam M, Wu X, Negron V, Gingery A, et al. ERβ1: characterization, prognosis, and evaluation of treatment strategies in ERα-positive and -negative breast cancer. BMC Cancer. 2014;14: 749. doi:10.1186/1471-2407-14-749
233. Wang J, Zhang C, Chen K, Tang H, Tang J, Song C, et al. ERβ1 inversely correlates with PTEN/PI3K/AKT pathway and predicts a favorable prognosis in triple-negative breast cancer. Breast Cancer Res Treat. 2015;152: 255–269. doi:10.1007/s10549-015-3467-3

234. Honma N, Horii R, Iwase T, Saji S, Younes M, Takubo K, et al. Clinical importance of estrogen receptor-beta evaluation in breast cancer patients treated with adjuvant tamoxifen therapy. J Clin Oncol. 2008;26: 3727–3734. doi:10.1200/JCO.2007.14.2968
235. Wimberly H, Han G, Pinnaduwage D, Murphy LC, Yang XR, Andrulis IL, et al. ERβ splice variant expression in four large cohorts of human breast cancer patient tumors. Breast Cancer Res Treat. 2014;146: 657–667. doi:10.1007/s10549-014-3050-3
236. Božović A, Markićević M, Dimitrijević B, Jovanović Ćupić S, Krajnović M, Lukić S, et al. Potential clinical significance of ERβ ON promoter methylation in sporadic breast cancer. Med Oncol. 2013;30: 642. doi:10.1007/s12032-013-0642-4
237. Al-Nakhle H, Burns PA, Cummings M, Hanby AM, Hughes TA, Satheesha S, et al. Estrogen receptor {beta}1 expression is regulated by miR-92 in breast cancer. Cancer Res. 2010;70: 4778–4784. doi:10.1158/0008-5472.CAN-09-4104
238. Chen J-Q, Russo J. ERα-negative and triple negative breast cancer: Molecular features and potential therapeutic approaches. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2009;1796: 162–175. doi:10.1016/j.bbcan.2009.06.003
239. Palmieri C, Cheng GJ, Saji S, Zelada-Hedman M, W√§rri A, Weihua Z, et al. Estrogen receptor beta in breast cancer. Endocrine-Related Cancer. 2002;9: 1–13.
240. Fuqua SAW, Schiff R, Parra I, Moore JT, Mohsin SK, Osborne CK, et al. Estrogen Receptor β Protein in Human Breast Cancer: Correlation with Clinical Tumor Parameters1. Cancer Research. 2003;63: 2434–2439.
241. Mandušić V, Nikolić-Vukosavljević D, Tanić N, Kanjer K, Neškovic-Konstantinović Z, Čeleketić D, et al. Expression of estrogen receptor β wt isoform (ERβ1) and ERβΔ5 splice variant mRNAs in sporadic breast cancer. J Cancer Res Clin Oncol. 2007;133: 571–579. doi:10.1007/s00432-007-0209-x
242. Haldosén L-A, Zhao C, Dahlman-Wright K. Estrogen receptor beta in breast cancer. Molecular and Cellular Endocrinology. 2014;382: 665–672. doi:10.1016/j.mce.2013.08.005
243. Omoto Y, Iwase H. Clinical significance of estrogen receptor β in breast and prostate cancer from biological aspects. Cancer Science. 2015;106: 337–343. doi:10.1111/cas.12613
244. Hopp TA, Weiss HL, Parra IS, Cui Y, Osborne CK, Fuqua SAW. Low levels of estrogen receptor beta protein predict resistance to tamoxifen therapy in breast cancer. Clin Cancer Res. 2004;10: 7490–7499. doi:10.1158/1078-0432.CCR-04-1114
245. Lu W, Katzenellenbogen BS. Estrogen Receptor-β Modulation of the ERα-p53 Loop Regulating Gene Expression, Proliferation, and Apoptosis in Breast Cancer. HORM CANC. 2017;8: 230–242. doi:10.1007/s12672-017-0298-1
246. Dey P, Wang A, Ziegler Y, Kumar S, Yan S, Kim SH, et al. Estrogen Receptor Beta 1: A Potential Therapeutic Target for Female Triple Negative Breast Cancer. Endocrinology. 2022;163: bqac172. doi:10.1210/endocr/bqac172
247. Lim W, Park Y, Cho J, Park C, Park J, Park Y-K, et al. Estrogen receptor beta inhibits transcriptional activity of hypoxia inducible factor-1 through the downregulation of arylhydrocarbon receptor nuclear translocator. Breast Cancer Res. 2011;13: R32. doi:10.1186/bcr2854
248. Park C, Lee Y. Overexpression of ERβ is sufficient to inhibit hypoxia-inducible factor-1 transactivation. Biochem Biophys Res Commun. 2014;450: 261–266. doi:10.1016/j.bbrc.2014.05.107
249. Tan W, Li Q, Chen K, Su F, Song E, Gong C. Estrogen receptor beta as a prognostic factor in breast cancer patients: A systematic review and meta-analysis. Oncotarget. 2016;7: 10373–10385. doi:10.18632/oncotarget.7219
250. Shanle EK, Zhao Z, Hawse J, Wisinski K, Keles S, Yuan M, et al. Research resource: global identification of estrogen receptor β target genes in triple negative breast cancer cells. Mol Endocrinol. 2013;27: 1762–1775. doi:10.1210/me.2013-1164
251. Shanle EK, Onitilo AA, Huang W, Kim K, Zang C, Engel JM, et al. Prognostic significance of full-length estrogen receptor beta expression in stage I-III triple negative breast cancer. Am J Transl Res. 2015;7: 1246–1259.
252. Sellitto A, D'Agostino Y, Alexandrova E, Lamberti J, Pecoraro G, Memoli D, et al. Insights into the Role of Estrogen Receptor β in Triple-Negative Breast Cancer. Cancers. 2020;12: 1477. doi:10.3390/cancers12061477
253. Zhou Y, Liu X. The role of estrogen receptor beta in breast cancer. Biomark Res. 2020;8: 39. doi:10.1186/s40364-020-00223-2

254. Warner M, Wu W-F, Montanholi L, Nalvarte I, Antonson P, Gustafsson J-A. Ventral prostate and mammary gland phenotype in mice with complete deletion of the ERβ gene. Proc Natl Acad Sci U S A. 2020;117: 4902–4909. doi:10.1073/pnas.1920478117
255. Shioda T, Fenner MH, Isselbacher KJ. msg1, a novel melanocyte-specific gene, encodes a nuclear protein and is associated with pigmentation. Proc Natl Acad Sci U S A. 1996;93: 12298–12303.
256. Fenner MH, Parrish JE, Boyd Y, Reed V, MacDonald M, Nelson DL, et al. *MSG1*(Melanocyte-Specific Gene 1): Mapping to Chromosome Xq13.1, Genomic Organization, and Promoter Analysis. Genomics. 1998;51: 401–407. doi:10.1006/geno.1998.5383
257. Rodriguez TA, Sparrow DB, Scott AN, Withington SL, Preis JI, Michalicek J, et al. Cited1 Is Required in Trophoblasts for Placental Development and for Embryo Growth and Survival. Mol Cell Biol. 2004;24: 228–244. doi:10.1128/MCB.24.1.228-244.2004
258. Howlin J, McBryan J, Napoletano S, Lambe T, McArdle E, Shioda T, et al. CITED1 homozygous null mice display aberrant pubertal mammary ductal morphogenesis. Oncogene. 2006;25: 1532–1542. doi:10.1038/sj.onc.1209183
259. Gerstner JR, Landry CF. Expression of the transcriptional coactivator CITED1 in the adult and developing murine brain. Dev Neurosci. 2007;29: 203–212. doi:10.1159/000096389
260. Yahata T, de Caestecker MP, Lechleider RJ, Andriole S, Roberts AB, Isselbacher KJ, et al. The MSG1 non-DNA-binding transactivator binds to the p300/CBP coactivators, enhancing their functional link to the Smad transcription factors. J Biol Chem. 2000;275: 8825–8834. doi:10.1074/jbc.275.12.8825
261. Yahata T, Shao W, Endoh H, Hur J, Coser KR, Sun H, et al. Selective coactivation of estrogen-dependent transcription by CITED1 CBP/p300-binding protein. Genes Dev. 2001;15: 2598–2612. doi:10.1101/gad.906301
262. Shi G, Boyle SC, Sparrow DB, Dunwoodie SL, Shioda T, de Caestecker MP. The transcriptional activity of CITED1 is regulated by phosphorylation in a cell cycle-dependent manner. J Biol Chem. 2006;281: 27426–27435. doi:10.1074/jbc.M602631200
263. McBryan J, Howlin J, Kenny PA, Shioda T, Martin F. ERalpha-CITED1 co-regulated genes expressed during pubertal mammary gland development: implications for breast cancer prognosis. Oncogene. 2007;26: 6406–6419. doi:10.1038/sj.onc.1210468
264. Dahlgren M, Lettiero B, Dalal H, Mårtensson K, Gaber A, Nodin B, et al. CITED1 as a marker of favourable outcome in anti-endocrine treated, estrogen-receptor positive, lymph-node negative breast cancer. BMC Res Notes. 2023;16: 105. doi:10.1186/s13104-023-06376-1
265. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human Breast Cancer: Correlation of Relapse and Survival with Amplification of the HER-2/neu Oncogene. Science. 1987;235: 177–182. doi:10.1126/science.3798106
266. 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. 1989;244: 707–712. doi:10.1126/science.2470152
267. King CR, Kraus MH, Aaronson SA. Amplification of a Novel v-erbB-Related Gene in a Human Mammary Carcinoma. Science. 1985;229: 974–976. doi:10.1126/science.2992089
268. Schechter AL, Stern DF, Vaidyanathan L, Decker SJ, Drebin JA, Greene MI, et al. The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. Nature. 1984;312: 513–516. doi:10.1038/312513a0
269. Akiyama T, Sudo C, Ogawara H, Toyoshima K, Yamamoto T. The product of the human c-erbB-2 gene: a 185-kilodalton glycoprotein with tyrosine kinase activity. Science. 1986;232: 1644–1646. doi:10.1126/science.3012781
270. Hua G, Bergon A, Cauchy P, Kahn-Perlès B, Bertucci F, Birnbaum D, et al. ERBB2b mRNA isoform encodes a nuclear variant of the ERBB2 oncogene in breast cancer. Journal of Cellular Biochemistry. 2020;121: 4870–4886. doi:10.1002/jcb.29762
271. Cho H-S, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, et al. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature. 2003;421: 756–760. doi:10.1038/nature01392
272. Graus-Porta D, Beerli RR, Daly JM, Hynes NE. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. EMBO J. 1997;16: 1647–1655. doi:10.1093/emboj/16.7.1647
273. Loibl S, Gianni L. HER2-positive breast cancer. The Lancet. 2017;389: 2415–2429. doi:10.1016/S0140-6736(16)32417-5

274. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol. 2001;2: 127–137. doi:10.1038/35052073
275. Ross JS, Fletcher JA. The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy. Stem Cells. 1998;16: 413–428. doi:10.1002/stem.160413
276. Yu Z, Ye S, Hu G, Lv M, Tu Z, Zhou K, et al. The RAF-MEK-ERK pathway: targeting ERK to overcome obstacles to effective cancer therapy. Future Med Chem. 2015;7: 269–289. doi:10.4155/fmc.14.143
277. Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene. 2007;26: 6469–6487. doi:10.1038/sj.onc.1210477
278. Wieduwilt MJ, Moasser MM. The epidermal growth factor receptor family: biology driving targeted therapeutics. Cell Mol Life Sci. 2008;65: 1566–1584. doi:10.1007/s00018-008-7440-8
279. Prenzel N, Fischer OM, Streit S, Hart S, Ullrich A. The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. Endocr Relat Cancer. 2001;8: 11–31. doi:10.1677/erc.0.0080011
280. Yarden Y, Pines G. The ERBB network: at last, cancer therapy meets systems biology. Nat Rev Cancer. 2012;12: 553–563. doi:10.1038/nrc3309
281. Kallioniemi OP, Kallioniemi A, Kurisu W, Thor A, Chen LC, Smith HS, et al. ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. Proc Natl Acad Sci U S A. 1992;89: 5321–5325. doi:10.1073/pnas.89.12.5321
282. Allred DC, Clark GM, Molina R, Tandon AK, Schnitt SJ, Gilchrist KW, et al. Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer. Hum Pathol. 1992;23: 974–979. doi:10.1016/0046-8177(92)90257-4
283. Park K, Han S, Kim HJ, Kim J, Shin E. HER2 status in pure ductal carcinoma in situ and in the intraductal and invasive components of invasive ductal carcinoma determined by fluorescence in situ hybridization and immunohistochemistry. Histopathology. 2006;48: 702–707. doi:10.1111/j.1365-2559.2006.02403.x
284. Mittendorf EA, Wu Y, Scaltriti M, Meric-Bernstam F, Hunt KK, Dawood S, et al. Loss of HER2 amplification following trastuzumab-based neoadjuvant systemic therapy and survival outcomes. Clin Cancer Res. 2009;15: 7381–7388. doi:10.1158/1078-0432.CCR-09-1735
285. Gutierrez MC, Detre S, Johnston S, Mohsin SK, Shou J, Allred DC, et al. Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase. J Clin Oncol. 2005;23: 2469–2476. doi:10.1200/JCO.2005.01.172
286. Guo S, Sonenshein GE. Forkhead box transcription factor FOXO3a regulates estrogen receptor alpha expression and is repressed by the Her-2/neu/phosphatidylinositol 3-kinase/Akt signaling pathway. Mol Cell Biol. 2004;24: 8681–8690. doi:10.1128/MCB.24.19.8681-8690.2004
287. Pegram M, Jackisch C, Johnston SRD. Estrogen/HER2 receptor crosstalk in breast cancer: combination therapies to improve outcomes for patients with hormone receptor-positive/HER2-positive breast cancer. npj Breast Cancer. 2023;9: 1–19. doi:10.1038/s41523-023-00533-2
288. Villman K, Sjöström J, Heikkilä R, Hultborn R, Malmström P, Bengtsson N-O, et al. TOP2A and HER2 gene amplification as predictors of response to anthracycline treatment in breast cancer. Acta Oncol. 2006;45: 590–596. doi:10.1080/02841860500543182
289. Gutierrez C, Schiff R. HER 2: Biology, Detection, and Clinical Implications. Arch Pathol Lab Med. 2011;135: 55–62. doi:10.1043/2010-0454-RAR.1
290. 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. JCO. 2018;36: 2105–2122. doi:10.1200/JCO.2018.77.8738
291. Farshid G, Bilous M, Morey A, Fox S, Lakhani S, Loi S, et al. ASCO/CAP 2018 breast cancer HER2 testing guidelines: summary of pertinent recommendations for practice in Australia. Pathology. 2019;51: 345–348. doi:10.1016/j.pathol.2019.02.004
292. Farshid G, Dhatrak D, Gilhotra A, Koszyca B, Nolan J. The impact of 2018 ASCO-CAP HER2 testing guidelines on breast cancer HER2 results. An audit of 2132 consecutive cases evaluated by immunohistochemistry and in situ hybridization. Mod Pathol. 2020;33: 1783–1790. doi:10.1038/s41379-020-0555-7
293. Zhang H, Moisini I, Ajabnoor RM, Turner BM, Hicks DG. Applying the New Guidelines of HER2 Testing in Breast Cancer. Curr Oncol Rep. 2020;22: 51. doi:10.1007/s11912-020-0901-4
294. Vranic S, Teruya B, Repertinger S, Ulmer P, Hagenkord J, Gatalica Z. Assessment of HER2 gene status in breast carcinomas with polysomy of chromosome 17. Cancer. 2011;117: 48–53. doi:10.1002/cncr.25580

295. Tse CH, Hwang HC, Goldstein LC, Kandalaft PL, Wiley JC, Kussick SJ, et al. Determining true HER2 gene status in breast cancers with polysomy by using alternative chromosome 17 reference genes: implications for anti-HER2 targeted therapy. J Clin Oncol. 2011;29: 4168–4174. doi:10.1200/JCO.2011.36.0107
296. Cresti N, Lee J, Rourke E, Televantou D, Jamieson D, Verrill M, et al. Genetic variants in the HER2 gene: Influence on HER2 overexpression and loss of heterozygosity in breast cancer. European Journal of Cancer. 2016;55: 27–37. doi:10.1016/j.ejca.2015.10.066
297. Stanton SE, Ward MM, Christos P, Sanford R, Lam C, Cobham MV, et al. Pro1170 Ala polymorphism in HER2-neu is associated with risk of trastuzumab cardiotoxicity. BMC Cancer. 2015;15: 267. doi:10.1186/s12885-015-1298-6
298. Gao Y, Tang X, Cao J, Rong R, Yu Z, Liu Y, et al. The Effect of HER2 Single Nucleotide Polymorphisms on Cervical Cancer Susceptibility and Survival in a Chinese Population. J Cancer. 2019;10: 378–387. doi:10.7150/jca.27976
299. Novillo A, Gaibar M, Romero-Lorca A, Malón D, Antón B, Moreno A, et al. HER2 and BARD1 Polymorphisms in Early HER2-Positive Breast Cancer Patients: Relationship with Response to Neoadjuvant Anti-HER2 Treatment. Cancers (Basel). 2023;15: 763. doi:10.3390/cancers15030763
300. Thanh TN, Tran BSN, Thi APH, Binh TT, Nguyen TB, Minh TL, et al. HER2Ile655Val Single Nucleotide Polymorphism Associated with Early-Onset Breast Cancer Susceptibility: A Systematic Review and Meta-Analysis. Asian Pac J Cancer Prev. 2021;22: 11–18. doi:10.31557/APJCP.2021.22.1.11
301. Meng P, Dalal H, Chen Y, Brueffer C, Gladchuk S, Alcaide M, et al. Digital PCR quantification of ultrahigh ERBB2 copy number identifies poor breast cancer survival after trastuzumab. npj Breast Cancer. 2024;10: 1–10. doi:10.1038/s41523-024-00621-x
302. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med. 2001;344: 783–792. doi:10.1056/NEJM200103153441101
303. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med. 2005;353: 1659–1672. doi:10.1056/NEJMoa052306
304. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med. 2005;353: 1673–1684. doi:10.1056/NEJMoa052122
305. Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med. 2012;367: 1783–1791. doi:10.1056/NEJMoa1209124
306. Swain SM, Baselga J, Kim S-B, Ro J, Semiglazov V, Campone M, et al. Pertuzumab, Trastuzumab, and Docetaxel in HER2-Positive Metastatic Breast Cancer. N Engl J Med. 2015;372: 724–734. doi:10.1056/NEJMoa1413513
307. Swain SM, Shastry M, Hamilton E. Targeting HER2-positive breast cancer: advances and future directions. Nat Rev Drug Discov. 2023;22: 101–126. doi:10.1038/s41573-022-00579-0
308. Gornowicz A, Szymanowski W, Czarnomysy R, Bielawski K, Bielawska A. Anti-HER2 monoclonal antibodies intensify the susceptibility of human gastric cancer cells to etoposide by promoting apoptosis, but not autophagy. PLoS One. 2021;16: e0255585. doi:10.1371/journal.pone.0255585
309. Geyer CE, Forster J, Lindquist D, Chan S, Romieu CG, Pienkowski T, et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. N Engl J Med. 2006;355: 2733–2743. doi:10.1056/NEJMoa064320
310. Press MF, Finn RS, Cameron D, Di Leo A, Geyer CE, Villalobos IE, et al. HER-2 gene amplification, HER-2 and epidermal growth factor receptor mRNA and protein expression, and lapatinib efficacy in women with metastatic breast cancer. Clin Cancer Res. 2008;14: 7861–7870. doi:10.1158/1078-0432.CCR-08-1056
311. Piccart-Gebhart M, Holmes E, Baselga J, de Azambuja E, Dueck AC, Viale G, et al. Adjuvant Lapatinib and Trastuzumab for Early Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer: Results From the Randomized Phase III Adjuvant Lapatinib and/or Trastuzumab Treatment Optimization Trial. J Clin Oncol. 2016;34: 1034–1042. doi:10.1200/JCO.2015.62.1797
312. Moolten FL, Cooperband SR. Selective destruction of target cells by diphtheria toxin conjugated to antibody directed against antigens on the cells. Science. 1970;169: 68–70. doi:10.1126/science.169.3940.68
313. Rowland GF, O'Neill GJ, Davies DA. Suppression of tumour growth in mice by a drug-antibody conjugate using a novel approach to linkage. Nature. 1975;255: 487–488. doi:10.1038/255487a0

314. Köhler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature. 1975;256: 495–497. doi:10.1038/256495a0
315. Junttila TT, Li G, Parsons K, Phillips GL, Sliwkowski MX. Trastuzumab-DM1 (T-DM1) retains all the mechanisms of action of trastuzumab and efficiently inhibits growth of lapatinib insensitive breast cancer. Breast Cancer Res Treat. 2011;128: 347–356. doi:10.1007/s10549-010-1090-x
316. Ogitani Y, Aida T, Hagihara K, Yamaguchi J, Ishii C, Harada N, et al. DS-8201a, A Novel HER2-Targeting ADC with a Novel DNA Topoisomerase I Inhibitor, Demonstrates a Promising Antitumor Efficacy with Differentiation from T-DM1. Clinical Cancer Research. 2016;22: 5097–5108. doi:10.1158/1078-0432.CCR-15-2822
317. Tamura K, Tsurutani J, Takahashi S, Iwata H, Krop IE, Redfern C, et al. Trastuzumab deruxtecan (DS-8201a) in patients with advanced HER2-positive breast cancer previously treated with trastuzumab emtansine: a dose-expansion, phase 1 study. Lancet Oncol. 2019;20: 816–826. doi:10.1016/S1470-2045(19)30097-X
318. Modi S, Saura C, Yamashita T, Park YH, Kim S-B, Tamura K, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Positive Breast Cancer. N Engl J Med. 2020;382: 610–621. doi:10.1056/NEJMoa1914510
319. Enhertu approved in the US as first tumour-agnostic HER2-directed therapy for previously treated patients with metastatic HER2-positive solid tumours. 6 Apr 2024. Available: https://www.astrazeneca.com/media-centre/press-releases/2024/enhertu-approved-in-the-us-as-first-tumour-agnostic-her2-directed-therapy-for-previously-treated-patients-with-metastatic-her2-positive-solid-tumours.html
320. Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, et al. HER2-Low Breast Cancer: Pathological and Clinical Landscape. JCO. 2020;38: 1951–1962. doi:10.1200/JCO.19.02488
321. Molinelli C, Jacobs F, Agostinetto E, Nader-Marta G, Ceppi M, Bruzzone M, et al. Prognostic value of HER2-low status in breast cancer: a systematic review and meta-analysis. ESMO Open. 2023;8. doi:10.1016/j.esmoop.2023.101592
322. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. New England Journal of Medicine. 2022;387: 9–20. doi:10.1056/NEJMoa2203690
323. Buzatto IPC, Ribeiro-Silva A, Andrade JM, Carrara HHA, Silveira WA, Tiezzi DG. Neoadjuvant chemotherapy with trastuzumab in HER2-positive breast cancer: pathologic complete response rate, predictive and prognostic factors. Braz J Med Biol Res. 2017;50: e5674. doi:10.1590/1414-431X20165674
324. Ross JS, Slodkowska EA, Symmans WF, Pusztai L, Ravdin PM, Hortobagyi GN. The HER-2 receptor and breast cancer: ten years of targeted anti-HER-2 therapy and personalized medicine. Oncologist. 2009;14: 320–368. doi:10.1634/theoncologist.2008-0230
325. Hunter FW, Barker HR, Lipert B, Rothé F, Gebhart G, Piccart-Gebhart MJ, et al. Mechanisms of resistance to trastuzumab emtansine (T-DM1) in HER2-positive breast cancer. Br J Cancer. 2020;122: 603–612. doi:10.1038/s41416-019-0635-y
326. Trowe T, Boukouvala S, Calkins K, Cutler RE Jr, Fong R, Funke R, et al. EXEL-7647 Inhibits Mutant Forms of ErbB2 Associated with Lapatinib Resistance and Neoplastic Transformation. Clinical Cancer Research. 2008;14: 2465–2475. doi:10.1158/1078-0432.CCR-07-4367
327. Hou Y, Nitta H, Li Z. HER2 Intratumoral Heterogeneity in Breast Cancer, an Evolving Concept. Cancers (Basel). 2023;15: 2664. doi:10.3390/cancers15102664
328. Elster N, Cremona M, Morgan C, Toomey S, Carr A, O'Grady A, et al. A preclinical evaluation of the PI3K alpha/delta dominant inhibitor BAY 80-6946 in HER2-positive breast cancer models with acquired resistance to the HER2-targeted therapies trastuzumab and lapatinib. Breast Cancer Res Treat. 2015;149: 373–383. doi:10.1007/s10549-014-3239-5
329. Luque-Cabal M, García-Teijido P, Fernández-Pérez Y, Sánchez-Lorenzo L, Palacio-Vázquez I. Mechanisms Behind the Resistance to Trastuzumab in HER2-Amplified Breast Cancer and Strategies to Overcome It. Clin Med Insights Oncol. 2016;10: 21–30. doi:10.4137/CMO.S34537
330. Kataoka Y, Mukohara T, Shimada H, Saijo N, Hirai M, Minami H. Association between gain-of-function mutations in PIK3CA and resistance to HER2-targeted agents in HER2-amplified breast cancer cell lines. Ann Oncol. 2010;21: 255–262. doi:10.1093/annonc/mdp304
331. Ferraro E, Drago JZ, Modi S. Implementing antibody-drug conjugates (ADCs) in HER2-positive breast cancer: state of the art and future directions. Breast Cancer Research. 2021;23: 84. doi:10.1186/s13058-021-01459-y

332. Mosele MF, Lusque A, Dieras V, Deluche E, Ducoulombier A, Pistilli B, et al. LBA1 Unraveling the mechanism of action and resistance to trastuzumab deruxtecan (T-DXd): Biomarker analyses from patients from DAISY trial. Annals of Oncology. 2022;33: S123. doi:10.1016/j.annonc.2022.03.277
333. Graham JD, Yeates C, Balleine RL, Harvey SS, Milliken JS, Bilous AM, et al. Characterization of progesterone receptor A and B expression in human breast cancer. Cancer Res. 1995;55: 5063–5068.
334. Giatti S, Romano S, Pesaresi M, Cermenati G, Mitro N, Caruso D, et al. Neuroactive steroids and the peripheral nervous system: An update. Steroids. 2015;103: 23–30. doi:10.1016/j.steroids.2015.03.014
335. Africander D, Storbeck K-H. Steroid metabolism in breast cancer: Where are we and what are we missing? Mol Cell Endocrinol. 2018;466: 86–97. doi:10.1016/j.mce.2017.05.016
336. Horwitz KB, McGuire WL. Estrogen control of progesterone receptor in human breast cancer. Correlation with nuclear processing of estrogen receptor. J Biol Chem. 1978;253: 2223–2228.
337. Horwitz KB, Sartorius CA. 90 YEARS OF PROGESTERONE: Progesterone and progesterone receptors in breast cancer: past, present, future. Journal of Molecular Endocrinology. 2020;65: T49–T63. doi:10.1530/JME-20-0104
338. Li Z, Wei H, Li S, Wu P, Mao X. The Role of Progesterone Receptors in Breast Cancer. Drug Des Devel Ther. 2022;16: 305–314. doi:10.2147/DDDT.S336643
339. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406: 747–752. doi:10.1038/35021093
340. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences. 2001;98: 10869–10874. doi:10.1073/pnas.191367098
341. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27: 1160–1167. doi:10.1200/JCO.2008.18.1370
342. Inic Z, Zegarac M, Inic M, Markovic I, Kozomara Z, Djurisic I, et al. Difference between Luminal A and Luminal B Subtypes According to Ki-67, Tumor Size, and Progesterone Receptor Negativity Providing Prognostic Information. Clin Med Insights Oncol. 2014;8: 107–111. doi:10.4137/CMO.S18006
343. Krishnamurti U, Silverman JF. HER2 in breast cancer: a review and update. Adv Anat Pathol. 2014;21: 100–107. doi:10.1097/PAP.0000000000000015
344. Zagami P, Carey LA. Triple negative breast cancer: Pitfalls and progress. npj Breast Cancer. 2022;8: 1–10. doi:10.1038/s41523-022-00468-0
345. Kim HK, Park KH, Kim Y, Park SE, Lee HS, Lim SW, et al. Discordance of the PAM50 Intrinsic Subtypes Compared with Immunohistochemistry-Based Surrogate in Breast Cancer Patients: Potential Implication of Genomic Alterations of Discordance. Cancer Res Treat. 2019;51: 737–747. doi:10.4143/crt.2018.342
346. Picornell AC, Echavarria I, Alvarez E, López-Tarruella S, Jerez Y, Hoadley K, et al. Breast cancer PAM50 signature: correlation and concordance between RNA-Seq and digital multiplexed gene expression technologies in a triple negative breast cancer series. BMC Genomics. 2019;20: 452. doi:10.1186/s12864-019-5849-0
347. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415: 530–536. doi:10.1038/415530a
348. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351: 2817–2826. doi:10.1056/NEJMoa041588
349. Wallden B, Storhoff J, Nielsen T, Dowidar N, Schaper C, Ferree S, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. BMC Med Genomics. 2015;8: 54. doi:10.1186/s12920-015-0129-6
350. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011;17: 6012–6020. doi:10.1158/1078-0432.CCR-11-0926
351. Jerevall P-L, Ma X-J, Li H, Salunga R, Kesty NC, Erlander MG, et al. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. Br J Cancer. 2011;104: 1762–1769. doi:10.1038/bjc.2011.145

352. Brueffer C, Vallon-Christersson J, Grabau† D, Ehinger A, Häkkinen J, Hegardt C, et al. Clinical Value of RNA Sequencing–Based Classifiers for Prediction of the Five Conventional Breast Cancer Biomarkers: A Report From the Population-Based Multicenter Sweden Cancerome Analysis Network—Breast Initiative. JCO Precis Oncol. 2018; 1–18. doi:10.1200/PO.17.00135
353. Staaf J, Häkkinen J, Hegardt C, Saal LH, Kimbung S, Hedenfalk I, et al. RNA sequencing-based single sample predictors of molecular subtype and risk of recurrence for clinical assessment of early-stage breast cancer. npj Breast Cancer. 2022;8: 1–17. doi:10.1038/s41523-022-00465-3
354. Klarenbach S, Sims-Jones N, Lewin G, Singh H, Thériault G, Tonelli M, et al. Recommendations on screening for breast cancer in women aged 40–74 years who are not at increased risk for breast cancer. CMAJ. 2018;190: E1441–E1451. doi:10.1503/cmaj.180463
355. Ren W, Chen M, Qiao Y, Zhao F. Global guidelines for breast cancer screening: A systematic review. Breast. 2022;64: 85–99. doi:10.1016/j.breast.2022.04.003
356. Ryser MD, Lange J, Inoue LYT, O'Meara ES, Gard C, Miglioretti DL, et al. Estimation of Breast Cancer Overdiagnosis in a U.S. Breast Screening Cohort. Ann Intern Med. 2022;175: 471–478. doi:10.7326/M21-3577
357. Gordon PB. The Impact of Dense Breasts on the Stage of Breast Cancer at Diagnosis: A Review and Options for Supplemental Screening. Curr Oncol. 2022;29: 3595–3636. doi:10.3390/curroncol29050291
358. Swedish Cancer Society Report - Segregated screening.pdf. Available: https://static-files.cancerfonden.se/Swedish%20Cancer%20Society%20Report%20-%20Segregated%20screening.pdf
359. Leong AS-Y, Zhuang Z. The Changing Role of Pathology in Breast Cancer Diagnosis and Treatment. Pathobiology. 2011;78: 99–114. doi:10.1159/000292644
360. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. Clin Med Insights Pathol. 2015;8: 23–31. doi:10.4137/CPath.S31563
361. Masood S. Breast Cancer Subtypes: Morphologic and Biologic Characterization. Womens Health (Lond Engl). 2016;12: 103–119. doi:10.2217/whe.15.99
362. Henry NL, Cannon-Albright LA. Breast cancer histologic subtypes show excess familial clustering. Cancer. 2019;125: 3131–3138. doi:10.1002/cncr.32198
363. Weigelt B, Geyer FC, Reis-Filho JS. Histological types of breast cancer: How special are they? Mol Oncol. 2010;4: 192–208. doi:10.1016/j.molonc.2010.04.004
364. WHO Classification of Tumours Editorial Board Breast Tumours. Lyon (France): International Agency for Research on Cancer; 2019 (WHO classification of tumours series, 5th ed.; vol. 2), p.87,102-104. https://publications.iarc.fr/581
365. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer. 1957;11: 359–377. doi:10.1038/bjc.1957.43
366. 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: 403–410. doi:10.1111/j.1365-2559.1991.tb00229.x
367. Ellis IO, Rakha EA, Tse GM, Tan PH. An international unified approach to reporting and grading invasive breast cancer. An overview of the International Collaboration on Cancer Reporting (ICCR) initiative. Histopathology. 2023;82: 189–197. doi:10.1111/his.14802
368. Fox SB, Webster F, Chen C-J, Chua B, Collins L, Foschini M-P, et al. Dataset for pathology reporting of ductal carcinoma in situ, variants of lobular carcinoma in situ and low-grade lesions: recommendations from the International Collaboration on Cancer Reporting (ICCR). Histopathology. 2022;81: 467–476. doi:10.1111/his.14725
369. Giuliano AE, Edge SB, Hortobagyi GN. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. Ann Surg Oncol. 2018;25: 1783–1785. doi:10.1245/s10434-018-6486-6
370. Koh J, Kim MJ. Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage. Korean J Radiol. 2019;20: 69–82. doi:10.3348/kjr.2018.0231
371. Morris EA, Schwartz LH, Drotman MB, Kim SJ, Tan LK, Liberman L, et al. Evaluation of pectoralis major muscle in patients with posterior breast tumors on breast MR images: early experience. Radiology. 2000;214: 67–72. doi:10.1148/radiology.214.1.r00ja1667
372. Cadiz F, Gormaz JG, Burotto M. Breast Cancer Staging: Is TNM Ready to Evolve? J Glob Oncol. 2017;4: JGO.17.00004. doi:10.1200/JGO.17.00004
373. Allred DC, Bustamante MA, Daniel CO, Gaskill HV, Cruz AB. Immunocytochemical analysis of estrogen receptors in human breast carcinomas. Evaluation of 130 cases and review of the

literature regarding concordance with biochemical assay and clinical relevance. Arch Surg. 1990;125: 107–113. doi:10.1001/archsurg.1990.01410130113018
374. Barnes DM, Harris WH, Smith P, Millis RR, Rubens RD. Immunohistochemical determination of oestrogen receptor: comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. Br J Cancer. 1996;74: 1445–1451. doi:10.1038/bjc.1996.563
375. Thomssen C, Balic M, Harbeck N, Gnant M. St. Gallen/Vienna 2021: A Brief Summary of the Consensus Discussion on Customizing Therapies for Women with Early Breast Cancer. Breast Care. 2021;16: 135–143. doi:10.1159/000516114
376. Curigliano G, Burstein HJ, Gnant M, Loibl S, Cameron D, Regan MM, et al. Understanding breast cancer complexity to improve patient outcomes: The St Gallen International Consensus Conference for the Primary Therapy of Individuals with Early Breast Cancer 2023. Annals of Oncology. 2023;34: 970–986. doi:10.1016/j.annonc.2023.08.017
377. nationellt-vardprogram-brostcancer.pdf. Available: https://www.swebcg.se/wp-content/uploads/2016/09/nationellt-vardprogram-brostcancer.pdf
378. Memon R, Prieto Granada CN, Harada S, Winokur T, Reddy V, Kahn AG, et al. Discordance Between Immunohistochemistry and In Situ Hybridization to Detect HER2 Overexpression/Gene Amplification in Breast Cancer in the Modern Age: A Single Institution Experience and Pooled Literature Review Study. Clinical Breast Cancer. 2022;22: e123–e133. doi:10.1016/j.clbc.2021.05.004
379. Curigliano G, Burstein HJ, P Winer E, Gnant M, Dubsky P, Loibl S, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. Ann Oncol. 2019;30: 1181. doi:10.1093/annonc/mdy537
380. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Nat Rev Dis Primers. 2019;5: 1–31. doi:10.1038/s41572-019-0111-2
381. Davidson BA, Croessmann S, Park BH. The breast is yet to come: current and future utility of circulating tumour DNA in breast cancer. Br J Cancer. 2021;125: 780–788. doi:10.1038/s41416-021-01422-w
382. Venetis K, Cursano G, Pescia C, D'Ercole M, Porta FM, Blanco MC, et al. Liquid biopsy: Cell-free DNA based analysis in breast cancer. The Journal of Liquid Biopsy. 2023;1: 100002. doi:10.1016/j.jlb.2023.100002
383. Kuderer NM, Burton KA, Blau S, Rose AL, Parker S, Lyman GH, et al. Comparison of 2 Commercially Available Next-Generation Sequencing Platforms in Oncology. JAMA Oncol. 2017;3: 996–998. doi:10.1001/jamaoncol.2016.4983
384. Krag DN, Anderson SJ, Julian TB, Brown AM, Harlow SP, Costantino JP, et al. Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the NSABP B-32 randomised phase 3 trial. Lancet Oncol. 2010;11: 927–933. doi:10.1016/S1470-2045(10)70207-2
385. McLaughlin SA. Surgical management of the breast: breast conservation therapy and mastectomy. Surg Clin North Am. 2013;93: 411–428. doi:10.1016/j.suc.2012.12.006
386. Buchholz TA, Mittendorf EA, Hunt KK. Surgical Considerations After Neoadjuvant Chemotherapy: Breast Conservation Therapy. J Natl Cancer Inst Monogr. 2015;2015: 11–14. doi:10.1093/jncimonographs/lgv020
387. De-Colle C, Kirby A, Russell N, Shaitelman SF, Currey A, Donovan E, et al. Adaptive radiotherapy for breast cancer. Clinical and Translational Radiation Oncology. 2023;39: 100564. doi:10.1016/j.ctro.2022.100564
388. Ebctcg (Early Breast Cancer Trialists' Collaborative Group). Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials. The Lancet. 2011;378: 1707–1716. doi:10.1016/S0140-6736(11)61629-2
389. Ebctcg (Early Breast Cancer Trialists' Collaborative Group). Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. The Lancet. 2014;383: 2127–2135. doi:10.1016/S0140-6736(14)60488-8
390. Golshan M, Cirrincione CT, Sikov WM, Carey LA, Berry DA, Overmoyer B, et al. Impact of neoadjuvant therapy on eligibility for and frequency of breast conservation in stage II-III HER2-

positive breast cancer: surgical results of CALGB 40601 (Alliance). Breast Cancer Res Treat. 2016;160: 297–304. doi:10.1007/s10549-016-4006-6
391. Rastogi P, Anderson SJ, Bear HD, Geyer CE, Kahlenberg MS, Robidoux A, et al. Preoperative chemotherapy: updates of National Surgical Adjuvant Breast and Bowel Project Protocols B-18 and B-27. J Clin Oncol. 2008;26: 778–785. doi:10.1200/JCO.2007.15.0235
392. von Minckwitz G, Untch M, Blohmer J-U, Costa SD, Eidtmann H, Fasching PA, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol. 2012;30: 1796–1804. doi:10.1200/JCO.2011.38.8595
393. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 2014;384: 164–172. doi:10.1016/S0140-6736(13)62422-8
394. von Minckwitz G, Huang C-S, Mano MS, Loibl S, Mamounas EP, Untch M, et al. Trastuzumab Emtansine for Residual Invasive HER2-Positive Breast Cancer. N Engl J Med. 2019;380: 617–628. doi:10.1056/NEJMoa1814017
395. De Wilde RL, Devassy R, Torres-de la Roche LA, Krentel H, Tica V, Cezar C. Guidance and Standards for Breast Cancer Care in Europe. J Obstet Gynaecol India. 2020;70: 330–336. doi:10.1007/s13224-020-01316-6
396. National Comprehensive Cancer Network (NCNN) Guidelines. Breast cancer (version 2.2024). Available: https://www.nccn.org/guidelines/guidelines-detail. Visited: 2024-04-23
397. Nitz U, Gluz O, Clemens M, Malter W, Reimer T, Nuding B, et al. West German Study PlanB Trial: Adjuvant Four Cycles of Epirubicin and Cyclophosphamide Plus Docetaxel Versus Six Cycles of Docetaxel and Cyclophosphamide in HER2-Negative Early Breast Cancer. J Clin Oncol. 2019;37: 799–808. doi:10.1200/JCO.18.00028
398. Poggio F, Tagliamento M, Ceppi M, Bruzzone M, Conte B, Fregatti P, et al. Adding a platinum agent to neoadjuvant chemotherapy for triple-negative breast cancer: the end of the debate. Ann Oncol. 2022;33: 347–349. doi:10.1016/j.annonc.2021.11.016
399. Francis PA, Pagani O, Fleming GF, Walley BA, Colleoni M, Láng I, et al. Tailoring Adjuvant Endocrine Therapy for Premenopausal Breast Cancer. N Engl J Med. 2018;379: 122–137. doi:10.1056/NEJMoa1803164
400. Ebctcg (Early Breast Cancer Trialists' Collaborative Group). Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. The Lancet. 2015;386: 1341–1352. doi:10.1016/S0140-6736(15)61074-1
401. Tjan-Heijnen VCG, van Hellemond IEG, Peer PGM, Swinkels ACP, Smorenburg CH, van der Sangen MJC, et al. Extended adjuvant aromatase inhibition after sequential endocrine therapy (DATA): a randomised, phase 3 trial. Lancet Oncol. 2017;18: 1502–1511. doi:10.1016/S1470-2045(17)30600-9
402. von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, et al. Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. N Engl J Med. 2017;377: 122–131. doi:10.1056/NEJMoa1703643
403. Tolaney SM, Barry WT, Dang CT, Yardley DA, Moy B, Marcom PK, et al. Adjuvant paclitaxel and trastuzumab for node-negative, HER2-positive breast cancer. N Engl J Med. 2015;372: 134–141. doi:10.1056/NEJMoa1406281
404. Tolaney SM, Guo H, Pernas S, Barry WT, Dillon DA, Ritterhouse L, et al. Seven-Year Follow-Up Analysis of Adjuvant Paclitaxel and Trastuzumab Trial for Node-Negative, Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer. J Clin Oncol. 2019;37: 1868–1875. doi:10.1200/JCO.19.00066
405. Martin M, Holmes FA, Ejlertsen B, Delaloge S, Moy B, Iwata H, et al. Neratinib after trastuzumab-based adjuvant therapy in HER2-positive breast cancer (ExteNET): 5-year analysis of a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol. 2017;18: 1688–1700. doi:10.1016/S1470-2045(17)30717-9
406. Chan A, Moy B, Mansi J, Ejlertsen B, Holmes FA, Chia S, et al. Final Efficacy Results of Neratinib in HER2-positive Hormone Receptor-positive Early-stage Breast Cancer From the Phase III ExteNET Trial. Clin Breast Cancer. 2021;21: 80-91.e7. doi:10.1016/j.clbc.2020.09.014
407. Cardoso F, Senkus E, Costa A, Papadopoulos E, Aapro M, André F, et al. 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4)†. Ann Oncol. 2018;29: 1634–1657. doi:10.1093/annonc/mdy192

408. Rugo HS, Rumble RB, Macrae E, Barton DL, Connolly HK, Dickler MN, et al. Endocrine Therapy for Hormone Receptor-Positive Metastatic Breast Cancer: American Society of Clinical Oncology Guideline. J Clin Oncol. 2016;34: 3069–3103. doi:10.1200/JCO.2016.67.1487
409. Finn RS, Crown JP, Lang I, Boer K, Bondarenko IM, Kulyk SO, et al. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. Lancet Oncol. 2015;16: 25–35. doi:10.1016/S1470-2045(14)71159-3
410. Turner NC, Slamon DJ, Ro J, Bondarenko I, Im S-A, Masuda N, et al. Overall Survival with Palbociclib and Fulvestrant in Advanced Breast Cancer. N Engl J Med. 2018;379: 1926–1936. doi:10.1056/NEJMoa1810527
411. Im S-A, Lu Y-S, Bardia A, Harbeck N, Colleoni M, Franke F, et al. Overall Survival with Ribociclib plus Endocrine Therapy in Breast Cancer. N Engl J Med. 2019;381: 307–316. doi:10.1056/NEJMoa1903765
412. Swain SM, Miles D, Kim S-B, Im Y-H, Im S-A, Semiglazov V, et al. Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA): end-of-study results from a double-blind, randomised, placebo-controlled, phase 3 study. Lancet Oncol. 2020;21: 519–530. doi:10.1016/S1470-2045(19)30863-0
413. Modi S, Saura C, Yamashita T, Park YH, Kim S-B, Tamura K, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Positive Breast Cancer. N Engl J Med. 2020;382: 610–621. doi:10.1056/NEJMoa1914510
414. Tolaney SM, Barroso-Sousa R, Jiang Z, Park YH, Rimawi M, Manich CS, et al. 328TiP Phase III study of trastuzumab deruxtecan (T-DXd) with or without pertuzumab vs a taxane, trastuzumab and pertuzumab in first-line (1L), human epidermal growth factor receptor 2–positive (HER2+) metastatic breast cancer (mBC): DESTINY-Breast09. Annals of Oncology. 2021;32: S507–S508. doi:10.1016/j.annonc.2021.08.611
415. Xu B, Yan M, Ma F, Hu X, Feng J, Ouyang Q, et al. Pyrotinib plus capecitabine versus lapatinib plus capecitabine for the treatment of HER2-positive metastatic breast cancer (PHOEBE): a multicentre, open-label, randomised, controlled, phase 3 trial. Lancet Oncol. 2021;22: 351–360. doi:10.1016/S1470-2045(20)30702-6
416. Schmid P, Rugo HS, Adams S, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol. 2020;21: 44–59. doi:10.1016/S1470-2045(19)30689-8
417. Emens LA, Adams S, Barrios CH, Diéras V, Iwata H, Loi S, et al. First-line atezolizumab plus nab-paclitaxel for unresectable, locally advanced, or metastatic triple-negative breast cancer: IMpassion130 final overall survival analysis. Ann Oncol. 2021;32: 983–993. doi:10.1016/j.annonc.2021.05.355
418. Bardia A, Hurvitz SA, Tolaney SM, Loirat D, Punie K, Oliveira M, et al. Sacituzumab Govitecan in Metastatic Triple-Negative Breast Cancer. N Engl J Med. 2021;384: 1529–1541. doi:10.1056/NEJMoa2028485
419. Litton JK, Rugo HS, Ettl J, Hurvitz SA, Gonçalves A, Lee K-H, et al. Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. N Engl J Med. 2018;379: 753–763. doi:10.1056/NEJMoa1802905
420. Litton JK, Hurvitz SA, Mina LA, Rugo HS, Lee K-H, Gonçalves A, et al. Talazoparib versus chemotherapy in patients with germline BRCA1/2-mutated HER2-negative advanced breast cancer: final overall survival results from the EMBRACA trial. Ann Oncol. 2020;31: 1526–1535. doi:10.1016/j.annonc.2020.08.2098
421. Cortes J, O'Shaughnessy J, Loesch D, Blum JL, Vahdat LT, Petrakova K, et al. Eribulin monotherapy versus treatment of physician's choice in patients with metastatic breast cancer (EMBRACE): a phase 3 open-label randomised study. Lancet. 2011;377: 914–923. doi:10.1016/S0140-6736(11)60070-6
422. Xu B, Sun T, Zhang Q, Zhang P, Yuan Z, Jiang Z, et al. Efficacy of utidelone plus capecitabine versus capecitabine for heavily pretreated, anthracycline- and taxane-refractory metastatic breast cancer: final analysis of overall survival in a phase III randomised controlled trial. Ann Oncol. 2021;32: 218–228. doi:10.1016/j.annonc.2020.10.600
423. Meyerholz DK, Beck AP. Principles and approaches for reproducible scoring of tissue stains in research. Lab Invest. 2018;98: 844–855. doi:10.1038/s41374-018-0057-0

424. Kim S-W, Roh J, Park C-S. Immunohistochemistry for Pathologists: Protocols, Pitfalls, and Tips. J Pathol Transl Med. 2016;50: 411–418. doi:10.4132/jptm.2016.08.08
425. Furrer D, Sanschagrin F, Jacob S, Diorio C. Advantages and Disadvantages of Technologies for HER2 Testing in Breast Cancer Specimens. American Journal of Clinical Pathology. 2015;144: 686–703. doi:10.1309/AJCPT41TCBUEVDQC
426. Wu LR, Dai P, Wang MX, Chen SX, Cohen EN, Jayachandran G, et al. Ensemble of nucleic acid absolute quantitation modules for copy number variation detection and RNA profiling. Nat Commun. 2022;13: 1791. doi:10.1038/s41467-022-29487-y
427. Nazir S. Medical diagnostic value of digital PCR (dPCR): A systematic review. Biomedical Engineering Advances. 2023;6: 100092. doi:10.1016/j.bea.2023.100092
428. Coccaro N, Tota G, Anelli L, Zagaria A, Specchia G, Albano F. Digital PCR: A Reliable Tool for Analyzing and Monitoring Hematologic Malignancies. Int J Mol Sci. 2020;21: 3141. doi:10.3390/ijms21093141
429. Prasad V, Kaestner V, Mailankody S. Cancer Drugs Approved Based on Biomarkers and Not Tumor Type-FDA Approval of Pembrolizumab for Mismatch Repair-Deficient Solid Cancers. JAMA Oncol. 2018;4: 157–158. doi:10.1001/jamaoncol.2017.4182
430. Raedler LA. Opdivo (Nivolumab): Second PD-1 Inhibitor Receives FDA Approval for Unresectable or Metastatic Melanoma. Am Health Drug Benefits. 2015;8: 180–183.
431. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, et al. Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children. N Engl J Med. 2018;378: 731–739. doi:10.1056/NEJMoa1714448
432. Wolff L, Kiesewetter B. Applicability of ESMO-MCBS and ESCAT for molecular tumor boards. memo. 2022;15: 190–195. doi:10.1007/s12254-022-00800-1
433. Subhan MA, Parveen F, Shah H, Yalamarty SSK, Ataide JA, Torchilin VP. Recent Advances with Precision Medicine Treatment for Breast Cancer including Triple-Negative Sub-Type. Cancers (Basel). 2023;15: 2204. doi:10.3390/cancers15082204
434. Saal LH, Vallon-Christersson J, Häkkinen J, Hegardt C, Grabau D, Winter C, et al. The Sweden Cancerome Analysis Network - Breast (SCAN-B) Initiative: a large-scale multicenter infrastructure towards implementation of breast cancer genomic analyses in the clinical routine. Genome Med. 2015;7: 20. doi:10.1186/s13073-015-0131-9
435. Rydén L, Loman N, Larsson C, Hegardt C, Vallon-Christersson J, Malmberg M, et al. Minimizing inequality in access to precision medicine in breast cancer by real-time population-based molecular analysis in the SCAN-B initiative. Br J Surg. 2018;105: e158–e168. doi:10.1002/bjs.10741
436. Xie Y, Davis Lynn BC, Moir N, Cameron DA, Figueroa JD, Sims AH. Breast cancer gene expression datasets do not reflect the disease at the population level. NPJ Breast Cancer. 2020;6: 39. doi:10.1038/s41523-020-00180-x
437. Staaf J, Glodzik D, Bosch A, Vallon-Christersson J, Reuterswärd C, Häkkinen J, et al. Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. Nat Med. 2019;25: 1526–1533. doi:10.1038/s41591-019-0582-4
438. Glodzik D, Bosch A, Hartman J, Aine M, Vallon-Christersson J, Reuterswärd C, et al. Comprehensive molecular comparison of BRCA1 hypermethylated and BRCA1 mutated triple negative breast cancers. Nat Commun. 2020;11: 3747. doi:10.1038/s41467-020-17537-2
439. Vallon-Christersson J, Häkkinen J, Hegardt C, Saal LH, Larsson C, Ehinger A, et al. Cross comparison and prognostic assessment of breast cancer multigene signatures in a large population-based contemporary clinical series. Sci Rep. 2019;9: 12184. doi:10.1038/s41598-019-48570-x
440. Persson H, Søkilde R, Häkkinen J, Pirona AC, Vallon-Christersson J, Kvist A, et al. Frequent miRNA-convergent fusion gene events in breast cancer. Nat Commun. 2017;8: 788. doi:10.1038/s41467-017-01176-1
441. Lundgren C, Bendahl P-O, Borg Å, Ehinger A, Hegardt C, Larsson C, et al. Agreement between molecular subtyping and surrogate subtype classification: a contemporary population-based study of ER-positive/HER2-negative primary breast cancer. Breast Cancer Res Treat. 2019;178: 459–467. doi:10.1007/s10549-019-05378-7
442. Søkilde R, Persson H, Ehinger A, Pirona AC, Fernö M, Hegardt C, et al. Refinement of breast cancer molecular classification by miRNA expression profiles. BMC Genomics. 2019;20: 503. doi:10.1186/s12864-019-5887-7
443. Veerla S, Hohmann L, Nacer DF, Vallon-Christersson J, Staaf J. Perturbation and stability of PAM50 subtyping in population-based primary invasive breast cancer. NPJ Breast Cancer. 2023;9: 83. doi:10.1038/s41523-023-00589-0

444. Dihge L, Vallon-Christersson J, Hegardt C, Saal LH, Häkkinen J, Larsson C, et al. Prediction of Lymph Node Metastasis in Breast Cancer by Gene Expression and Clinicopathological Models: Development and Validation within a Population-Based Cohort. Clinical Cancer Research. 2019;25: 6368–6381. doi:10.1158/1078-0432.CCR-19-0075
445. Brueffer C, Gladchuk S, Winter C, Vallon-Christersson J, Hegardt C, Häkkinen J, et al. The mutational landscape of the SCAN-B real-world primary breast cancer transcriptome. EMBO Molecular Medicine. 2020;12: e12118. doi:10.15252/emmm.202012118
446. Dahlgren M, George AM, Brueffer C, Gladchuk S, Chen Y, Vallon-Christersson J, et al. Preexisting Somatic Mutations of Estrogen Receptor Alpha (ESR1) in Early-Stage Primary Breast Cancer. JNCI Cancer Spectrum. 2021;5: pkab028. doi:10.1093/jncics/pkab028
447. Förnvik D, Aaltonen KE, Chen Y, George AM, Brueffer C, Rigo R, et al. Detection of circulating tumor cells and circulating tumor DNA before and after mammographic breast compression in a cohort of breast cancer patients scheduled for neoadjuvant treatment. Breast Cancer Res Treat. 2019;177: 447–455. doi:10.1007/s10549-019-05326-5
448. Ringnér M, Fredlund E, Häkkinen J, Borg Å, Staaf J. GOBO: Gene Expression-Based Outcome for Breast Cancer Online. PLOS ONE. 2011;6: e17911. doi:10.1371/journal.pone.0017911
449. 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: R65. doi:10.1186/bcr2124
450. O´Leary PC, Penny SA, Dolan RT, Kelly CM, Madden SF, Rexhepaj E, et al. Systematic antibody generation and validation via tissue microarray technology leading to identification of a novel protein prognostic panel in breast cancer. BMC Cancer. 2013;13: 175. doi:10.1186/1471-2407-13-175
451. Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. Science. 1998;281: 363, 365. doi:10.1126/science.281.5375.363
452. Buermans HPJ, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2014;1842: 1932–1941. doi:10.1016/j.bbadis.2014.06.015
453. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975;94: 441–448. doi:10.1016/0022-2836(75)90213-2
454. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74: 5463–5467. doi:10.1073/pnas.74.12.5463
455. Maxam AM, Gilbert W. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 1980;65: 499–560. doi:10.1016/s0076-6879(80)65059-9
456. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431: 931–945. doi:10.1038/nature03001
457. Qin D. Next-generation sequencing and its clinical application. Cancer Biol Med. 2019;16: 4–10. doi:10.20892/j.issn.2095-3941.2018.0055
458. Balasubramanian S. Sequencing nucleic acids: from chemistry to medicine. Chem Commun (Camb). 2011;47: 7281–7286. doi:10.1039/c1cc11078k
459. Slatko BE, Gardner AF, Ausubel FM. Overview of Next Generation Sequencing Technologies. Curr Protoc Mol Biol. 2018;122: e59. doi:10.1002/cpmb.59
460. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13: 341. doi:10.1186/1471-2164-13-341
461. Gyarmati P, Song Y, Hällman J, Käller M. Chemical fragmentation for massively parallel sequencing library preparation. Journal of Biotechnology. 2013;168: 95–100. doi:10.1016/j.jbiotec.2013.08.020
462. Ribarska T, Bjørnstad PM, Sundaram AYM, Gilfillan GD. Optimization of enzymatic fragmentation is crucial to maximize genome coverage: a comparison of library preparation methods for Illumina sequencing. BMC Genomics. 2022;23: 92. doi:10.1186/s12864-022-08316-y
463. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res. 2009;37: e123. doi:10.1093/nar/gkp596
464. Khan J, Saal LH, Bittner ML, Jiang Y, Gooden GC, Glatfelter AA, et al. Gene expression profiling in cancer using cDNA microarrays. Methods Mol Med. 2002;68: 205–222. doi:10.1385/1-59259-135-3:205

465. Mills JD, Kawahara Y, Janitz M. Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. Curr Genomics. 2013;14: 173–181. doi:10.2174/1389202911314030003
466. Negi A, Shukla A, Jaiswar A, Shrinet J, Jasrotia RS. Chapter 6 - Applications and challenges of microarray and RNA-sequencing. In: Singh DB, Pathak RK, editors. Bioinformatics. Academic Press; 2022. pp. 91–103. doi:10.1016/B978-0-323-89775-4.00016-X
467. Rao MS, Van Vleet TR, Ciurlionis R, Buck WR, Mittelstadt SW, Blomme EAG, et al. Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. Front Genet. 2019;9: 636. doi:10.3389/fgene.2018.00636
468. Corchete LA, Rojas EA, Alonso-López D, De Las Rivas J, Gutiérrez NC, Burguillo FJ. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. Sci Rep. 2020;10: 19737. doi:10.1038/s41598-020-76881-x
469. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. Cold Spring Harb Protoc. 2015;2015: 951–969. doi:10.1101/pdb.top084970
470. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced Applications of RNA Sequencing and Challenges. Bioinform Biol Insights. 2015;9: 29–46. doi:10.4137/BBI.S28991
471. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biology. 2016;17: 13. doi:10.1186/s13059-016-0881-8
472. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet. 2016;17: 257–271. doi:10.1038/nrg.2016.10
473. Slatko BE, Gardner AF, Ausubel FM. Overview of Next Generation Sequencing Technologies. Curr Protoc Mol Biol. 2018;122: e59. doi:10.1002/cpmb.59
474. Häkkinen J, Nordborg N, Månsson O, Vallon-Christersson J. Implementation of an Open Source Software solution for Laboratory Information Management and automated RNAseq data analysis in a large-scale Cancer Genomics initiative using BASE with extension package Reggie. bioRxiv; 2016. p. 038976. doi:10.1101/038976
475. Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg Å, Peterson C. BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. Genome Biology. 2002;3: software0003.1. doi:10.1186/gb-2002-3-8-software0003
476. Vallon-Christersson J, Nordborg N, Svensson M, Häkkinen J. BASE--2nd generation software for microarray data management and analysis. BMC Bioinformatics. 2009;10: 330. doi:10.1186/1471-2105-10-330
477. Picard toolkit. Broad Institute, GitHub repository. Broad Institute; 2019. Available: http://broadinstitute.github.io/picard/
478. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
479. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2: automated genomic discovery of transposable element families. Genomics; 2019 Nov. doi:10.1101/856591
480. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359. doi:10.1038/nmeth.1923
481. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
482. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37: 907–915. doi:10.1038/s41587-019-0201-4
483. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002;12: 996–1006. doi:10.1101/gr.229102
484. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33: 290–295. doi:10.1038/nbt.3122
485. Zhao S, Ye Z, Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. RNA. 2020;26: 903–909. doi:10.1261/rna.074922.120
486. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research; 2016. doi:10.12688/f1000research.7563.1
487. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11: 1650–1667. doi:10.1038/nprot.2016.095

488. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47. doi:10.1093/nar/gkv007
489. Differential expression of FPKM from RNA-seq data using limma and voom(). [cited 5 May 2024]. Available: https://support.bioconductor.org/p/56275/#56299
490. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology. 2014;15: R29. doi:10.1186/gb-2014-15-2-r29
491. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2021; 060012. doi:10.1101/060012
492. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Research. 2017;45: W130–W137. doi:10.1093/nar/gkx356
493. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol. 1986;51 Pt 1: 263–273. doi:10.1101/sqb.1986.051.01.032
494. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 1988;239: 487–491. doi:10.1126/science.2448875
495. Chien A, Edgar DB, Trela JM. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. J Bacteriol. 1976;127: 1550–1557.
496. Majumdar N, Banerjee S, Pallas M, Wessel T, Hegerich P. Poisson Plus Quantification for Digital PCR Systems. Sci Rep. 2017;7: 9617. doi:10.1038/s41598-017-09183-4
497. Quan P-L, Sauzade M, Brouzes E. dPCR: A Technology Review. Sensors (Basel). 2018;18: 1271. doi:10.3390/s18041271
498. Sánchez-Martín V, López-López E, Reguero-Paredes D, Godoy-Ortiz A, Domínguez-Recio ME, Jiménez-Rodríguez B, et al. Comparative study of droplet-digital PCR and absolute Q digital PCR for ctDNA detection in early-stage breast cancer patients. Clinica Chimica Acta. 2024;552: 117673. doi:10.1016/j.cca.2023.117673
499. Taylor SC, Laperriere G, Germain H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. Sci Rep. 2017;7: 2409. doi:10.1038/s41598-017-02217-x
500. Krasnov GS, Kudryavtseva AV, Snezhkina AV, Lakunina VA, Beniaminov AD, Melnikova NV, et al. Pan-Cancer Analysis of TCGA Data Revealed Promising Reference Genes for qPCR Normalization. Front Genet. 2019;10. doi:10.3389/fgene.2019.00097
501. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, et al. Redefine statistical significance. Nat Hum Behav. 2018;2: 6–10. doi:10.1038/s41562-017-0189-z
502. Winters R, Winters A, Amedee RG. Statistics: A Brief Overview. Ochsner J. 2010;10: 213–216.
503. Gourgou-Bourgade S, Cameron D, Poortmans P, Asselain B, Azria D, Cardoso F, et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials)†. Ann Oncol. 2015;26: 873–879. doi:10.1093/annonc/mdv106
504. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53: 457–481. doi:10.2307/2281868
505. Rich JT, Neely JG, Paniello RC, Voelker CCJ, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. Otolaryngol Head Neck Surg. 2010;143: 331–336. doi:10.1016/j.otohns.2010.05.007
506. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B (Methodological). 1972;34: 187–220.
507. Zhang M-J. Cox Proportional Hazards Regression Models for Survival Data in Cancer Research. In: Beam C, editor. Biostatistical Applications in Cancer Research. Boston, MA: Springer US; 2002. pp. 59–70. doi:10.1007/978-1-4757-3571-0_4
508. Lewoniewska S, Oscilowska I, Forlino A, Palka J. Understanding the Role of Estrogen Receptor Status in PRODH/POX-Dependent Apoptosis/Survival in Breast Cancer Cells. Biology (Basel). 2021;10: 1314. doi:10.3390/biology10121314
509. Pons DG, Torrens-Mas M, Nadal-Serrano M, Sastre-Serra J, Roca P, Oliver J. The presence of estrogen receptor β modulates the response of breast cancer cells to therapeutic agents. Int J Biochem Cell Biol. 2015;66: 85–94. doi:10.1016/j.biocel.2015.07.014

510. Song P, Li Y, Dong Y, Liang Y, Qu H, Qi D, et al. Estrogen receptor β inhibits breast cancer cells migration and invasion through CLDN6-mediated autophagy. J Exp Clin Cancer Res. 2019;38: 354. doi:10.1186/s13046-019-1359-9
511. Nelson AW, Groen AJ, Miller JL, Warren AY, Holmes KA, Tarulli GA, et al. Comprehensive assessment of estrogen receptor beta antibodies in cancer cell line models and tissue reveals critical limitations in reagent specificity. Mol Cell Endocrinol. 2017;440: 138–150. doi:10.1016/j.mce.2016.11.016
512. Hawse JR, Carter JM, Aspros KGM, Bruinsma ES, Koepplin JW, Negron V, et al. Optimized immunohistochemical detection of estrogen receptor beta using two validated monoclonal antibodies confirms its expression in normal and malignant breast tissues. Breast Cancer Res Treat. 2020;179: 241–249. doi:10.1007/s10549-019-05441-3
513. Wegler C, Ölander M, Wiśniewski JR, Lundquist P, Zettl K, Åsberg A, et al. Global variability analysis of mRNA and protein concentrations across and within human tissues. NAR Genomics and Bioinformatics. 2020;2: lqz010. doi:10.1093/nargab/lqz010
514. Meng W, Liao Y, Chen J, Wang Y, Meng Y, Li K, et al. Upregulation of estrogen receptor beta protein but not mRNA predicts poor prognosis and may be associated with enhanced translation in non-small cell lung cancer: a systematic review and meta-analysis. J Thorac Dis. 2021;13: 4281–4300. doi:10.21037/jtd-21-658
515. Reese JM, Bruinsma ES, Nelson AW, Chernukhin I, Carroll JS, Li Y, et al. ERβ-mediated induction of cystatins results in suppression of TGFβ signaling and inhibition of triple-negative breast cancer metastasis. Proc Natl Acad Sci U S A. 2018;115: E9580–E9589. doi:10.1073/pnas.1807751115
516. Baba AB, Rah B, Bhat GhR, Mushtaq I, Parveen S, Hassan R, et al. Transforming Growth Factor-Beta (TGF-β) Signaling in Cancer-A Betrayal Within. Frontiers in Pharmacology. 2022;13. Available: https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2022.791272
517. Goutsouliak K, Veeraraghavan J, Sethunath V, De Angelis C, Osborne CK, Rimawi MF, et al. Towards personalized treatment for early stage HER2-positive breast cancer. Nat Rev Clin Oncol. 2020;17: 233–250. doi:10.1038/s41571-019-0299-9
518. Perez EA, Romond EH, Suman VJ, Jeong J-H, Sledge G, Geyer CE, et al. Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. J Clin Oncol. 2014;32: 3744–3752. doi:10.1200/JCO.2014.55.5730
519. Bradley R, Braybrooke J, Gray R, Hills R, Liu Z, Peto R, et al. Trastuzumab for early-stage, HER2-positive breast cancer: a meta-analysis of 13 864 women in seven randomised trials. The Lancet Oncology. 2021;22: 1139–1150. doi:10.1016/S1470-2045(21)00288-6
520. Salvo EM, Ramirez AO, Cueto J, Law EH, Situ A, Cameron C, et al. Risk of recurrence among patients with HR-positive, HER2-negative, early breast cancer receiving adjuvant endocrine therapy: A systematic review and meta-analysis. Breast. 2021;57: 5–17. doi:10.1016/j.breast.2021.02.009
521. Rönnlund C, Sifakis EG, Schagerholm C, Yang Q, Karlsson E, Chen X, et al. Prognostic impact of HER2 biomarker levels in trastuzumab-treated early HER2-positive breast cancer. Breast Cancer Research. 2024;26: 24. doi:10.1186/s13058-024-01779-9
522. Yahata T, Shao W, Endoh H, Hur J, Coser KR, Sun H, et al. Selective coactivation of estrogen-dependent transcription by CITED1 CBP/p300-binding protein. Genes Dev. 2001;15: 2598–2612. doi:10.1101/gad.906301
523. McBryan J, Howlin J, Kenny PA, Shioda T, Martin F. ERalpha-CITED1 co-regulated genes expressed during pubertal mammary gland development: implications for breast cancer prognosis. Oncogene. 2007;26: 6406–6419. doi:10.1038/sj.onc.1210468
524. Curigliano G, Dent R, Earle H, Modi S, Tarantino P, Viale G, et al. Open questions, current challenges, and future perspectives in targeting human epidermal growth factor receptor 2-low breast cancer. ESMO Open. 2024;9. doi:10.1016/j.esmoop.2024.102989
525. Tarantino P, Viale G, Press MF, Hu X, Penault-Llorca F, Bardia A, et al. ESMO expert consensus statements (ECS) on the definition, diagnosis, and management of HER2-low breast cancer. Annals of Oncology. 2023;34: 645–659. doi:10.1016/j.annonc.2023.05.008
526. Gennari A, André F, Barrios CH, Cortés J, de Azambuja E, DeMichele A, et al. ESMO Clinical Practice Guideline for the diagnosis, staging and treatment of patients with metastatic breast cancer. Ann Oncol. 2021;32: 1475–1495. doi:10.1016/j.annonc.2021.09.019
527. Rossi C, Cicalini I, Cufaro MC, Consalvo A, Upadhyaya P, Sala G, et al. Breast cancer in the era of integrating "Omics" approaches. Oncogenesis. 2022;11: 1–13. doi:10.1038/s41389-022-00393-8

528. Orsini A, Diquigiovanni C, Bonora E. Omics Technologies Improving Breast Cancer Research and Diagnostics. International Journal of Molecular Sciences. 2023;24: 12690. doi:10.3390/ijms241612690
529. Cheng C, Fei Z, Xiao P. Methods to improve the accuracy of next-generation sequencing. Front Bioeng Biotechnol. 2023;11. doi:10.3389/fbioe.2023.982111
530. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012: 251364. doi:10.1155/2012/251364
531. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30: 418–426. doi:10.1016/j.tig.2014.07.001
532. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, et al. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nat Commun. 2019;10: 754. doi:10.1038/s41467-019-08734-9
533. Zheng P, Zhou C, Ding Y, Liu B, Lu L, Zhu F, et al. Nanopore sequencing technology and its applications. MedComm (2020). 2023;4: e316. doi:10.1002/mco2.316
534. Brinkmann A, Ulm S-L, Uddin S, Förster S, Seifert D, Oehme R, et al. AmpliCoV: Rapid Whole-Genome Sequencing Using Multiplex PCR Amplification and Real-Time Oxford Nanopore MinION Sequencing Enables Rapid Variant Identification of SARS-CoV-2. Front Microbiol. 2021;12: 651151. doi:10.3389/fmicb.2021.651151
535. Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A. Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics. Life (Basel). 2021;12: 30. doi:10.3390/life12010030
536. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biology. 2020;21: 30. doi:10.1186/s13059-020-1935-5
537. Ren X, Zhang L, Zhang Y, Li Z, Siemers N, Zhang Z. Insights Gained from Single-Cell Analysis of Immune Cells in the Tumor Microenvironment. Annu Rev Immunol. 2021;39: 583–609. doi:10.1146/annurev-immunol-110519-071134
538. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science. 2016;353: 78–82. doi:10.1126/science.aaf2403
539. Andersson A, Larsson L, Stenbeck L, Salmén F, Ehinger A, Wu SZ, et al. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. Nat Commun. 2021;12: 6012. doi:10.1038/s41467-021-26271-2
540. Bassiouni R, Idowu MO, Gibbs LD, Robila V, Grizzard PJ, Webb MG, et al. Spatial Transcriptomic Analysis of a Diverse Patient Cohort Reveals a Conserved Architecture in Triple-Negative Breast Cancer. Cancer Research. 2023;83: 34–48. doi:10.1158/0008-5472.CAN-22-2682
541. Wang XQ, Danenberg E, Huang C-S, Egle D, Callari M, Bermejo B, et al. Spatial predictors of immunotherapy response in triple-negative breast cancer. Nature. 2023;621: 868–876. doi:10.1038/s41586-023-06498-3
542. Romanens L, Chaskar P, Marcone R, Ryser S, Tille J-C, Genolet R, et al. Clonal expansion of intra-epithelial T cells in breast cancer revealed by spatial transcriptomics. International Journal of Cancer. 2023;153: 1568–1578. doi:10.1002/ijc.34620
543. Jiménez-Santos MJ, García-Martín S, Rubio-Fernández M, Gómez-López G, Al-Shahrour F. Spatial Transcriptomics in Breast Cancer Reveals Tumour Microenvironment-Driven Drug Responses and Clonal Therapeutic Heterogeneity. bioRxiv; 2024. p. 2024.02.18.580660. doi:10.1101/2024.02.18.580660
544. Coutant A, Cockenpot V, Muller L, Degletagne C, Pommier R, Tonon L, et al. Spatial Transcriptomics Reveal Pitfalls and Opportunities for the Detection of Rare High-Plasticity Breast Cancer Subtypes. Laboratory Investigation. 2023;103: 100258. doi:10.1016/j.labinv.2023.100258
545. Wu SZ, Al-Eryani G, Roden D, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. Nat Genet. 2021;53: 1334–1347. doi:10.1038/s41588-021-00911-1
546. Huang Y, Zeng P, Zhong C. Classifying breast cancer subtypes on multi-omics data via sparse canonical correlation analysis and deep learning. BMC Bioinformatics. 2024;25: 132. doi:10.1186/s12859-024-05749-y
547. Choi JM, Chae H. moBRCA-net: a breast cancer subtype classification framework based on multi-omics attention neural networks. BMC Bioinformatics. 2023;24: 169. doi:10.1186/s12859-023-05273-5

548. Chai H, Zhou X, Zhang Z, Rao J, Zhao H, Yang Y. Integrating multi-omics data through deep learning for accurate cancer prognosis prediction. Comput Biol Med. 2021;134: 104481. doi:10.1016/j.compbiomed.2021.104481
549. Huang Z, Zhan X, Xiang S, Johnson TS, Helm B, Yu CY, et al. SALMON: Survival Analysis Learning With Multi-Omics Neural Networks on Breast Cancer. Front Genet. 2019;10. doi:10.3389/fgene.2019.00166
550. Jiang L, Xu C, Bai Y, Liu A, Gong Y, Wang Y-P, et al. Autosurv: interpretable deep learning framework for cancer survival analysis incorporating clinical and multi-omics data. npj Precis Onc. 2024;8: 1–16. doi:10.1038/s41698-023-00494-6
551. Mandair D, Reis-Filho JS, Ashworth A. Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. npj Breast Cancer. 2023;9: 1–11. doi:10.1038/s41523-023-00518-1
552. Gamble P, Jaroensri R, Wang H, Tan F, Moran M, Brown T, et al. Determining breast cancer biomarker status and associated morphological features using deep learning. Commun Med. 2021;1: 1–12. doi:10.1038/s43856-021-00013-3
553. Sammut S-J, Crispin-Ortuzar M, Chin S-F, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature. 2022;601: 623–629. doi:10.1038/s41586-021-04278-5
554. Somasundaram DB, Aravindan S, Yu Z, Jayaraman M, Tran NTB, Li S, et al. Droplet digital PCR as an alternative to FISH for MYCN amplification detection in human neuroblastoma FFPE samples. BMC Cancer. 2019;19: 106. doi:10.1186/s12885-019-5306-0
555. Suryavanshi M, Jaipuria J, Mehta A, Kumar D, Panigrahi MK, Verma H, et al. Droplet digital polymerase chain reaction offers an improvisation over conventional immunohistochemistry and fluorescent in situ hybridization for ascertaining Her2 status of breast cancer. South Asian J Cancer. 2019;8: 203–210. doi:10.4103/sajc.sajc_344_18
556. Taylor SC, Laperriere G, Germain H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. Sci Rep. 2017;7: 2409. doi:10.1038/s41598-017-02217-x
557. Magbanua MJM, Swigart LB, Wu H-T, Hirst GL, Yau C, Wolf DM, et al. Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. Annals of Oncology. 2021;32: 229–239. doi:10.1016/j.annonc.2020.11.007
558. Peng Y, Mei W, Ma K, Zeng C. Circulating Tumor DNA and Minimal Residual Disease (MRD) in Solid Tumors: Current Horizons and Future Perspectives. Front Oncol. 2021;11. doi:10.3389/fonc.2021.763790
559. Serafini MS, Molteni E, Nicolò E, Gerratana L, Reduzzi C, Cristofanilli M. Cellular residual disease (CRD) in early breast cancer –Expanding the concept of minimal residual disease monitoring? The Journal of Liquid Biopsy. 2024;3: 100132. doi:10.1016/j.jlb.2023.100132
560. Baksh M, Mahajan B, Dufresne MM, Shoukry MM, Nussbaum S, Abbaszadeh-Kasbi A, et al. Circulating tumor DNA for breast cancer: Review of active clinical trials. Cancer Treatment and Research Communications. 2022;32: 100609. doi:10.1016/j.ctarc.2022.100609

Studies I - V