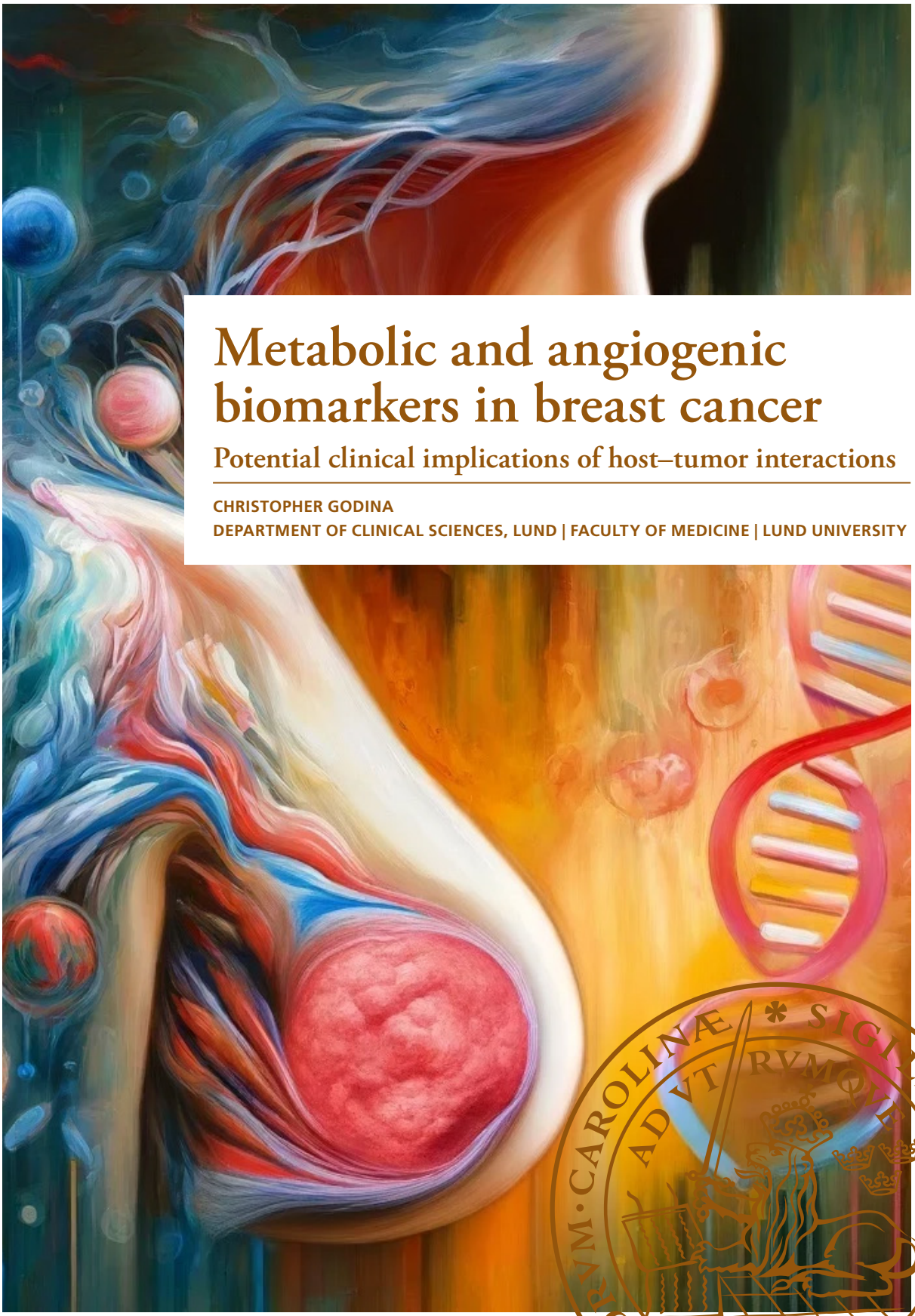
Metabolic and angiogenic biomarkers in breast cancer: potential clinical implications of host–tumor interactions

Godina, Christopher

2024

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



Metabolic and angiogenic biomarkers in breast cancer

Potential clinical implications of host–tumor interactions

CHRISTOPHER GODINA
DEPARTMENT OF CLINICAL SCIENCES, LUND | FACULTY OF MEDICINE | LUND UNIVERSITY

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

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences, Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2024:87
ISBN 978-91-8021-582-4
ISSN 1652-8220

Metabolic and angiogenic biomarkers in breast cancer: potential clinical implications of host–tumor interactions

Metabolic and angiogenic biomarkers in breast cancer

Potential clinical implications of host–tumor interactions

Christopher Godina, MD

LUND UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on the 18th of June 2024 at 13.00 in Torsten Landberg Hall, the Radiotherapy building, Klinikgatan 5, Lund

Faculty opponent
Professor Therese Sørlie, PhD

Oslo University and Oslo University Hospital, Norway

Organization: LUND UNIVERSITY

Document name: Doctoral Dissertation **Date of issue:** 2024-06-18

Author(s): Christopher Godina **Sponsoring organization:**

Title and subtitle: Metabolic and angiogenic biomarkers in breast cancer: potential clinical implications of host–tumor interactions

Abstract:

Background: Both caveolin-1 (CAV1) and insulin-like growth factor binding protein 7 (IGFBP7) have been linked to angiogenesis, insulin-like growth factor (IGF) receptor 1 (IGF-1R) signaling, and the tumor microenvironment. However, CAV1 and IGFBP7 have not yet been adequately characterized and investigated as potential prognostic or treatment-predictive biomarkers at the genomic, transcriptomic, and proteomic level in breast cancer.

Methods: CAV1 and IGFBP7 were investigated in two large prospective population-based cohorts: the Breast Cancer and Blood (BC-Blood) cohort and the Sweden Cancerome Analysis Network – Breast Initiative (SCAN-B) cohort, which both comprised early-stage breast cancer patients. Additionally, the roles of both CAV1 and IGFBP7 in breast cancer were explored in various public databases. IGFBP7 was further examined in the Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2 (I-SPY2), an adaptively randomized phase II clinical trial on neoadjuvant therapy for early-stage breast cancer.

Results: In the BC-Blood cohort, the prognostic impact of CAV1 protein expression varied based on its localization, anthropometric factors, and tumor characteristics. Notably, CAV1 protein expression in malignant cells predicted a high incidence of breast cancer events among patients with tumors categorized as low-risk, while also indicating metachronous contralateral disease. Additionally, *CAV1* polymorphisms were linked to an elevated risk of locoregional recurrence and contralateral breast cancer. On the other hand, low protein levels of tumor-specific IGFBP7 suggested a favorable prognosis. However, the prognostic significance of high levels of tumor-specific IGFBP7 depended on host factors and treatment. In the SCAN-B cohort, high *CAV1* gene expression emerged as an independent prognostic factor in triple-negative breast cancer. Moreover, the molecular profile associated with high *CAV1* gene expression implicated a potential role in chemoresistance and fostering a tumor-promoting tumor microenvironment. Similarly, elevated *IGFBP7* gene expression was predictive of poor outcomes in breast cancer and correlated with a tumor-promoting tumor microenvironment. Conversely, low *IGFBP7* gene expression identified a subset of breast cancer patients with a favorable response to ganitumab in the I-SPY2 trial.

Conclusions
Both CAV1 and IGFBP7 have been identified as potential prognostic markers in breast cancer, although their significance may vary depending on the specific context. Notably, CAV1 appears to play a particularly crucial role in triple-negative breast cancer. Furthermore, the expression of the *IGFBP7* gene shows promise in predicting the efficacy of treatment targeting IGF-1R using monoclonal antibodies.

Key words: breast cancer, caveolin-1, insulin-like growth factor binding protein-7

Language: English **ISSN and key title**: 1652-8220

ISBN: 978-91-8021-582-4

Number of pages: 160

Signature Date 2024-05-02

Metabolic and angiogenic biomarkers in breast cancer

Potential clinical implications of host–tumor interactions

Christopher Godina, MD

LUND
UNIVERSITY

Supervisor

Professor Helena Jernström, PhD

Co-supervisors

Associate professor Karolin Isaksson, MD, PhD

Doctor Ana Bosch, MD, PhD

Faculty of Medicine
Department of Clinical Sciences in Lund

ISBN 978-91-8021-582-4
ISSN 1652-8220

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

In memory of my grandmother, Edith

Per aspera ad astra

Table of Contents

Abstract

Background: Both caveolin-1 (CAV1) and insulin-like growth factor bindings protein 7 (IGFBP7) have been linked to angiogenesis, insulin-like growth factor (IGF) receptor 1 (IGF-1R) signaling, and the tumor microenvironment. However, CAV1 and IGFBP7 have not yet been adequately characterized and investigated as potential prognostic or treatment-predictive biomarkers at the genomic, transcriptomic, and proteomic level in breast cancer.
Methods: CAV1 and IGFBP7 were investigated in two large prospective population-based cohorts: the Breast Cancer and Blood (BC-Blood) cohort and the Sweden Cancerome Analysis Network – Breast (SCAN-B) cohort, which both comprised early-stage breast cancer patients. Additionally, the roles of both CAV1 and IGFBP7 in breast cancer were explored in various public databases. IGFBP7 was further examined in the Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2 (I-SPY2), an adaptively randomized phase II clinical trial on neoadjuvant therapy for early-stage breast cancer.
Results: In the BC-Blood cohort, the prognostic impact of CAV1 protein expression varied based on its localization, anthropometric factors, and tumor characteristics. Notably, CAV1 protein expression in malignant cells predicted a high incidence of breast cancer events among patients with tumors categorized as low-risk, while also indicating metachronous contralateral disease. Additionally, *CAV1* polymorphisms were linked to an elevated risk of locoregional recurrence and contralateral breast cancer. On the other hand, low protein levels of tumor-specific IGFBP7 suggested a favorable prognosis. However, the prognostic significance of high levels of tumor-specific IGFBP7 depended on host factors and treatment. In the SCAN-B cohort, high *CAV1* gene expression emerged as an independent prognostic factor in triple-negative breast cancer. Moreover, the molecular profile associated with high *CAV1* gene expression implicated a potential role in chemoresistance and fostering a tumor-promoting tumor microenvironment. Similarly, elevated *IGFBP7* gene expression was predictive of poor outcomes in breast cancer and correlated with a tumor-promoting tumor microenvironment. Conversely, low *IGFBP7* gene expression identified a subset of breast cancer patients with a favorable response to ganitumab in the I-SPY2 trial.
Conclusion: Both CAV1 and IGFBP7 have been identified as potential prognostic markers in breast cancer, although their significance may vary depending on the specific context. Notably, CAV1 appears to play a particularly crucial role in triple-negative breast cancer. Furthermore, the messenger ribonucleic acid (mRNA) expression of the *IGFBP7* gene shows promise in predicting the efficacy of treatment targeting IGF-1R using monoclonal antibodies.

Populärvetenskaplig sammanfattning

Bröstcancer är den vanligaste cancerformen bland kvinnor i Sverige och världen över. Allt fler kvinnor insjuknar i bröstcancer och behandlingen har blivit alltmer komplex med flera olika behandlingsalternativ som är tillgängliga. Det är därför avgörande att varje patient får rätt behandling för att minska risken för återfall samtidigt som biverkningarna av behandlingen minimeras. Detta har lett till idén precision-medicin, som syftar till att erbjuda bästa möjliga behandling till varje patient. Trots framsteg och intensiv forskning inom området har detta mål ännu inte uppnåtts inom bröstcancer. För att ytterligare förbättra behandlingen behövs nya så kallade tumörmarkörer som kan förutspå prognos och/eller behandlingssvar. Inom cancerforskning används tumörmarkörer ofta för att återspegla tumörens egenskaper, även om samspel mellan tumören och dess värd inte har utforskats väl. Denna avhandling fokuserade på två tumörmarkörer, Caveolin-1 och IGFBP7, båda relaterade till ämnesomsättning och kärlbildning i kroppen. Både Caveolin-1 och IGFBP7 undersöktes i flera populationer av bröstcancerpatienter för att fastställa deras potential att förutspå prognos eller behandlingssvar. IGFBP7 undersöktes också i en randomiserad studie av bröstcancer med hög risk för återfall. Caveolin-1 och IGFBP7 i tumören undersöktes både gällande nivåer av protein och så kallat mRNA (som kodar för proteiner). Även normalvarianter i Caveolin-1-genen undersöktes. Både Caveolin-1 och IGFBP7 var relaterade till liknande tumöregenskaper som främjar tumörcellernas förmåga att sprida sig i kroppen och dessa egenskaper visar även en framträdande roll i tumörcellernas omgivande miljö. Höga proteinnivåer av Caveolin-1 var kopplat till ökad risk för återfall i bröstet, området runt bröstet och det andra "friska" bröstet, beroende på vilka celler i tumören som hade höga nivåer av Caveolin-1. Dessutom var vissa normalvarianter i Caveolin-1-genen kopplade till ökad risk för återfall i bröstet, området runt bröstet och det andra "friska" bröstet. I en specifik typ av bröstcancer, kallad trippelnegativ bröstcancer, var höga mRNA nivåer av Caveolin-1 kopplat till ökad risk för återfall. Höga nivåer av IGFBP7, både mRNA och protein, var också kopplat till ökad risk för återfall. Dessutom kunde IGFBP7 förutspå om patienter svarar på ganitumab, ett läkemedel som riktar sig mot receptorn för tillväxtfaktorn IGF-1 som bidrar till ökad tumörtillväxt och spridning. Dessa fynd ger insikt om två nya potentiella tumörmarkörer som förtjänar vidare forskning och bekräftelse av fynden i framtida studier.

List of Papers

Paper I

Godina C, Indira Chandran V, Barbachowska M, Tryggvadottir H, Nodin B, Visse E, Borgquist S, Jirström K, Isaksson K, Bosch A, Belting M, Jernström H. Interplay between Caveolin-1 and body and tumor size affects clinical outcomes in breast cancer. Translational Oncology. 2022;22:101464

Paper II

Godina C, Tryggvadottir H, Bosch A, Borgquist S, Belting M, Isaksson K, Jernström H. Caveolin-1 genotypes as predictor for locoregional recurrence and contralateral disease in breast cancer. Breast Cancer Research and Treatment. 2023;199(2):335-347

Paper III

Godina C, Khazaei S, Belting M, Vallon-Christersson J, Nodin B, Jirström K, Isaksson K, Bosch A, Jernström H. High caveolin-1 expression in triple-negative breast cancer is associated with an aggressive tumor microenvironment, chemoresistance and poor clinical outcome. *Manuscript in revision*.

Paper IV

Godina C, Khazaei S, Tryggvadottir H, Visse E, Nodin B, Jirström K, Borgquist S, Bosch A, Isaksson K, Jernström H. Prognostic impact of tumor-specific insulin-like growth factor binding protein 7 (IGFBP7) levels in breast cancer: a prospective cohort study. Carcinogenesis. 2021;42(11):1314-1325

Paper V

Godina C, Pollak MN, Jernström H. Low IGFBP7 expression identifies a subset of breast cancers with favorable prognosis and sensitivity to IGF-1 receptor targeting with ganitumab: Data from I-SPY2 and SCAN-B. *Manuscript in revision*.

Author's contribution to the papers

Paper I

Christopher Godina is the sole first author. I contributed to the conceptualization, methodology, statistical analysis, investigation, and writing of the manuscript. I was involved in all stages of the publication process.

Paper II

Christopher Godina is the sole first author. I contributed to the conceptualization, methodology, statistical analysis, investigation, and writing of the manuscript. I was involved in all stages of the publication process.

Paper III

Christopher Godina is the sole first author. I contributed to the conceptualization, methodology, data analysis, investigation, and writing of the manuscript. I will take the main responsibility in all stages of the publication process.

Paper IV

Christopher Godina is the sole first author. I contributed to the conceptualization, methodology, statistical analysis, investigation, and writing of the manuscript. I was involved in all stages of the publication process.

Paper V

Christopher Godina is the sole first author. I contributed to the conceptualization, methodology, data analysis, investigation, and writing of the manuscript. I will take the main responsibility in all stages of the publication process.

Other published papers not included in this thesis

Godina, C., Belting, M., Vallon-Christersson, J., Isaksson, K., Bosch, A., & Jernström, H. (2024). Caveolin-1 gene expression provides additional prognostic information combined with PAM50 risk of recurrence (ROR) score in breast cancer. Scientific reports, 14(1), 6675. https://doi.org/10.1038/s41598-024-57365-8

Lindgren, H., Ademi, D., Godina, C., Tryggvadottir, H., Isaksson, K., Jernström, H. (2024). Potential interplay between tumor size and vitamin D receptor (VDR) polymorphisms in breast cancer prognosis: a prospective cohort study. Cancer causes & control, 10.1007/s10552-023-01845-1. Online head of print. https://doi.org/10.1007/s10552-023-01845-1

Jujić, A., Godina, C., Belting, M., Melander, O., Juul Holst, J., Ahlqvist, E., Gomez, M. F., Nilsson, P. M., Jernström, H., & Magnusson, M. (2023). Endogenous incretin levels and risk of first incident cancer: a prospective cohort study. Scientific reports, 13(1), 382. https://doi.org/10.1038/s41598-023-27509-3

Godina, C., Ottander, E., Tryggvadottir, H., Borgquist, S., Isaksson, K., & Jernström, H. (2020). Prognostic Impact of Menopausal Hormone Therapy in Breast Cancer Differs According to Tumor Characteristics and Treatment. Frontiers in oncology, 10, 80. https://doi.org/10.3389/fonc.2020.00080

Thesis at a glance

Paper	Research question	Methods	Results
I Levels of CAV1 protein expression	What is the role of CAV1 in different cellular localizations of breast cancer in relation to clinical outcomes?	In the BC-Blood cohort, TMAs were stained with a polyclonal CAV1 antibody. CAV1 was investigated in relation to prognosis.	CAV1 in malignant cells predicted high incidence of breast cancer events in a group of patients with small body size and tumors. CAV1 in malignant cells predicted contralateral disease.
II *CAV1* polymorphisms	Do *CAV1* genotypes and haplotypes impact prognosis in breast cancer?	In the BC-blood cohort, *CAV1* genotypes and haplotypes from OncoArray were investigated in relation to prognosis.	*CAV1* polymorphisms were associated with an increased incidence for locoregional recurrence and contralateral breast cancer.
III CAV1 in TNBC	Are there associations between gene and protein expression of CAV1 in TNBC and molecular features, tumor microenvironment composition, and clinical outcome?	In the SCAN-B cohort, TMAs were stained with a polyclonal CAV1 antibody, and *CAV1* gene expression from RNA-seq was used. CAV1 was investigated in relation to prognosis.	High *CAV1* gene expression was an independent prognostic factor in TNBC with molecular features suggesting a role in chemoresistance and a tumor-promoting TME.
IV Levels of IGFBP7 protein expression	Are IGFBP7 protein levels associated with the patient and tumor characteristics and prognosis in breast cancer?	Within the BC-blood cohort, TMAs were stained with a polyclonal CAV1 antibody. IGFBP7 was investigated in relation to prognosis.	Low levels of tumor-specific IGFBP7 were a potential marker of good prognosis. The association between high levels of tumor-specific IGFBP7 and prognosis is dependent on host factors and treatment.
V *IGFBP7* gene expression as a predictor for IGF-1R targeting agents	Does IGFBP7 compete with IGF-1R monoclonal antibody binding to IGF-1R, decreasing its efficacy and at the same time promoting tumor growth and metastatic potential?	In both I-SPY2 and SCAN-B gene expression profiling of tumors was performed, and *IGFBP7* mRNA expression was obtained. *IGFBP7* expression was investigated in relation to pCR and prognosis.	A subset of breast cancer patients that have a good response to ganitumab can be identified by low *IGFBP7* gene expression. High *IGFBP7* gene expression was predictive of poor outcome in breast cancer.

Abbreviations

ACAT1	Acetyl coenzyme A cholesterol acyltransferase 1
AKT	Protein kinase B
AI	Aromatase inhibitor
AJCC	American Joint Committee on Cancer
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BC-Blood	Breast cancer and blood
BL	Basal-like
BMI	Body mass index
BSA	Body surface area
c	Clinical
CAF	Cancer-associated fibroblast
CAV1	Caveolin-1
CBC	Contralateral breast cancer
CDK4/6	Cyclin-dependent kinase 4 and 6
CE	Carcinoma ecotype
CI	Confidence interval
CIF	Cumulative incidence function
DC	Dendritic cells
DGE	Differential gene expression
DNA	Deoxyribonucleic acid
DPD	Dihydropyrimidine dehydrogenase
DSB	Double-stranded break
GEO	Gene Expression Omnibus
GnRH	Gonadotropin-releasing hormone
GO	Gene ontology
GOBO	Gene Expression-based Outcome for Breast Cancer Online
GSEA	Gene set enrichment analysis

GWAS	Genome-wide association study
H&E	Hematoxylin and eosin
HWE	Hardy-Weinberg equilibrium
E1	Estrone
E2	Estradiol
E3	Estriol
ECM	Extra-cellular matrix
EDR	Endocrine-sensitive disease rate
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor
ERα	Estrogen receptor alpha
ERβ	Estrogen receptor beta
FDR	False discovery rate
FFPE	Formalin-fixed paraffin-embedded
FPKM	Fragments per kilobase of exon model per million mapped reads
GLUT3	Glucose transporter 3
HER2	Human epidermal growth factor 2
HR	Hazard ratio
ICI	Immune checkpoint inhibitor
IHC	Immunohistochemistry
IGF	Insulin-like growth factor
IGF-1R	Insulin-like growth factor receptor 1
IGFBP	Insulin like growth factor binding protein
IGFBP7	Insulin-like growth factor binding protein 7
IM	Immunomodulatory
ISH	In situ hybridization
I-SPY2	Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2
InsR	Insulin receptor
IRS1	Insulin receptor substrate 1

KEGG	Kyoto Encyclopedia of Genes and Genomes
Ki67	Antigen Kiel 67
KM	Kaplan-Meier
LAR	Luminal androgen receptor
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LumHR	Luminal hormone-responsive
LumSec	Luminal secretory
M	Mesenchymal
MAPK	Mitogen-activated protein kinase
MAR	Missing at random
MCAR	Missing completely at random
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MICE	Multiple imputation by chained equations
MNAR	Missing not at random
mRNA	Messenger ribonucleic acid
MSL	Mesenchymal stem-like
MP	MammaPrint®
NK	Natural killer
NKBC	Swedish National Breast Cancer Quality Register
NST	Invasive carcinoma of no special type
OR	Odds ratio
p	Pathological
PAM50	Prediction analysis of microarray 50
PARP	Poly (adenosine diphosphate–ribose) polymerase
pCR	Pathological complete response
PD-1	Programmed death 1
PD-L1	Programmed death ligand 1
PI3K	Phosphatidylinositol 3-kinase

PR	Progesterone receptor
RAF	Rapidly accelerated fibrosarcoma
RCT	Randomized clinical trial
RERI	Relative risk to due interaction
RIN	RNA integrity number
RNA	Ribonucleic acid
RNA-seq	Massive parallel sequencing of RNA
ROR	Risk of recurrence
ROS	Reactive oxygen species
SCAN-B	Swedish Cancerome Analysis Network - Breast
SERM	Selective estrogen receptor modulator
SGLT2	Sodium-glucose co-transporter-2
SNP	Single nucleotide polymorphism
STAT3	Signal transducer and activator of transcription 3
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
T-DM1	Trastuzumab-emtansine
TAM	Tumor-associated macrophage
TCGA	The cancer genome atlas
TET	Ten-eleven translocation
TGF- β	Transforming growth factor beta
TMA	Tissue microarray
TME	Tumor microenvironment
TNM	Tumor Node Metastasis
TPM	Transcripts per million reads
WC	Waist circumference
WHO	World Health Organization
WHR	Waist-to-hip ratio
yp	Yield pathological

Introduction

"Research is to see what everybody else has seen, and to think what nobody else has thought"

—Albert Szent-Györgyi

Cancer predates humans (*Homo sapiens*), with the oldest documented cancer dating back approximately 1.7 million years ago in South Africa (1). It was an osteosarcoma found in the foot of a hominin, a human ancestor (1). This suggests that malignant neoplasms have been a part of human life for as long as our species has existed. The susceptibility of humans to develop cancer has persisted, and the historical lethality of cancer has endured, most likely because humans typically develop cancer after childbearing age, thereby minimizing the influence of natural selection.

The earliest recorded instance of cancer in a human can be traced back to approximately 4,000 BC, with evidence suggesting that a Scythian king likely had disseminated prostate cancer (2, 3). The first account of breast cancer is found in the Edwin Smith Surgical Papyrus dating back to 1,600 BC, although the papyrus is likely a copy of even older texts from around 3,000 BC (2, 3). It was not until 400 BC that Hippocrates coined the term "carcinoma", which is derived from the Greek word for crab, "karkinos" (2, 3).

What is highly fascinating is the treatment principles from ancient times, such as having a wide margin of excision and only removing tumors of limited extent (2, 3). It was also understood that cancer is a systemic disease and may not always be wise to treat (2, 3). It was even postulated that the cessation of menstruation was somehow linked to cancer (2, 3). These ancient practices mirror modern-day oncological clinical practice and in some respects are not too far from the modern understanding of cancer.

Today, cancer remains a formidable and ongoing challenge to human health and profoundly impacts individuals, families, and societies worldwide. It is increasing globally (4). In Sweden, it is estimated that at least one-third of the population will be diagnosed with cancer (5). Thus, there is an urgent need to understand and treat it better. Due to its multifaceted nature, cancer can affect virtually any organ or tissue in the human body, which adds another layer of difficulty.

This thesis focuses on breast cancer, the second most common cancer worldwide. For many patients, breast cancer is now a curable disease due to advancements in diagnostics, screening, and treatment (6-10). However, breast cancer is the leading cause of death from cancer among women worldwide, and a substantial portion of patients relapse (4, 6, 7). Sadly, those who do relapse often face poor outcomes (4, 6, 7).

Cancer and tumorigenesis

The word “cancer” was initially used to describe the pattern of varicose, swollen blood vessels, and hardness reminiscent of crustaceans observed in certain cancers that have grown large (2). The original principle was that “hidden” cancer—i.e., cancer without breakthrough to the skin—should not be treated as the benefits would outweigh the harms (2, 3). Stemming from Latin, the term “tumor” means “swelling” or “lump” and refers to a mass of abnormally growing cells (3). A crucial distinction is that tumors can be benign, lacking the capacity to damage or invade surrounding tissues, and are not considered cancer. However, tumors can become malignant when the cells are more aggressive, damage surrounding tissue, and can invade and metastasize, which is referred to as cancer (11).

Cancer can be categorized based on the cell types of origin as four major types. Carcinomas originate from epithelial cells of the internal or external lining of the body and account for 80–90% of all cancer cases (11). Breast cancer is a type of carcinoma (11). Other types include sarcoma, which originates from non-hematopoietic mesenchymal cells in supportive and connective tissues such as bones; leukemia, which originates from hematopoietic cells maturing blood; and lymphomas, which originate from hematopoietic cells maturing in the lymphatic system (11). Some cancers do not fit into this classification and remain unclassified, such as melanoma of the skin and glioblastoma (11).

According to current understanding, cancer is thought to develop from the accumulation of cellular and genomic damage over generations of cell division (12-14). This damage gradually disables and alters intrinsic control mechanisms, enabling cells to acquire tumorigenic properties and transform into cancer cells (15). As cancer cells proliferate and change the surrounding microenvironment through paracrine signaling, the tumor becomes a micro-ecosystem where aggressive and proliferative properties undergo positive selection. This process spans several years, with an estimated average of over 20 years for a tumor to evolve from a single cell to a clinically detectable state, implying prolonged periods of cellular and genomic damage (12, 14). Many precursor lesions succumb to cellular and genomic damage, undergoing apoptosis or being cleared away by the immune system before developing into cancer, while those that remain are potentially dangerous.

Eventually, when cancer is sufficiently developed, it will likely cause morbidity and mortality (15, 16). Several mechanisms are at play. Locally, cancer can harm the integrity and function of the tissue or organ of origin through inflammation, tissue destruction, and hijacking normal cell function via paracrine signaling (15, 16). Due to altered metabolism, tumor cells can consume nutrients to such an extent that the host is starved, which is a condition called cachexia (17). Tumors can disrupt the homeostasis of the entire body by releasing different signaling molecules into the bloodstream, including inflammatory factors and hormones (18). As cancer spreads throughout the body, the tumor burden increases, worsening symptoms, and affecting additional tissues, and potentially leading to multiple organ failure and eventual death (18, 19).

Hallmarks of cancer

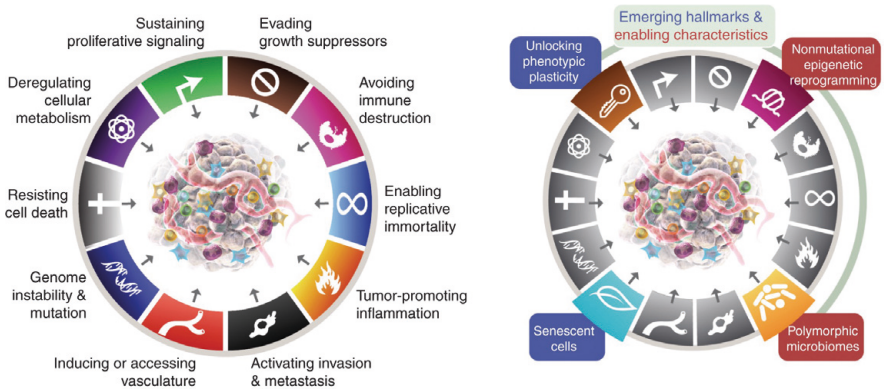


Figure 1
The hallmarks of cancer according to the latest update in 2022 (20). © 2022 American Association for Cancer Research. Reprinted with permission from the American Association for Cancer Research.

In 2000, Douglas Hanahan and Robert A. Weinberg authored a seminal paper in cancer biology titled, "The Hallmarks of Cancer", which presented a conceptual model of cancer and tumorigenesis (15). This comprehensive model describes six critical functional capabilities of cancer that are needed for the malignant transformation of normal cells (15). Cancer must achieve self-sufficiency in growth signals, meaning it must generate the necessary signals for its own unbridled growth (15). Additionally, it must develop reduced sensitivity to growth-inhibitory signals, enabling it to resist external and internal factors that would otherwise impede its proliferation (15). Furthermore, cancer cells must possess the ability to avoid programmed cell death, or apoptosis, ensuring their survival even when the normal cellular mechanisms would dictate otherwise (15, 21, 22). Another essential capability outlined in the paper is the acquisition of unlimited replicative potential (15). Cancer cells must overcome the usual constraints on cell division, allowing

them to proliferate indefinitely (23). As a tumor expands in size, it becomes crucial for cancer cells to induce vascular supply through angiogenesis (15, 24). This process ensures an adequate blood supply to sustain the growing tumor (15, 24). Equally significant is the capability of malignant tumors to invade surrounding tissues (15). Cancer cells acquire invasive properties, allowing them to breach the boundaries of their original site and infiltrate neighboring tissues (18, 19). Moreover, these cells must gain the capacity to spread to distant sites within the body via the bloodstream or lymphatic system, facilitating metastasis (18, 19, 25, 26). The acquisition of these functional capabilities happens in parallel through a clonal process and not in a step-wise manner (15).

Hanahan and Weinberg have updated their model twice to incorporate new information from an enhanced understanding of cancer. The initial update in 2011 introduced two novel emerging hallmarks and two "enabling characteristics" that are crucial for cells to acquire hallmarks (16). The two additional hallmarks involve reprogramming energy metabolism and evading the immune system (16). Cancer cells must adapt their metabolism to sustain chronic and often uncontrolled cell proliferation, which was initially described by Otto Warburg (27, 28). In this context, instead of oxygen-dependent oxidative phosphorylation in the mitochondria, the energy metabolism of cancer cells primarily relies on glycolysis, even in the presence of oxygen, which is called "aerobic glycolysis" (27). Evading the immune system is a critical aspect as the it acts as a significant barrier to tumor formation by recognizing and eliminating abnormal cells (16). However, cancers that do form often escape immune surveillance through various mechanisms, including immunosuppression by regulatory immune cells, defective antigen presentation, and the action of immune suppressive mediators (16, 29).

The enabling characteristics encompass genomic instability and mutation, along with tumor-promoting inflammation (16). Genomic instability and an increased mutational rate in cancer enhance the likelihood of acquiring mutations, epigenetic changes, or genomic alterations in key oncogenes and tumor suppressors (16). This contributes to the rapid development of subpopulations of cancer cells with increased fitness and additional hallmarks necessary for the transformation into malignant tumors (30, 31). Tumor-promoting inflammation caused by innate immune cells in the tumor facilitates angiogenesis, invasion, and metastasis (16). Inflamed immune cells secrete growth factors, proangiogenic factors, and extracellular matrix (ECM)-modifying enzymes, which promote the tumor's progression (16, 32, 33). Paradoxically, the immune system's inflammatory response, which is typically employed by the body to combat pathogens and aid in tissue repair, inadvertently supports tumor development by providing the necessary conditions for tumor-promoting capabilities in normal cells (16, 32, 33).

The most recent update in 2022 by Douglas Hanahan introduced two novel emerging hallmarks and two "enabling characteristics": unlocking phenotypic

plasticity and understanding senescent cells (20). Phenotypic plasticity refers to the remarkable ability of cancer cells to alter their phenotype, denoting a shift in tissue lineage (20, 34). Typically, fully differentiated cells exhibit sustained anti-proliferative signaling; however, many cancer cells subvert this by altering their transcriptional program towards another cell fate or an earlier progenitor cell (34-36). This adaptation enables cancer cells to acquire tissue-specific traits that were not predetermined by their normal cells of origin (20, 35). Intriguingly, cellular senescence, which has traditionally been seen as a protective mechanism against neoplasia, can be the opposite. Senescent cells can promote tumor phenotypes through paracrine signaling, conferring hallmark capabilities to viable cancer cells (37, 38). Studies have demonstrated that senescent cells contribute to proliferative signaling, evading apoptosis, angiogenesis, invasion, metastasis, and suppressing tumor immunity (37, 38).

The newly proposed enabling characteristics are non-mutational epigenetic reprogramming and the polymorphic microbiome (20). In contrast to instability and mutation of genomic deoxyribonucleic acid (DNA), non-mutational epigenetic reprogramming refers to epigenetically regulated changes in gene expression independent of genomic alterations (20). Emerging evidence suggests that not all acquired capabilities of cancer cells stem from genomic alterations, as the microenvironment can induce epigenetic changes, driving phenotypic selection of hallmark capabilities (20, 39). For instance, hypoxia reduces the activity of ten-eleven translocation (TET) demethylases, leading to substantive changes in the methylome, particularly hypermethylation (40). Additionally, epigenetic changes may account for the dynamic transcriptomic heterogeneity increasingly documented in cells within malignant tumor microenvironments (39, 41). The concept of the polymorphic microbiome emphasizes the impact of ecosystems created by resident bacteria and fungi on cancer phenotype (20). It has been proposed that specific bacterial species can directly stimulate proliferative signaling in the colonic epithelium and modulate growth suppression by altering tumor suppressor activity (42-44).

Cancer is beyond a mere accumulation of rapidly dividing cells; it is a complex tissue emerging from several concordant processes required for tumor promotion (15, 16, 20). A cancer tumor represents a dynamic ecosystem where cancer cells interact with normal cells, fostering an environment conducive to its development and survival (15, 16, 20). However, this framework of cancer biology fails to comprehensively account for the extensive interactions between an advancing tumor and remote organs within the host, as well as the influence of host pathophysiology, germline genetic variations, and environmental exposures on cancer initiation and progression (45). Recently, Swanton and colleagues defined the concept of "hallmarks of systemic disease". These were meant to describe and encourage the exploration of cancer as a systemic disease resulting from the intricate interplay

among host genome diversity and a history of human behavior leading to multifaceted environmental exposures (45). One of the main focuses of this thesis is to elucidate this interplay.

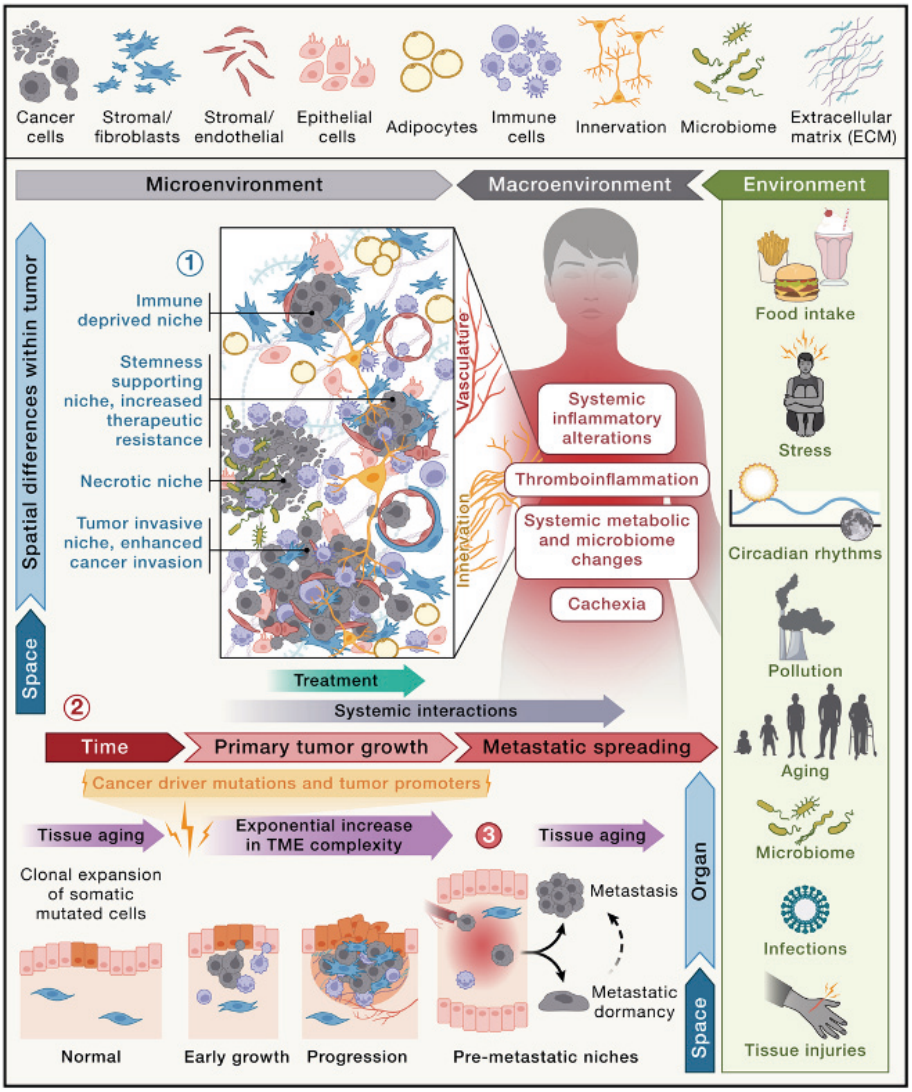


Figure 2.
Hallmarks of systemic disease as illustrated in Swanton et al. (45). © 2024 Elsevier. Reprinted with permission from Elsevier.

Invasion and metastasis

The hallmark of cancer, responsible for the majority of cancer-related deaths, is invasion and metastasis, with over 90% of cancer deaths attributed to metastasis to organs distant from the primary site (19, 25). Metastasis is the formation of secondary tumors in distant body parts and is a crucial distinction between benign and malignant tumors in clinical practice (19, 26). A notable aspect of metastasis is the daily release of large number of cancer cells into the circulation and lymphatic system in patients with cancer. However, only a tiny fraction (< 0.1%) of these released cells successfully metastasize to other parts of the body (46), which underscores the arduous that journey cancer cells must undergo in order to form a distant metastasis. The process involves intricate steps, including intravasation, circulation, extravasation, and colonization, each presenting challenges for cancer cells seeking to establish a secondary tumor in a distant organ (19, 26). Adding complexity in the metastatic process, it has recently been demonstrated that normal epithelial cells can "metastasize" without being cancerous, uncoupling metastasis from tumorigenesis (47, 48).

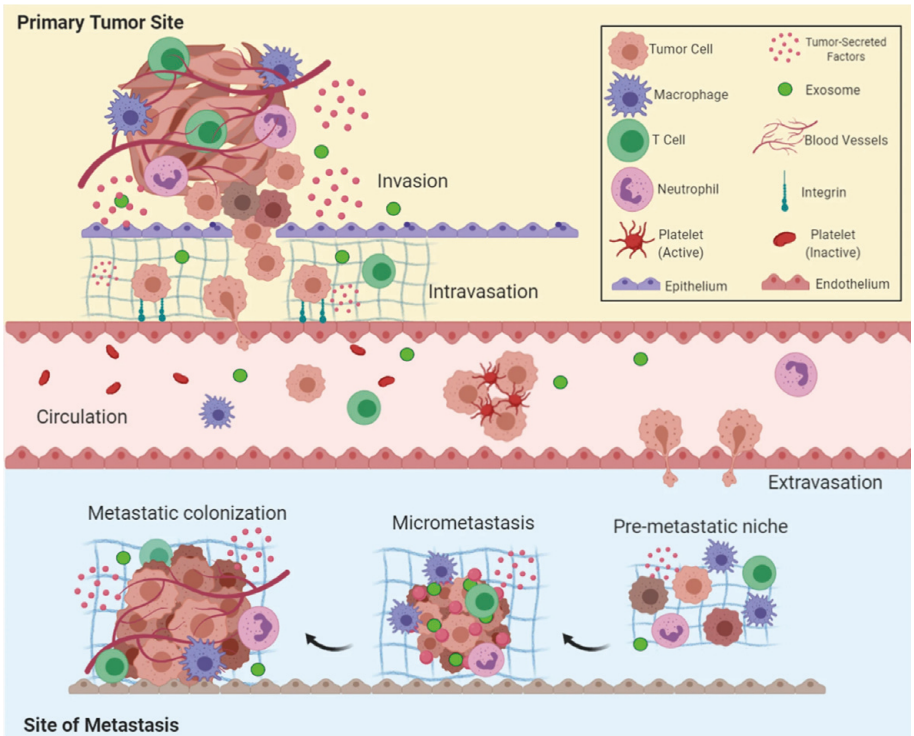


Figure 3.
The principles stages of metastasis from Fares et al. (19). It illustrates the main steps of intravasation, circulation, and extravasation. © 2020 Fares et al. Open access.

Intravasation is the first step and requires cancer cells to leave their site of origin and invade the bloodstream (16, 19). For this to happen, the cancer cells must be primed and usually undergo epithelial-mesenchymal transition (EMT) (19, 26, 49). Under EMT, cancer cells develop the ability to invade, resist stress, and disseminate of the immobile epithelial cancer cells tightly bound to each other and to the neighboring ECM (19, 26, 49). The ECM is also remodeled to support their growth and enhance their metastatic capability, enabling them to invade into the bloodstream (19, 26). Subsequently, cancer cells circulate in the bloodstream, facing harsh conditions, and must attach and invade a new site (26). Cancer cells can circulate as single cells or travel in clusters. These circulating clusters are much more likely to form metastases (50). The circulating clusters contain stromal cells and immune components from the original microenvironment that contribute to survival in the circulation (50, 51). The attachment of platelets to these clusters leads to the formation of a coating shield around cancer cells that prevents detection by immune cells and provides the structure needed to bear the physical stresses of circulation (52). When cancer cells passes through small capillaries, they become entrapped (19). This forces the cells to undergo extravasation (26). Extravasation is a complex process that involves ligand–receptor interactions, chemokines, and circulating nontumor cells. Generally, cancer cells induce programmed necrosis of endothelial cells that causes the metastatic cancer cells to extravasate (19, 26).

Certain organs with highly permeable sinusoidal vessels, such as the liver and bone, show a higher rate of metastasis (19). The concept of organotropism, pertaining as to why different cancer types have varying preferences of spreading to distant metastasis, dates back to the 19th century, when Paget concluded that metastasis to distant organs is not a random process (53). Paget proposed the “seed and soil” hypothesis for metastatic dissemination (53). In fact, Paget was actually quite right that primary tumors can prime the host microenvironment of distant organs, creating a premetastatic niche even before the initiation of metastasis (54). The development of a premetastatic niche is a multistep process involving secretory factors and extracellular vesicles that induce vascular leakage, ECM remodeling, and immunosuppression (54).

Initially, upon extravasation at the target site, metastatic cancer cells confront adverse conditions that pose challenges to their survival (19, 26). The subsequent colonization phase requires the cancer cells to adapt and eventually integrate into their new surroundings, creating a supportive microenvironment akin to establishing a vascular network and ecosystem conducive to tumorigenesis (19). The intricacies of the colonization process are complex and remain inadequately understood, involving interactions between cancer cells and host cells mediated by secreted tumor-derived factors and exosomes (26).

Another facet of colonization that has gathered substantial attention from the research community is the phenomenon of cancer dormancy. Cancer dormancy represents a phase of arrested progression during primary tumor formation or

following invasion into secondary sites (19, 25, 26). Specifically, metastatic dormancy arises from the delayed adaptation of disseminating cancer cells to their secondary niches, impacting both individual invading cells and cancer clusters post-circulation (19, 25, 26). In numerous cancer survivors, dormant cancer cells persist long after the radical removal of the primary tumor and adjuvant treatment been administered, posing significant challenges for treatment, and potentially contributing to delayed relapses (26, 55). Dormancy encompasses different states, including quiescence, angiogenic dormancy characterized by a balance between dividing and dying (vasculature-lacking) cancer cells, and immune-mediated dormancy, where immune-cell cytotoxicity preserves the tumor mass (19, 25, 26).

Tumor heterogeneity and evolution

The complex heterogeneity of cancer presents a formidable challenge for precision medicine, and the often unpredictable initiation of cancer underscores the dynamic and multifaceted nature of its development and progression (12, 13). Unlike a predetermined course, cancer represents an intricate destabilization of crucial cellular processes and continually evolves even after malignant transformation. This ongoing evolution increases intratumoral heterogeneity bulk tumors are molecularly diverse with distinct cellular ecosystems exhibiting spatial and temporal variations in genetic and molecular makeup (12, 13).

Intratumoral heterogeneity observed across diverse types of cancer is crucial in driving cancer evolution and fostering drug resistance (56). Diverse tumor cell populations within a single tumor increase the likelihood of some cells surviving therapy, while ongoing diversification during treatment enables adaptation to selective pressures, leading to *de novo* resistance and eventual relapse (56, 57). Conversely, intertumoral heterogeneity reflects differences between patients with tumors of the same histological type and arises from patient-specific factors like germline genetic variations, somatic mutation profiles, and environmental influences (56, 57). These subtle differences impact treatment outcomes and complicate the search for optimal treatments. In diseases like breast cancer, where survival outcomes vary, identifying subgroups of patients who respond and those who do not beyond the current understanding remains a challenging goal (56-58). Efforts to uncover the underlying factors determining treatment response require exploring multiple dimensions of the tumor, including proteomic, transcriptomic, epigenomic, genomic, and spatial dimensions (56-58).

Breast cancer is a success story in terms of addressing intertumoral heterogeneity, with landmark studies by Perou and Sørlie revealing the clinical implications of transcriptomic classification (59, 60). However, the complexity of intertumoral heterogeneity remains a formidable challenge demanding comprehensive information and powerful tools for relevant tumor classification. Adding to the complexity is the often underappreciated aspect of host heterogeneity, which has

not been as compressively profiled as tumor heterogeneity (57). The pursuit of understanding cancer's intricate heterogeneity underscores the evolving landscape of cancer research, and the aim is to enhance our ability to tailor treatments to the unique characteristics of each patient's tumor (58). Additionally, investigating host–tumor interactions offers further opportunities to refine treatment strategies for cancer patients (57).

Tumor microenvironment

The current understanding acknowledges that cancer is more than just a genetic disorder; it is an intricate ecosystem involving a diverse array of non-cancerous cells and their intricate interactions within the tumor (61, 62). While genetic alterations are crucial, they alone are not adequate for the initiation and progression of cancer (61-66). Cancers are intricate, organized ecosystems that consist of tumor cells and a myriad of non-cancerous cells collectively known as the tumor microenvironment (TME), which occurs within the setting of an altered, vascularized ECM (61-63).

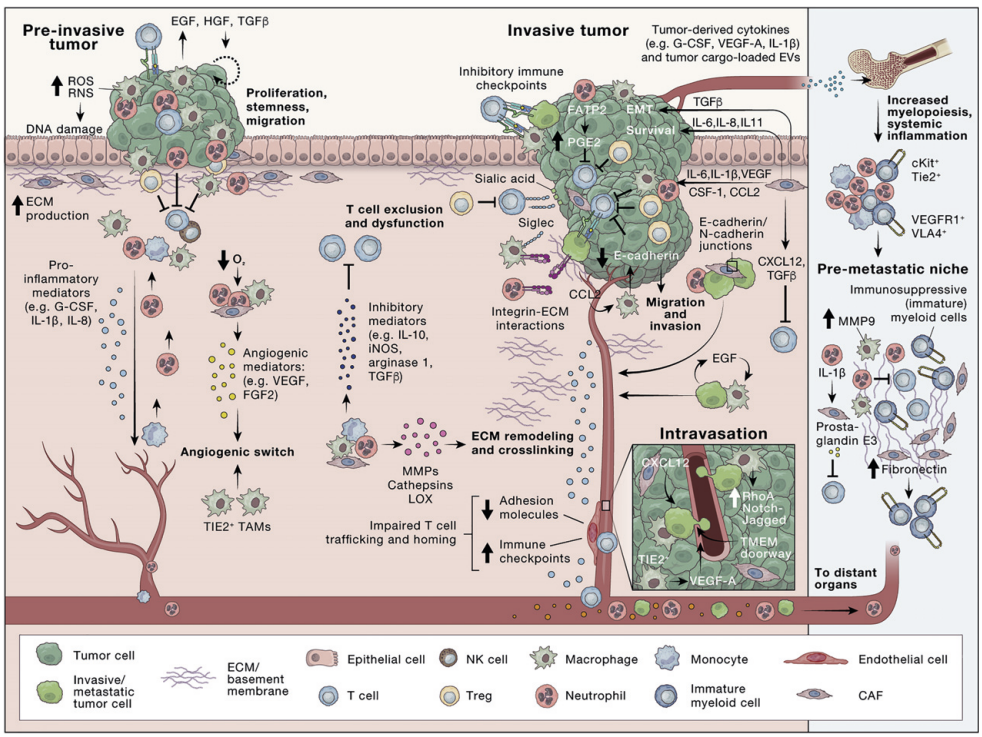


Figure 4.
The complex crosstalk between the malignant cells of the tumor and their microenvironment as illustrated by de Visser et al (61). © 2023 Elsevier. Reprinted with permission from Elsevier.

The TME is characterized by a diverse array of immune cells (61, 62), including T and B lymphocytes, tumor-associated macrophages (TAMs), neutrophils, dendritic cells (DCs), and natural killer (NK) cells (61, 62). Additionally, stromal components such as cancer-associated fibroblasts (CAFs), pericytes, and mesenchymal stromal cells form a part of the TME (61). Other cell types, including adipocytes and neurons, may vary depending on the tissue context (62). The TME also encompasses the ECM and various secreted molecules, such as growth factors, cytokines, chemokines, and extracellular vesicles (62). The intricate network includes the blood and lymphatic vascular systems, which establish dynamic communication with each other and with the heterogeneous population of cancer cells (61, 62).

Initially perceived as bystanders in tumorigenesis, host cells within the TME are now recognized as playing pivotal roles in the pathogenesis of cancer (61, 63). The cellular composition and functional state of the TME vary depending on factors such as the organ of tumor's origin, intrinsic characteristics of cancer cells, tumor stage, and patient-specific attributes (61). Notably, various cells within the TME can exhibit either tumor-suppressive or tumor-supportive functions (61). The regular tissue microenvironment has the capacity to restrain the unchecked growth of cancer by leveraging the suppressive actions of immune cells, fibroblasts, and the ECM (61, 63). For a cancer to progress, it must elude these constraints and induce non-cancerous cells within the TME to adopt a tumor-promoting phenotype, leading to heightened proliferation, invasion, and intravasation at the primary site (61, 63). Cancer cells actively shape a supportive environment by recruiting and reprogramming host cells, and orchestrating vasculature and ECM alterations (61, 63).

The constitution and functional condition of the TME can significantly differ among individuals, even within the same cancer type, which is influenced by patient-specific factors such as age, gender, lifestyle, and the microbiome (61). Additionally, the organ in which the tumor originates contributes to TME variation (61). Most importantly, the cancer cell itself emerges as a pivotal regulator of the TME, as shown by distinct immune landscapes in gliomas, which originate in the brain, compared to brain metastases from extracranial tumors (61, 67). It is increasingly evident that intrinsic features within cancer cells encompassing altered (epi)genetics, metabolic reprogramming, and dysregulated signaling serve as crucial determinants influencing how tumors shape their microenvironment (61, 62). In recent years, strategies to therapeutically target the TME have emerged as a promising avenue in cancer treatment (62). This focus is driven by the recognition of the critical roles played by the TME in regulating tumor progression and influencing the response to standard-of-care therapies (62).

Breast cancer

Breast cancer is the most common cancer among women globally, with 2.3 million diagnoses reported in 2020, including in most countries outside of sub-Saharan Africa (4). Sweden is part of this trend, with approximately 8,500 new breast cancer patients diagnosed annually, and estimates suggest that 1 in 9 women in Sweden will receive a breast cancer diagnosis during her lifetime (5, 68). As the incidence of breast cancer continues to rise, this ratio may escalate, presenting a substantial health challenge. Over 100,000 women previously diagnosed with breast cancer are living in Sweden, making this diagnosis a public health challenge (5). Globally, breast cancer claims the highest percentage (15.5%) of cancer-related deaths among women, surpassing even lung cancer (4). Disparities in breast cancer care, screening, and the prevalence of various subtypes contribute to these statistics (4).

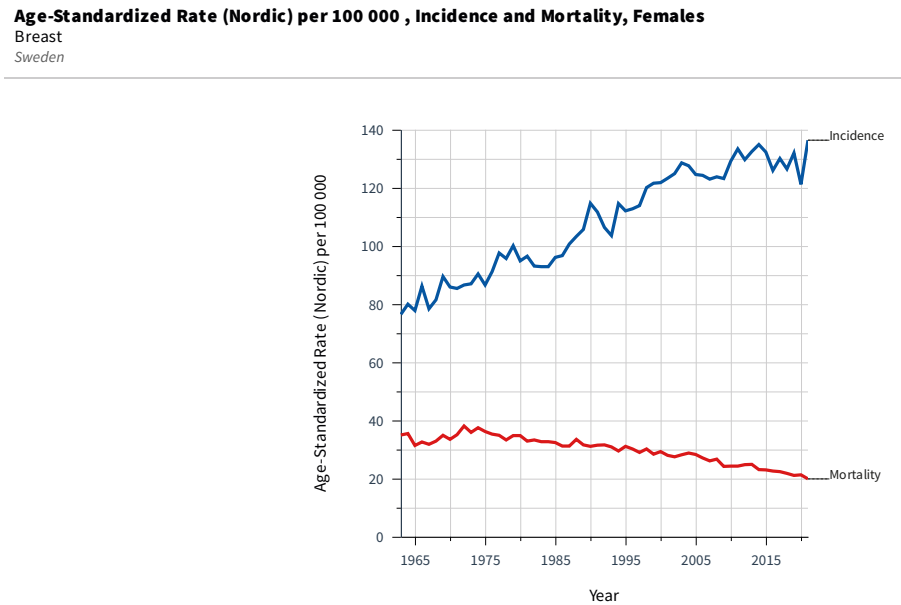


Figure 5.
Age-standardized breast cancer incidence (blue line) and mortality (red line) in Swedish women of all ages, 1952-2021. The rates are presented per 100,000 individuals. © NORDCAN database, provided by the International Agency for Research on Cancer.

Breast cancer mortality has steadily decreased in Sweden and other Western nations over the past decades (69-72), with a notable improvement in the 10-year survival rate from approximately 60% to 86%, but challenges persist. Despite optimal treatment according to current guidelines (68, 73), around 1,300 female breast cancer patients succumb to the disease annually in Sweden alone (5, 68). On a global

scale, an estimated 680,000 women lost their lives to breast cancer in 2020 (4). Challenges include uneven reductions in breast cancer across different stages and age groups (69, 71, 74, 75). Mortality rates remain elevated in subgroups with higher-stage disease and tumors exhibiting aggressive molecular characteristics (69, 70, 74, 75). Additionally, both older patients (age > 70 years) and younger patients (age < 40 years) face relatively high breast cancer mortality (69, 70, 74). Finally, the risk of late relapse poses a formidable challenge, particularly for patients diagnosed with estrogen receptor (ER)-positive breast cancer, in which the risk of relapse plateaus at 15 years and beyond (55). In the intricate landscape of breast cancer, these nuances underscore the ongoing complexity and necessity for comprehensive strategies to enhance prevention, diagnosis, and treatment outcomes.

Normal breast and breast cancer development

The breast is an apocrine organ with a vital physiological role in producing milk for infant nourishment post-birth (76, 77). The breast mainly develops during puberty when the ductal tree of ducts and lobules is formed (78). The formation of the ductal tree is coordinated by a specialized structure known as the terminal end bud (78), (78), which serves as a regulatory control point for various processes, including branching, angiogenesis, and pattern formation within the mammary gland (78). Female sex hormones and growth hormones are key players in controlling the growth of the buds (78).

The human breast is composed of 15–20 milk-producing lobules connected to ducts that serve as conduits for transporting milk to the nipple (76, 77). Different lobules (types one to four) represent various stages of development based on morphology of the breast of post-pubertal women (79). In nulliparous women, lobule type one predominates across all age groups, whereas in parous women, lobule type four is most commonly observed (79). The lobules and the ducts form the epithelial part of the breast. The inner layer consists of luminal cells, and the outer layer comprises basal-myoepithelial cells (76, 77). This epithelial system is integrated into adipose-rich tissue and enveloped by a dense network of vasculature and lymphatic vessels (76, 77). From a histological perspective, the human breast is characterized by four major regions: terminal ductal lobular units with densely packed, branched epithelium; mostly bilayered epithelium in tubular ducts; adipose-rich regions; and ECM-rich connective tissue (76, 77).

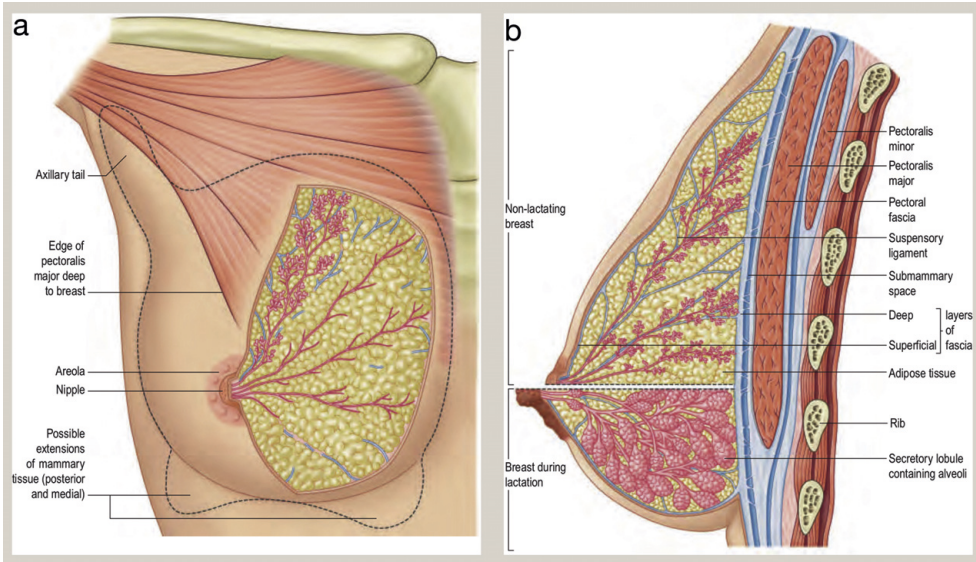


Figure 6.
A schematic image with an overview of (A) the breast and (B) its lobular strucuture by lactation status. "Anatomy and physiology of the breast" by Peter J. Bazira and colleagues (80). © 2022 Elsevier. Reprinted with permission from Elsevier

There are three distinct mammary epithelial cell types: luminal hormone-responsive (LumHR), luminal secretory (LumSec), and basal-myoepithelial (basal) cells (77, 81, 82). LumSec cells synthesize milk components during lactation, which are collected in the alveoli and duct lumen, while LumHR cells integrate endocrine signals to regulate mammary growth and differentiation (77). The luminal cells are usually the ones that give rise to breast cancer (83). Basal cells contract to aid milk secretion during lactation and provide structural support to the two luminal cell populations (77). Other significant cell types identified in the breast include fibroblasts, perivascular cells, adipocytes, lymphatic cells, vascular cells, myeloid cells, T-cells, and B-cells (81, 82). Depending on age, menopausal status, and ethnicity, the composition of the breast can vary in terms of breast-cell type and cell state (81, 82). Furthermore, breast composition can vary with additional host factors, exogenous hormone use, reproductive history, breast density, and obesity (81, 82).

A pivotal link between breast cancer tumorigenesis and normal development is the female sex hormone estrogen (84-86). In women, there are three primary endogenous forms of physiological estrogens: estrone (E1), estradiol (E2), and estriol (E3) (86, 87). After menopause, E1 assumes significance as it is produced in adipose tissue and is the main source of estrogen. E2 is considered the most potent product of estrogen biosynthesis, playing a major role in premenopausal women (84, 87). In premenopausal women, estrogen is mostly produced in the ovaries. The

least prevalent estrogen form, E3, is derived from E1 or E2 and is notable during pregnancy when the placenta produces it in substantial quantities (86, 87).

ERs are responsible for mediating estrogen's actions and functions and has two isoforms, ERα and ERβ (86, 88). This dimeric nuclear protein acts as a transcription factor, binds to DNA, regulates gene expression, and initiates a transcriptional program (88). ERα is a transcription factor for genes linked to cell survival, proliferation, and tumor growth and plays a crucial role in hormone-dependent breast cancer growth (84, 89, 90). Apart from genomic signaling, estrogen initiates non-genomic pathways by inducing cytoplasmic ERα, which forms a complex with Src and phosphatidylinositol 3-kinase (PI3K), activating protein kinase B (AKT) and thereby stimulating proliferation and cell survival (84, 89, 90). In contrast, the role of ERβ remains controversial, displaying both proliferative and antiproliferative characteristics(89, 90). ERα will be hereafter referred to as ER.

Akin to ER, progesterone receptor (PR) is a hormone-dependent nuclear transcription factor (91). PR mediates the impact of progesterone on mammary gland development and is expressed in about two-thirds of all ER-positive breast cancers (91). PR is a transcriptional target of ER, and ER signaling activation leads to increased PR expression, which redirects ER binding to the genome initiating a unique transcriptional program (90, 92). Approximately 20% of all breast cancers also exhibit overexpression of human epidermal growth factor 2 (HER2), which is attributed to amplification or activating mutations of the HER2 gene (*ERBB2*), indicating a more clinically aggressive disease (7, 93-96). HER2 is a member of the epidermal growth factor receptor family and requires dimerization for activation, although no ligand has been identified to date (93, 94). Dimerization activates the cytoplasmic kinase domain, leading to phosphorylation of a specific tyrosine kinase and activation of intracellular signaling pathways involved in cell proliferation and survival (93, 94). Estrogens play a crucial role in the normal development of breast epithelium by stimulating proliferation and ductal morphogenesis (85). However, exposure of luminal cells to high levels of estrogens lead to a pro-proliferative effect, causing the accumulation of replication errors, mutations, and the development of breast cancer (84, 86, 97). In response to estrogen stimulation, proliferating cells experience increased energy demands, leading to heightened mitochondrial activity and elevated levels of reactive oxygen species (ROS) as a byproduct of cellular respiration (86, 98). Moreover, evidence suggests that estrogen can exert carcinogenic effects independently of ER (84). Estrogens can undergo metabolism to catechol metabolites, followed by further oxidation to semi-quinones and quinones through a redox cycling process that generates ROS (99, 100). This is significant in tumorigenesis as estrogen quinones possess mutagenic properties and can directly interact with DNA, forming adducts that constitute a form of DNA damage (84).

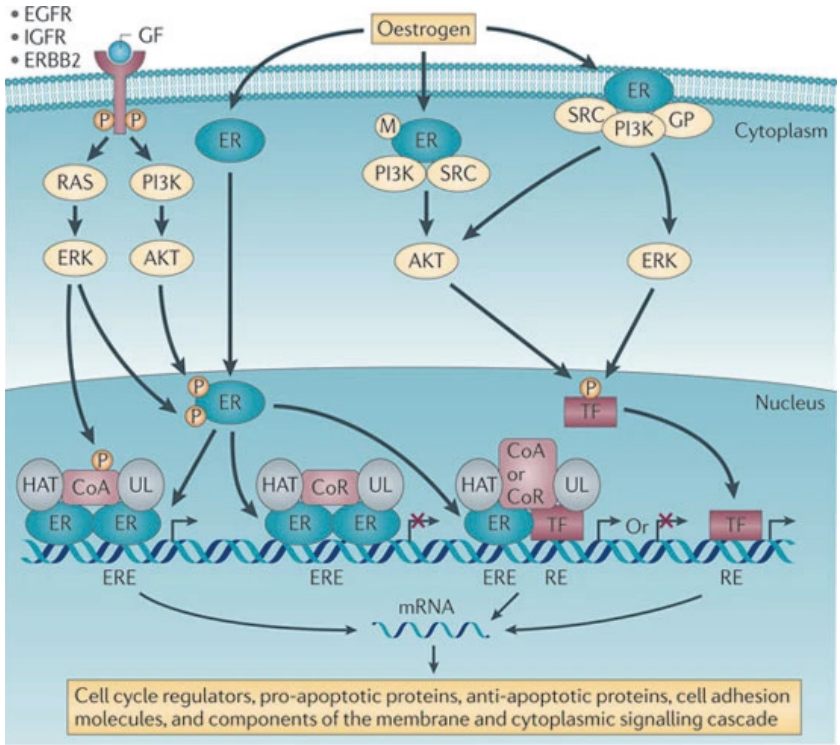


Figure 7.
Schematic overview of the canonical (genomic) effects of ER and the crosstalk between ER and HER2. As illustrated in Thomas et al. (101) © 2011 Nature Publishing Group. Reprinted with permission from Nature Publishing Group.

Numerous studies have demonstrated that treatment of normal breast epithelial cells with estrogen metabolites results in an increase in intracellular ROS, leading to oxidative DNA damage (99, 100, 102, 103). Importantly, by directly interacting with DNA, estrogen metabolites do not rely on the ER to exert their mutagenic effects (84, 103). This may elucidate the role of estrogen in promoting some ER-negative breast cancers as estrogen metabolites can induce double-strand breaks (DSBs) in both normal breast epithelial cells and ER-negative breast cancer cells (84, 103).

Other hormones and growth factors, such as insulin-like growth factor (IGF-1), also significantly influence the development of breast cancer (104-106). IGF-1 plays a pivotal role as a key mediator in the formation of mammary ducts during development (106, 107). It exerts various effects, including mitogenic, anti-apoptotic, and cell cycle initiation effects, which are primarily mediated by the transmembrane tyrosine kinase receptor IGF-1R (104-106). Upon ligand binding, IGF-1R undergoes phosphorylation, activating two major signaling cascades

through insulin receptor substrate 1 (IRS-1): the PI3K/AKT pathway and the rapidly accelerated fibrosarcoma kinase (RAF)/mitogen-activated protein kinase (MAPK) pathway (104-106). These pathways stimulate proliferation and provide protection against apoptosis (104-106). In both normal mammary glands and malignant breast tissues, IGF-1 is predominantly expressed by stromal cells, with only occasional expression by epithelial cells (106, 108). In contrast, IGF-1R is primarily expressed in the mammary epithelium (106, 108).

Elevated levels of circulating IGF-1 have consistently shown an association with an increased risk of breast cancer, particularly for ER-positive breast cancer (109-112). The bioavailability and half-life of circulating IGF-1 are regulated by a family of six IGF-binding proteins (IGFBPs) (113), and some related IGFBPs, such as IGFBP-related protein 1, which is also called IGFBP7 (114). Each IGFBP exhibits a high affinity for binding to IGF-1 and is subject to regulation by specific IGFBP proteases (106, 115). IGFBP-related proteins have a lower affinity for IGF-1 (116). Approximately 1% of circulating IGF-1 remains unbound, with the majority primarily binding to IGFBP3, forming a complex with an acid-labile subunit (106, 115). Adding another level of complexity is the bi-directional crosstalk between the IGF-1 system and the ER (106). Although the exact mechanism is not currently known, it is evident that breast cancer cells have a differential response to IGF-1 with regard to both proliferation and survival, depending on their ER status (104-106). Specifically, cells expressing both IGF-1R and ER demonstrate synergistic or additive growth effects in response to simultaneous administration of ligands (IGF-1, E2) (117). Many components of the IGF-1 system are also under the transcriptional control by ER (106).

Epidemiology and breast cancer risk factors

The risk of each individual developing breast cancer is based on a complex interplay between several lifestyle, reproductive, genetic, and environmental factors (6, 7, 118). There are several established risk factors for breast cancer, and the two most important are female biological sex (> 99% cases occur in women) and increasing age (6, 7, 118). These risk factors are non-modifiable and are called determinants. Beyond determinants, risk factors can also be classified as markers (e.g., socioeconomic status) and modifiable risk factors (alcohol, physical activity, exogenous hormone use, and obesity). When speaking of primary prevention on a population level, the modifiable risk factors are generally targeted. Primary prevention concerns preventing the onset of a health condition or disease in a population that has not yet been affected. It is estimated that approximately 30–40% of breast cancer can be prevented by a healthy lifestyle (119). Secondary prevention refers to methods of early detection and intervention to minimize the impact of a disease or condition, such as screening. Tertiary prevention involves managing and

improving the quality of life for individuals who are already affected by the disease by preventing further complications, relapses, and enhancing rehabilitation.

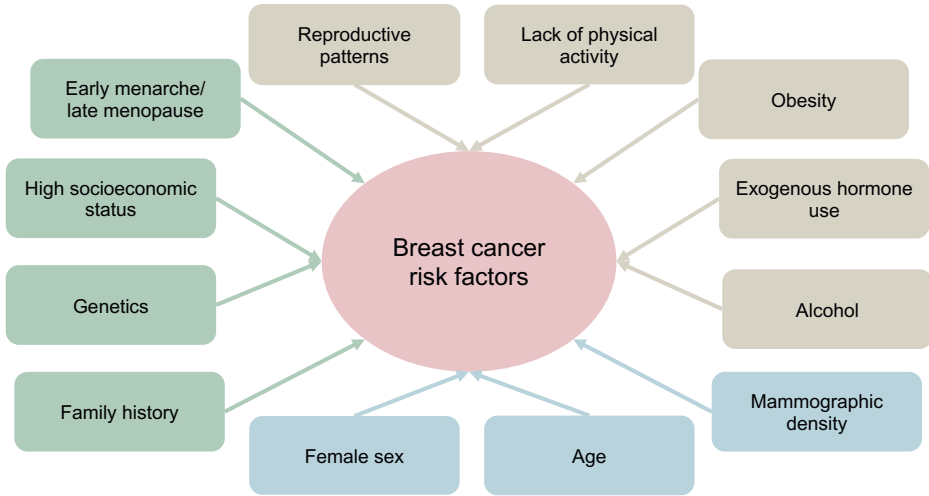


Figure 8.
An illustration of several key risk factors for breast cancer. Blue indicates the three most key risk factors for breast cancer, green indicates determinants and markers, and beige indicates modifiable risk factors.

Other important risk factors (determinants) of breast cancer risk are a history of benign breast disease and high mammographic density (120-122). For instance, women with atypical hyperplasia have an almost fivefold increase in risk for breast cancer (122, 123). Having dense breasts characterized by a higher proportion of fibroglandular tissue relative to fatty tissue is associated with a fourfold to sixfold increase in risk (120, 121).

Familial breast cancer and breast cancer genetics

Genetics and familial factors play a significant role in breast cancer risk (7). Approximately 10–15% of all breast cancers are considered familial; i.e., close female relatives have previously been diagnosed with breast cancer (124). Having one first-degree female relative (e.g., sister, mother, or daughter) with breast cancer roughly doubles a woman's risk of breast cancer compared to the general population, and the risk increases with the number of affected first-degree female relatives, especially when diagnosed at a young age (125, 126). Second-degree relatives (e.g., grandmother and aunt), having a male relative with breast cancer, and a family history of both ovarian and breast cancer are also familial risk factors for breast cancer (125, 126). It is essential to note that familial breast cancer and cancer in general extend beyond mere heritability (i.e., the influence of genetic factors on disease risk) (127, 128). Shared environmental factors such as socioeconomic status, lifestyle choices, and exposures *in utero*, during childhood, and adolescence, as well

as screening patterns in adult life contribute to the complexity of familial breast cancer (129).

In the 1990s, the identification and sequencing of the *BRCA1* and *BRCA2* genes revealed that certain genetic variants in these genes confer a substantial lifetime risk of developing breast cancer (130-132). The encoded proteins, BRCA1 and BRCA2, are integral components of a larger complex responsible for repairing DSB through homologous recombination (133, 134). Dysfunction in these genes leads to the accumulation of DSBs, contributing to tumorigenesis (133, 134). While pathogenic germline variants in *BRCA1* and *BRCA2* are rare, women carrying such variants face a significantly elevated lifetime risk of breast cancer (135, 136). Estimates indicate a lifetime risk in the range of 55–72% for those with *BRCA1* variants and 45–69% for those with *BRCA2* variants (135-138). Additionally, pathogenic variants of both *BRCA1* (44% estimated lifetime risk) and *BRCA2* (17% estimated lifetime risk) are linked to an increased risk of ovarian cancer (136, 138, 139). *BRCA2* variants are also associated with heightened risks of pancreatic, male breast, and prostate cancer (139, 140). Breast cancers resulting from pathogenic variants of *BRCA1* are typically characterized as ER-negative and/or triple-negative breast cancer (TNBC) (135, 141). *BRCA2*-associated breast cancer tends to be an aggressive ER-positive and HER2-negative subtype with high grade (135, 141-144).

Elevated risk of breast cancer is also associated with pathogenic variants in other genes with high penetrance (lifetime risk > 30%) to moderate penetrance (lifetime risk 17–30%) involved in homologous recombination repair and/or cell cycle regulation (145). Examples include *PALB2*, *CHEK2*, *ATM*, *BARD1*, *RAD51C*, and *RAD51D* (135, 146). Additionally, rare variants in syndrome genes like *TP53*, *PTEN*, *STK11*, and *CDH1*, which encode tumor suppressors, are linked to increased breast cancer risk (135, 141, 142, 146). However, carriers of variants in syndrome genes face heightened risks for various cancers, with breast cancer forming part of a broader syndrome.

Over the past decade, genome-wide association studies (GWAS) have identified multiple loci in the genome associated with a slightly increased risk for breast cancer (135, 141-143, 147-149). The most common genetic alterations identified are single-nucleotide polymorphisms (SNPs), which are substitutions of single nucleotides in DNA that are present in over 1% of the population. Several hundred SNPs have been pinpointed, which each individually confer a modest increase in risk that can depend on the subtype (141, 143, 147, 148). The cumulative effect of multiple variants in SNPs can lead to a significant overall risk, which is often summarized with a polygenic risk score (150, 151). The latest validated polygenic risk score for breast cancer comprises 313 SNPs and offers predictive capabilities for risk of breast cancer across diverse populations (150, 151). However, the predictive precision of such scores varies based on the ancestral background of the population (152-154). Furthermore, a notable limitation arises from the tendency of scores to primarily identify individuals who are prone to developing low-risk, low-

grade, small tumors, and non-aggressive ER-positive breast cancer (155). Gene–gene interactions and changes in the epigenome may further explain part of the heritability of breast cancer (149, 156). Several large cohort studies have found that epimutations are associated with increased risk of breast and/or ovarian cancer (157, 158). Further refinement of the polygenic risk score may improve its ability to identify individuals who may benefit from early detection measures.

Reproductive risk factors

Reproductive factors influencing breast cancer risk are closely tied to the lifetime exposure to the female sex hormone estrogen, particularly before the first pregnancy, when the breast is still developing and more sensitive to endocrine influences (7). Factors that extend the duration of estrogen exposure, such as early onset of menstruation and late menopause, are correlated with an elevated risk of breast cancer (159, 160). Additionally, age at first childbirth, the number of children, and breastfeeding are associated with breast cancer risk (160, 161). In both premenopausal and postmenopausal women, higher systemic levels of sex hormones are positively linked to an increased risk of breast cancer (162-164).

Endogenous hormones are not the sole contributors to breast cancer risk; menopausal hormone therapy, particularly formulations containing both estrogen and progestogen, can also elevate the risk of breast cancer (165). The use of oral contraceptives is an established risk factor for breast cancer, especially among younger women (6). The use of oral contraceptives before the age of 20 years is linked to the highest risk (166). Among women aged 20–44 years, the use of oral contraceptives (both contemporary and older formulations) has been associated with a slightly increased risk correlating with the number of years of use (166, 167). This heightened risk persists for up to 10 years after cessation (168, 169). Additionally, the use of hormonal intrauterine devices has also been associated with an increased risk of breast cancer (166).

Other lifestyle risk factors

A widely recognized risk factor for breast cancer is body fatness, which is often assessed through indicators such as the body mass index (BMI), waist circumference, or waist-to-hip ratio (6, 7). According to the World Health Organization (WHO), abdominal obesity is characterized by a waist–to–hip ratio exceeding 0.85 for females (or 0.90 for males). Alternatively, waist circumference can be utilized, with which central overweight is defined as ≥ 80 cm, while central obesity is defined as ≥ 88 cm. General body fatness is classified as overweight with BMI ≥ 25 kg/m^2 or obesity with BMI ≥ 30 kg/m^2 (170). High body fatness is linked to an increased risk of breast cancer among postmenopausal women, particularly for ER-positive breast cancer (171-173). Conversely, obesity is inversely associated with the risk of ER-positive breast cancer in premenopausal women (171, 174-176). Additionally, regardless of menopausal status, obese breast cancer patients tend to

have a poorer prognosis compared to those with normal weight (177, 178). Obesity-associated conditions, including metabolic syndrome and type 2 diabetes mellitus, are all also associated with an increased susceptibility to breast cancer (179, 180).

The precise mechanisms behind these phenomena are not yet fully understood (181, 182). Nevertheless, adipose tissue is not merely a passive reservoir for energy storage; instead, it serves as a highly metabolically active endocrine organ that is responsible for regulating metabolic substrates (such as free fatty acids, cholesterol, and triglycerides), as well as synthesizing and secreting various substances like adipokines, estrogen, adiponectin, cytokines, and leptin (181, 182). The disruption of adipose tissue homeostasis due to excess body fat leads to increased secretion of leptin and free fatty acids (182). Elevated leptin levels in adipose tissue have been reported to diminish the effectiveness of immune checkpoint inhibitors (ICIs), promote a pro-metastatic phenotype through the upregulation of EMT-associated genes, and enhance tumor growth, invasion, and metastasis (183, 184). Cancers (including breast cancer) utilize free fatty acids for proliferation and migration and store them within lipid droplets (182). Additionally, obesity alters the body's energy-balance signaling network, resulting in elevated systemic levels of insulin and estrogen (182).

The development of obesity triggers inflammation due to hypertrophic remodeling of adipose tissue, adipocyte necrosis, and dysregulated fatty acid flux from heightened adipocyte lipolysis (181). During adipose tissue expansion, rapid adipocyte hypertrophy may lead to insufficient angiogenesis, preventing proper tissue vascularization and resulting in hypoxic regions (182). Hypoxia activates hypoxia-inducible transcription factors (HIFs), which hinder preadipocyte differentiation and initiate adipose tissue fibrosis (185). In conjunction with hypoxia, stressed adipose tissue fosters immune cell infiltration and stimulates the release of inflammatory cytokines and chemokines from resident macrophages in adipose tissue (185). This chronic low-grade inflammation escalates the risk of mammary carcinogenesis (181, 182).

Insufficient physical activity and prolonged sedentary behavior are proposed as risk factors for both pre- and postmenopausal women (6). Alcohol consumption is also linked to breast cancer risk, which increases in a dose-dependent manner with higher intake (186). This is potentially due to alcohol-induced inflammation and elevated levels of endogenous estrogens (187). Regarding smoking, the association is less clear, possibly due to confounding by teenage contraceptive use. There is likely a small risk associated with smoking, which is partially mitigated by the anti-estrogenic properties of cigarettes and tobacco (188). Women with a higher socioeconomic status exhibit a greater incidence of breast cancer (189), which may possibly be attributed to lifestyle choices such as regular mammographic screening, the use of exogenous hormones, and reproductive patterns with delayed childbearing. Similarly, women with higher education also face an increased risk of breast cancer compared to those with lower education (190).

Diagnosis and detection

Sweden and most Western countries have a population-based screening program for early breast cancer detection (8). The program was fully implemented in Sweden in 1997. All women aged 40–74 years in Sweden are invited for screening every 18–24 months, and the attendance rate is slightly over 80% (191). About 60% of breast cancer cases among women in the screening-age group are detected through screening (5, 191). Interval cancers refer to cases diagnosed between scheduled screening intervals for women who regularly attend screenings. The current screening method is (digital) mammography involving X-rays taken from two projections (craniocaudal and mediolateral oblique) of the breasts (191). Findings prompting further investigation lead to recalls.

Breast cancer diagnosis involves triple assessments (clinical examination, imaging, and biopsy) for both recalled individuals and those with clinical symptoms (191). The common symptoms of breast cancer are palpable lumps or masses. Additional signs include swelling or thickening of the breast, alterations in the appearance, size, or shape of the breast or nipple, and peau d'orange (191). Clinical examination consists of palpation of the breasts and local lymph nodes (191). Imaging can be conducted through various methods, such as mammography, ultrasound, breast tomosynthesis, and magnetic resonance imaging. Generally, all women undergoing investigation for breast cancer receive both mammography (complementary in cases of recall) and ultrasound examinations of both the breast and the axilla. Biopsies are typically obtained using core-needle approaches, while fine-needle aspiration is the primary choice for examining axillary lymph nodes suspected of containing invasive tumors cells (191). In rare instances, biopsy results and pathology fail to provide conclusive evidence of breast cancer, but radiological findings strongly suggest it, so a surgical biopsy is performed using a breast-conserving surgical approach (191).

Tumor classification and molecular profiling

Breast cancer is recognized as a highly diverse disease, characterized by various subtypes exhibiting distinct biological features, clinical outcomes, and responses to treatment. In clinical practice, the classification of breast cancer uses histological, pathological, and molecular markers. This approach allows for a more accurate prognosis prediction and facilitates the tailoring of treatment modalities to suit each patient.

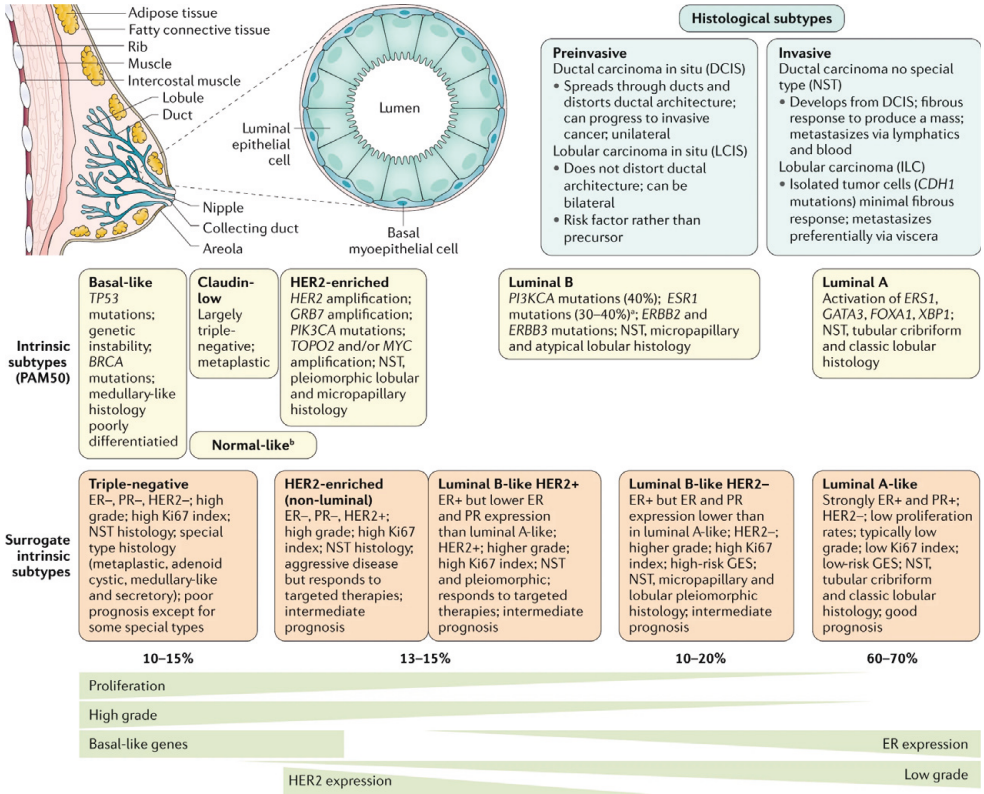


Figure 9.
An illustration of the complex classifcation of breast cancer according to molecular and histological features from Harbeck et al. (6). The differences between surrogate intrinsic subtypes and intrinsic subtypes are highlighted as well as the key features of each subtype. © 2019 Springer Nature. Reprinted with permission from Springer Nature.

Histological classification

Per the 2012 WHO classification, invasive breast carcinomas are categorized into 19 major subtypes (192). Approximately 70% of all breast cancers fall under the designation of "invasive carcinoma of no special type" (NST), which was previously known as ductal carcinoma (192). This histological type displays a highly variable morphology and does not align with a specific type, hence its descriptive name. The most prevalent special type is lobular carcinoma, which constitutes around 20% of all breast cancers. Additionally, there are less common types, such as tubular, mucinous, cribriform, metaplastic, and micropapillary carcinoma (192). Prognosis and molecular features vary among histological subtypes, particularly for some of the rarer types (192). However, the clinical utility of this classification system is limited due to the rarity of many special histological subtypes.

Histological grade

The morphological evaluation of tumors involves assessing the degree of differentiation in breast cancer (193), essentially gauging how closely it resembles normal, noncancerous breast tissue. This assessment follows the Nottingham histological grade, which considers three histomorphological features to establish the final grade: tubule formation, nuclear pleomorphism, and mitotic count (193). Grade I indicates well-differentiated tumors, Grade II signifies moderately differentiated ones, and Grade III represents poorly differentiated tumors, reflecting aggressiveness of the tumor (193).

Tumor stage

The Tumor Node Metastasis (TNM) staging system is applicable to all solid tumors and measures the extent of the cancer burden. Patients are stratified into four prognostic groups (I-IV) based on three key parameters: the size of the primary tumor (T), involvement of regional lymph nodes by cancer (N), and the presence of distant metastases (M) (194, 195). The prognostic value of these parameters varies across different cancers (194, 195).

For breast cancer, the parameters are classified as follows. Invasive breast cancer size (T) is categorized into four groups: T1: ≤ 20 mm; T2: > 20–50 mm; T3: > 50 mm; and T4: involvement of skin or muscles, irrespective of size. The lymph nodes (N) are also grouped into four categories: no positive nodes, 1–3 positive nodes, 4–9 positive nodes, and 10 or more positive nodes (194, 195). The N stage is also influenced by the location of the pathological lymph nodes (194, 195). The M stage is dichotomous, indicating either the presence or absence of apparent distant metastases (194, 195). Staging can be based on clinical pre-surgery parameters ("c" as a designator), or information derived from surgery and pathological assessment ("p" as a designator). In cases involving neoadjuvant treatment, the designator "yp" standing for "yield pathological" is used for post-chemotherapy staging (194, 195). To align with the current understanding of cancer and provide a more nuanced prognostication of patients, the 8th and latest version of the TNM-staging system of the American Joint Committee on Cancer (AJCC) Prognostic Stage Group incorporates non-anatomical factors such as tumor grade and tumor receptor status (ER, (PR), and HER2) alongside the traditional TNM variables in determining the prognostic stage group (194, 195).

Immunohistochemical biomarkers

In clinical practice, ER, PR, HER2, and antigen Kiel 67 (Ki67) are four routinely employed immunohistochemical biomarkers at the time of diagnosis (145, 196). These markers hold international recognition and are deemed crucial for guiding therapy decisions (145, 196). These markers are analyzed in tissues obtained from surgical specimens and/or samples obtained during pre-surgical core needle biopsies (191). In Sweden, a tumor is deemed ER-positive if 10% or more nuclei express ER,

whereas the international standard employs a more common cutoff value of more than 1% positive nuclei (145, 191, 196). In Sweden, a tumor is considered PR-positive if 10% or more nuclei express PR, while the international standard uses a cutoff of over 1% positive nuclei (145, 191, 196). Clinical assessment of HER2 status commonly utilizes two methods: immunohistochemistry (IHC) and in situ hybridization (ISH). IHC is a semi-quantitative method that categorizes patients into four groups (0, 1+, 2+, and 3+), while ISH quantitatively measures the number of HER2 gene copies in each tumor cell, and the result is reported as either positive or negative (94). A tumor is classified as HER2-positive when scoring 3+ by IHC and is characterized by strong, complete membrane staining (145, 191, 196, 197). Tumors scoring 2+, with weak to moderate complete membrane staining in > 10% of tumor cells, are considered equivocal and are subjected to ISH testing according to Swedish national guidelines to determine HER2 status (191, 197).

Approximately 15% of breast cancer tumors lack expression of the three immunohistochemical markers (ER, PR, and HER2) and are categorized as TNBC (198-201). TNBC is recognized for its biological aggressiveness, high-grade and highly proliferative cancer cells, and it often presents as invasive ductal carcinoma (198-201). TNBCs exhibit considerable variation, characterized by complex genomes, heightened genetic instability, and both intertumor and intratumor heterogeneity (198-201). The predominant gene mutation is *TP53*, although several other genes show mutations at lower frequencies (198-201). Clinically, TNBC patients face a heightened risk of early metastasis and breast cancer-related mortality within five years of diagnosis (198-201). Limited treatment options contribute to the challenging prognosis, making TNBC the subtype with the poorest outcomes among breast cancers (198-201).

Ki67 is a nuclear protein expressed in all phases of the cell cycle, excluding the inactive G0 phase. Ki67 expression is associated with the proliferative rate of tumor cells and is therefore a proliferation marker (202). However, there is a lack of international consensus on the most appropriate method for Ki67 scoring and cutoffs, and when coupled with inter- and intra-laboratory variabilities, pose challenges to its standardization (203, 204). In accordance with Swedish national guidelines, the counting of 200 tumor cells within a hotspot region is recommended, and there are laboratory-specific cutoff values (191).

Molecular subtypes

High-throughput technologies have revolutionized the characterization of breast cancer, provided unprecedented understanding of its biology and heterogeneity, and provided significant clinical implications. The landmark study by Perou and colleagues introduced an independent classification of breast cancer based on gene expression profiles and four distinct subtypes: luminal epithelial/ER-positive, basal epithelial, normal-breast like, and HER2-enriched subtypes (59). The follow-up study confirmed and expanded these findings, subdivided the luminal epithelial/ER-

positive cluster into A and B, and established the clinical implications and prognostic value of gene expression patterns (60). Numerous studies across diverse settings consistently validate the conservation of subtypes, which correlate with specific biological and molecular features and offering independent prognostic information (205-207). Further, the intrinsic molecular subtypes harbor distinct features in several different omic layers, including genomic with distinct mutation patterns and copy-number aberrations, epigenomic with different methylation patterns and microRNA profiles, and distinct proteomic patterns (207).

The luminal subtypes are predominantly characterized by ER and PR positivity. Luminal A tumors exhibit elevated expression of ER-related genes such as *GATA3* and *FOXA1*, along with mutations in the *PIK3CA* gene (207). Luminal A tumors have relatively decreased proliferation and the most favorable prognosis among all intrinsic subtypes (207, 208). In contrast, luminal B subtype tumors have lower expression of ER-related genes than luminal A, an increased proliferation rate, and a distinct mutation pattern, including more *TP53* mutations and *ATM* loss (207, 208). The HER2-enriched subtype prominently expresses genes related to the HER2 signaling pathway and frequently presents amplification of the *ERBB2* gene and low expression of ER-related genes (207), which correspond to HER2-positive/ER-negative tumors (207). The basal-like subtype frequently correlates with the IHC definition of TNBC and has higher expression of basal markers from the myoepithelium in normal breast, such as *KRT5*, *KRT17*, and *EGFR* (207, 208). This subtype also exhibits a high frequency of TP53 mutations and the loss of other tumor suppressors *RB1* and *BRCA1* leading to high genomic instability (207). The basal-like subtype is generally considered the most distinct among the intrinsic subtypes (207). The normal-like subtype is identified by low expression of proliferation-related genes and a gene expression pattern resembling that of normal epithelial breast cells (207, 209). A sixth subtype known as the claudin-low subtype has been identified based on gene expression characteristics and is primarily characterized by low cell–cell adhesion, EMT, and stem cell-like features (207, 210-212). Claudin-low tumors exhibit significant immune and stromal cell infiltration but are otherwise remarkably heterogeneous in many aspects (210, 211). Recently, there has been a suggestion that claudin-low might be more appropriately described as a breast cancer phenotype rather than an intrinsic subtype (222, 223). Unlike other subtypes, both normal-like and claudin-low tumors do not exhibit specific genomic aberrations, leading to discussions about whether they represent true subtypes (207, 209, 213). Therefore, the commonly referenced molecular intrinsic subtypes are luminal A, luminal B, HER2-enriched, and basal subtypes.

To facilitate the integration of intrinsic molecular subtypes into clinical practice, Parker and colleagues created a concise gene list of 50 genes using the Prediction Analysis of Microarray (PAM) method, which simplifies the classification of breast cancers (214). Subsequently, they developed the risk of recurrence (ROR) score using the correlations with four of the five molecular intrinsic subtypes (basal-like,

HER2-enriched, luminal A, and luminal B), alongside clinicopathological features to predict the risk of distant metastasis within 10 years of breast cancer diagnosis (214). This approach underwent further refinement, culminating in the development of the Prosigna® assay. The PAM50 subtypes, PAM50 ROR, and Prosigna® assay have been extensively validated, especially for postmenopausal ER-positive/HER2-negative node-negative patients (215-223).

Other efforts to classify breast cancer have been made by integrating gene expression data and offer prognostic information or by a more thorough genomic approach integrating chromosomal aberrations and gene expression profile patterns (212, 224-236). The latter is known as the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) integrative cluster classification. To date, however, it has not been developed as a clinical tool to implement therapeutical decisions. The integrated cluster classification divides breast cancer based on copy-number aberrations into 10 groups that have distinct gene expression profiles and mutations patterns and clinical outcomes (212, 234-236). Akin to the PAM50 ROR score, additional gene signatures have been developed to predict the risk of relapse. Two stand out prominently. The 70-gene signature developed by Van't Veer et al. by agnostic comparison of the gene expression profiles of tumors that relapsed and those that did not (237). (237). Additionally, the 21-gene signature discovered by Paik and colleagues is derived from an *a priori* list of genes hypothesized to predict relapse (238). These assays have been extensively validated and developed into MammaPrint® and OncotypeDx®, respectively (224-230).

Until recently, in Swedish clinical practice, an approximate surrogate classification for determining the intrinsic molecular subtypes of breast cancer based on histology and IHC biomarkers has mainly been used (191). However, certain issues remain the proxy classification is not the best approximation and for example struggle to distinguish luminal A and B tumors (239, 240). Importantly, these intrinsic molecular subtypes not only differ based on different gene expression profiles, they also present independent prognostic value apart from the more commonly used IHC biomarkers (215-223, 241). Gene expression profiling according to international guidelines can only be used to guide treatment choices in postmenopausal women diagnosed with an ER+/HER2- breast cancer with up to three positive axillary lymph nodes (191). Gene expression profiling may also be used in cases of an ER+/HER2- breast cancer with intermediate risks such as a grade II tumors with intermediate Ki67 (191).

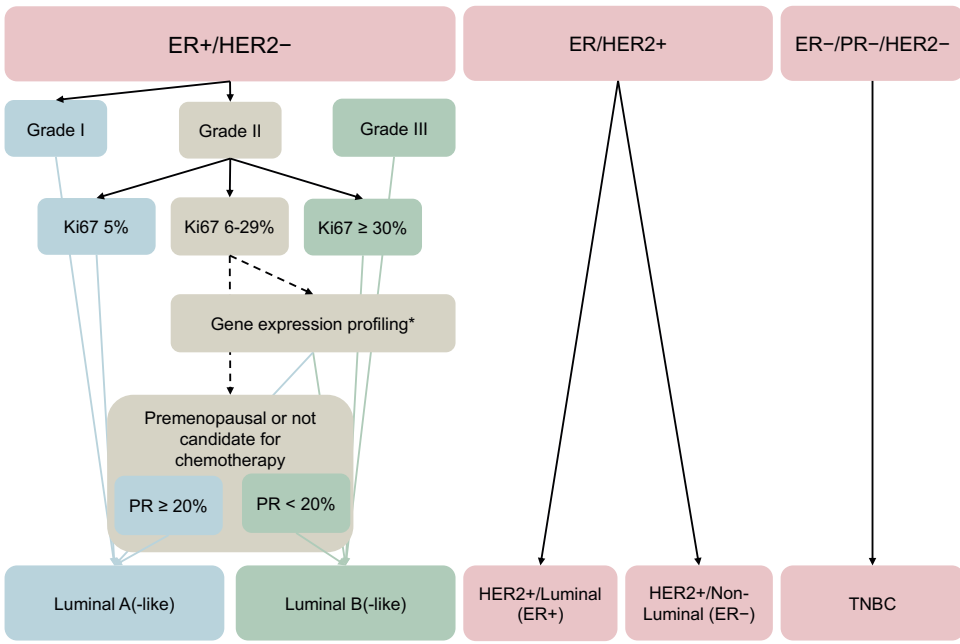


Figure 10.
Classification of breast cancer by surrogate intrinsic subtype according to Swedish guidelines (191). The classification of surrogate subtypes into Luminal A-like and B-like should consistently include an assessment of the plausibility of such categorization (191). In instances of discrepancy, a reevaluation of the case is warranted.*Gene expression profiling should be utilized in postmenopausal women with ER-positive/HER2-negative breast cancer when uncertainty exists regarding tumor risk categorization before chemotherapy selection (191). Additionally, gene expression profiling may be warranted when IHC evaluation alone dictates the decision to use chemotherapy (191).

Beyond general breast cancer classifications, various subtype-specific schemas have been developed for ER-positive/HER2-negative, HER2-positive, and TNBC cancer (242, 243). TNBC in particular, has garnered significant interest for classification due to its heterogeneity, leading to proposals of several schemas over the years (198-201). The most well-known classification is the Lehman subtypes, which include basal-like (BL1 and BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL), and luminal androgen receptor (LAR) subtypes, further refined into BL1, BL2, M, and LAR (244, 245). Other classification systems primarily employ the four features of basal, mesenchymal, luminal/androgen, and immune activation for classification (246-248). However, it is worth noting that none of these classification schemas are currently utilized in clinical practice (145).

Contralateral breast cancer

Contralateral breast cancer (CBC) is characterized by occurrence in both breasts either simultaneously (with the second breast cancer diagnosed within six months

to a year) or sequentially (with the second breast cancer diagnosed a year after the initial diagnosis). In essence, CBC is typically treated as a new primary tumor and is diagnosed and managed accordingly (191). However, certain studies investigating the genomic relationship between the first and second breast cancers propose that a subset of CBC cases may represent a metastasis from the initial primary breast cancer (249, 250). Breast cancer patients have elevated risk of developing a new tumor in the contralateral breast compared to individuals without a history of breast cancer (251). Studies also indicate that the occurrence of CBC confers a higher risk of breast cancer-related death compared to other breast cancer patients, particularly if the second tumor arises within four to five years or is synchronous (252, 253).

Prognostic and treatment-predictive factors

Prognostic and treatment-predictive factors are important considerations in medical decision-making, particularly in the context of managing diseases like breast cancer. Prognostic factors are characteristics or variables associated with the natural course and outcome of a disease, independent of any specific treatment intervention. They help predict the likely course of the disease and the patient's overall outcome. Therefore, they aid clinicians in estimating the patient's prognosis and can influence treatment decisions. Treatment-predictive factors, on the other hand, are characteristics that help predict how a patient is likely to respond to a specific treatment. They are indicators of the effectiveness of a particular therapeutic intervention. Treatment predictive factors guide clinicians in selecting the most appropriate and effective treatment for an individual patient. By identifying these factors, therapies can be tailored to maximize benefits while minimizing potential side-effects. Certain factors can be both.

Host factors

The most firmly established and arguably the most influential prognostic host factor for breast cancer is age. Women diagnosed with breast cancer at a very young age, specifically under 35 or 40 years, exhibit a higher risk of recurrence and experience lower survival compared to those who are older (254, 255). The mechanism underlying this phenomenon is not entirely clear, but there is a suggestion that the age at diagnosis is associated with certain molecular characteristics that may confer a worse prognosis (254, 256-258). Additionally, it is well established that older women, particularly those above 80 years, have poorer outcomes, which is partially explained by increased comorbidities that limit treatment options (255, 259).

The field of pharmacogenomics has gathered increasing attention, given the belief that individual responses to drugs are substantially influenced by genetic variations, particularly in genes coding for proteins responsible for drug absorption, distribution, metabolism, and elimination (260). Currently, clinical testing involves examining variants in the *dihydropyrimidine dehydrogenase (DPD*) gene related to

the metabolism of capecitabine or 5-fluorouracil (191). A randomized clinical trial (RCT) has demonstrated that genotype-guided dose reduction of capecitabine or 5-fluorouracil significantly reduces the risk of adverse events (261, 262). There is ongoing potential for the discovery and validation of additional genomic variants that could further enhance treatment guidance (260). Another critical host factor is obesity, which has been consistently linked to poorer prognosis and increased risk of distant metastasis, although it is not currently utilized in clinical settings (177, 178). One plausible explanation could be that chemotherapy doses are capped to a maximum body-surface area (BSA) of 2.0 m^2, potentially resulting in underdosing for obese or overweight patients (263, 264).

Standard clinicopathological factors¨

Perhaps the most important individual way to classify breast cancer in terms of prognosis is the TNM stage. It is likely the most important factor for determining prognosis and is very influential for determining whether a case of breast cancer is high-risk or low-risk, which is highly influential in treatment decisions (194, 195). In addition, The Nottingham histological grade is another key prognostic tumor characteristic, and together with the stage, it forms the basis of most clinical prediction models like "PREDICT" (193-195, 265-267).

Beyond these factors, the four IHC tumor markers are the most widely recognized and important as both prognostic and treatment-predictive markers in breast cancer (268). ER is both a prognostic factor associated with better prognosis and a treatment-predictive factor for endocrine treatment, and all types of endocrine treatment are directed at ER-positive tumors (70, 269). According to our current understanding, PR is not considered to be treatment predictive; instead, it is a prognostic factor associated with improved outcome in ER-positive and ER-negative disease (70, 269-271). Since PR is associated with improved outcome comparable to ER-positive tumors, in many countries, tumors are classified as hormone-receptor positive, meaning ER and/or PR-positive tumors, which are treated as ER-positive tumors (268). HER2 is a treatment-predictive factor for HER-targeting agents but also a negative prognostic factor (96, 197, 272).

In recent years, a key prognostic factor that has gained prominence is the pathological complete response (pCR), which is defined as "the absence of invasive cancer in the breast and axillary nodes" (i.e. ypT0/Tis ypN0) (145). It serves as a crucial indicator of treatment response and has received FDA approval as an endpoint for clinical trials. It is linked to a significantly lower likelihood of recurrence and death in breast cancer (145, 273). Notably, the association with pCR is subtype-dependent, and the strongest correlations are observed in HER2-positive/ER-negative tumors, followed by TNBC (273). In contrast, for ER-positive/HER2-negative disease, pCR is not as strongly predictive, and neoadjuvant treatment is not the standard practice for this subtype (191, 273). Nevertheless, a similar metric of response known as the endocrine-sensitive disease rate (EDR) has

been devised (274, 275). EDR combines the assessment of Ki67 and ER expression in tumors following neoadjuvant endocrine treatment. Even when the tumor has not completely disappeared, those with low Ki67 and high ER levels exhibit an excellent outcome and a favorable response to endocrine treatment (274, 275).

Gene expression assays such as Mammaprint®, OncotypeDx®, and Prosigna® also play a key role in classifying ER-positive/HER2-negative tumors into molecular high-risk and low-risk categories (145). (145). Notably, the results from the MINDACT, RxPONDER, and TAILORx trials have demonstrated that chemotherapy may be omitted for tumors that are clinically deemed high-risk but molecularly categorized as low-risk by Mammaprint® and OncotypeDx® (215, 222, 229, 276-279). This represents a substantial treatment de-escalation and emphasizes that molecular profiling goes beyond identifying potential treatment targets. Currently, the OPTIMA trial (ISRCTN42400492) is investigating whether the Prosigna® assay can be utilized in a similar manner.

Treatment

Multi-disciplinary conference

In accordance with Swedish national guidelines, it is recommended that a multidisciplinary conference be held to discuss treatment plans for all breast cancer patients. The aim of this collaborative approach is to optimize individualized treatment strategies and ensure the inclusion of diverse perspectives in the decision-making process (191). Typically, the team comprises surgeons, oncologists, pathologists, radiologists, and nurses, who discuss each patient's case both before and after surgery (191).

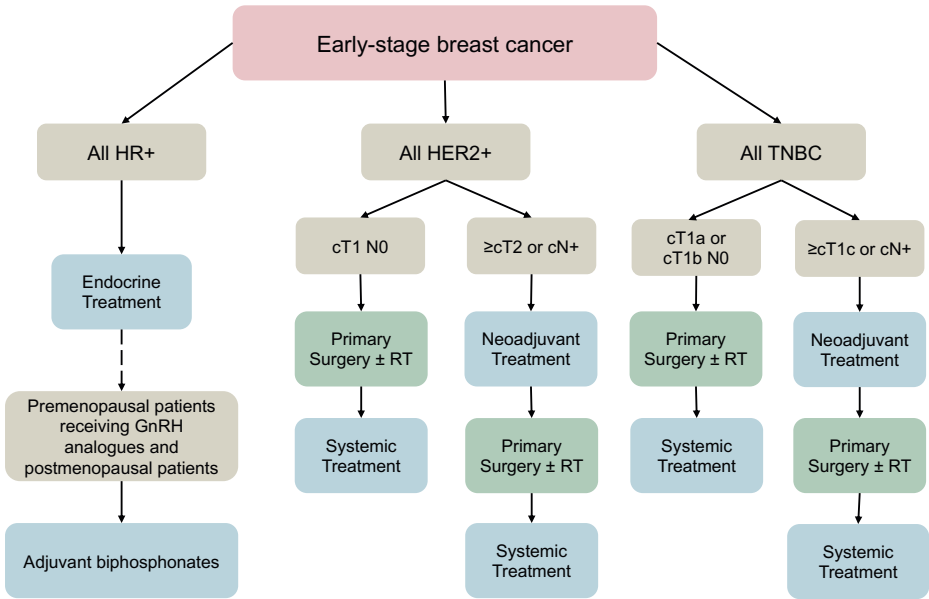


Figure 11.
Schematic overview of the early-stage breast cancer treatment according to ESMO guidelines (196). The key points of breast cancer treatment are summarized here. Further detail are provided in figures 11-13. GnRH: gonadotropin-releasing hormone. RT: radiotherapy.

Surgery

Surgery plays a pivotal role in the treatment of breast cancer. Developed in the late 19th century, radical mastectomy involves the removal of the entire breast, pectoral muscles, and axillary lymph nodes. This was the first surgical approach for breast cancer and was described by Halsted (280). Presently, the preferred surgical approach is breast-conserving surgery when feasible to achieve tumor removal by partial mastectomy with favorable cosmetic outcomes (145, 191). Results from RCTs with more than 20 years of follow-up indicate that breast-conserving surgery

followed by postoperative radiotherapy provides comparable survival outcomes to mastectomy (281-283). Population-based registry studies have also demonstrated improved breast cancer survival with breast-conserving surgery compared to mastectomy, although there could be potential biases and residual confounding (284-286). The importance of surgical margins cannot be overstated. Positive surgical margins indicate tumor growth in the resection margin, which is referred to as "tumor on ink" and indicates a doubled risk of local recurrence (6, 287). Even with postoperative oncological treatment, this risk will persist, although it will decline. To mitigate the need for re-operation due to positive margins, a macroscopic margin of 10 mm is recommended during the surgical procedure (191, 287).

Various oncoplastic breast-conserving techniques can be performed to allow for more extensive resections needed for larger tumors while still having good cosmetic outcomes (288). These advancements have expanded the indications for breast-conserving surgery (288), and as a result, more women with breast cancer now undergo this procedure (191). For a minority of women, breast-conserving surgery is not feasible due to various reasons such as contraindications for postoperative radiotherapy or the presence of multifocal tumors where good aesthetic outcomes cannot be achieved (191, 288). For these women, immediate or delayed reconstruction after mastectomy is an option (191, 288).

Similarly, axillary surgery has undergone de-escalation (145). As the axillary lymph nodes are often the initial site of breast cancer metastasis, the goal of axillary surgery is to stage the nodes or remove preoperatively identified metastases. Sentinel node biopsy has become the primary method and involves the identification and removal of the first lymph nodes that drain the breast tumor for pathological analysis (145, 191). This technique has replaced axillary dissection as the standard procedure and offers comparable breast cancer outcomes with significantly fewer associated side-effects.

There has been further discussion about increased de-escalation of axillary after the ACOSOG Z0011 trial showed non-inferiority of omission of axillary dissection after positive (macro-metastases) sentinel lymph node biopsy in patients with clinically node-negative breast undergoing breast-conserving surgery (289-291). Ensuing trials have confirmed the findings. The AMAROS trial demonstrated non-inferiority of axillary radiotherapy to axillary lymph node dissection for patients with a positive sentinel node (292). In Sweden, the SENOMAC trial also confirmed that it is safe to omit axillary dissection for wide range of patients with one to two positive sentinel nodes, including patients treated with mastectomy (293). Currently, if one to two sentinel node biopsies are positive for macro-metastases, complementary radiotherapy targeting the axilla is administered based on the results of the AMAROS and SENOMAC trials (191, 292, 293). Recently, the SOUND trial demonstrated that patients with small breast cancer (invasive tumor size < 2 cm) and node-negative preoperative axillary ultrasonography, sentinel node biopsy can

be safely omitted without affecting clinical outcome (294). Historically, axillary dissection has been performed for all patients with preoperatively determined axillary lymph-node involvement and macro-metastases (> 2 mm) detected in one or more sentinel node biopsies.

Axillary dissection is now reserved for cases involving preoperatively determined axillary lymph node engagement and no planned neoadjuvant treatment, macro-metastases (> 2 mm) in more than two sentinel node biopsies, or viable residual metastasis after preoperative treatment (191). In the case of neoadjuvant treatment, a targeted axillary node dissection and sentinel node biopsy are recommended, and only if viable tumor cells persist can a complementary axillary dissection be performed according to Swedish national guidelines (191).

Radiotherapy

Postoperative radiotherapy primarily serves as a local treatment to reduce the likelihood of a local relapse, abscopal effects not withstanding (295). Postoperative radiotherapy reduces the risk of recurrence and breast cancer mortality (145, 296). According to Swedish guidelines, postoperative radiotherapy is a standard treatment for all patients treated with breast-conserving surgery targeting residual breast tissue (191). Nevertheless, ongoing research is exploring the feasibility of omitting postoperative radiotherapy for low-risk tumors (297). In cases of mastectomy, postoperative radiotherapy directed towards the thoracic wall is generally recommended for tumors larger than 50 mm (191, 298). Irrespective of the surgical method, when lymph-node involvement is present, radiotherapy targeting regional lymph nodes is incorporated into the treatment (295). To mitigate side-effects and morbidity, recent advancements have shifted from conventional fractioning to hypo-fractionated approaches involving higher fractions per dose (145, 298). This modification shortens the typical radiotherapy duration from five to three weeks while maintaining comparable clinical results and potentially reducing side-effects (295). Recently, the FAST-Forward trial reported that five fractions over one week have similar 5-year local control to that of the standard 3-week protocol (299, 300).

Endocrine treatment

The initial realization of estrogen's pivotal role in certain breast cancers emerged inadvertently when George Thomas Beatson reported the beneficial effects of oophorectomy for patients with inoperable breast cancer in 1896, despite the hormone's existence being unknown at the time (301). Presently, there are two primary types of endocrine treatment for early-stage breast cancer: tamoxifen and aromatase inhibitors (AIs) (191). In Sweden, endocrine treatment is administered to nearly all patients with ER-positive tumors (around 70% of patients), excluding those with the smallest tumors (< 10 mm) and no lymph-node involvement, for whom it may be omitted given the excellent outcome of these tumors with just surgical excision (191). Tamoxifen is a selective estrogen receptor modulator that

exhibits both estrogen antagonistic and agonistic effects. It was discovered in the late 1960s and was gradually introduced for breast cancer treatment in the 1970s (302-304). It has proven to reduce the risk of relapse by 30% of tumors and is still the standard of care for mainly premenopausal women with low risk of recurrence for whom ovarian suppression is either not needed or contraindicated (6, 73). As the other primary endocrine treatment, AIs are primarily prescribed to postmenopausal women and inhibit aromatase, an enzyme that synthesizes estrogen in the liver, muscle, and fat tissue (97). In several trials (ATAC, TEAM, and BIG 1-98), AIs have demonstrated a small but superior effect to that of tamoxifen as an adjuvant treatment for postmenopausal breast cancer patients (305-310). Currently, the standard duration of endocrine treatment is five years with either AI or tamoxifen (6). However, for high-risk patients, such as those with axillary lymph-node involvement, the treatment duration may be extended to 10 years (6, 305). Postmenopausal patients are advised to combine bisphosphonates with AIs to lower the risk of skeletal metastasis and marginally improve survival (311).

For breast cancer patients under the age of 40 years who have a heightened risk of recurrence warranting adjuvant chemotherapy, ovarian suppression is added using gonadotropin-releasing hormone (GnRH) analogues to tamoxifen or AIs (312, 313). These analogues which hinder ovarian production of estradiol (312, 313). This approach is recommended for higher-risk patients since it comes with a battery of side-effects that can affect the long-term health of younger women (191).

Recently, the MonarchE trial demonstrated improved disease-free survival outcomes with the addition of cyclin-dependent kinase 4 and 6 (CDK4/6) inhibitors to the endocrine treatment for early-stage, high-risk ER-positive/HER2-negative, breast cancer patients (314-316). Both US and European guidelines endorse two years of adjuvant abemaciclib in combination with endocrine therapy as a standard of care for high-risk ER-positive early breast cancer. High-risk patients are defined as having four or more positive axillary lymph nodes or lower nodal involvement but larger tumors or tumors with more aggressive characteristics (191, 314-316). The recent NATALEE study used ribociclib as an CDK4/6 inhibitor in the first three years of adjuvant endocrine treatment and included patients with a lower risk profile (317). The results could possibly lead to the approval of this regime in an adjuvant setting, which would further increase the already complex algorithm for adjuvant endocrine treatment (317).

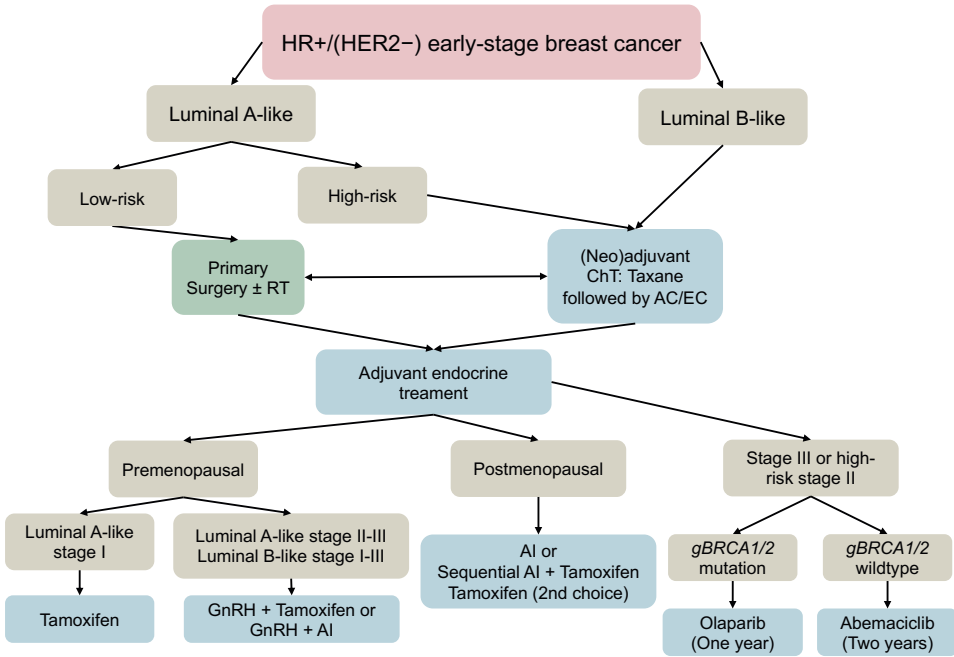


Figure 12.
A brief overview of endocrine treatment for hormone receptor (ER and/or PR)-positive breast cancer according to ESMO guidelines (196). "Low-risk" implies low-risk molecular score (preferred; e.g. MammaPrint® "Low"; OncotypeDX® ≤15; Prosigna® ≤60; luminal A) and/or lower-risk features on traditional pathological analysis including lower-grade histology and lower measures of proliferation. "High-risk" implies high-risk molecular score (MammaPrint® "High"; OncotypeDX® ≥26; Prosigna® >60; luminal B) and/or higher-risk features on traditional pathological analysis including higher-grade histology and higher measures of proliferation. The length of adjuvant endocrine treatment is decided based on a combination of anatomical stage, pathological, and molecular features. HR: hormone receptor. RT: radiotherapy. G: germline.

Chemotherapy

Chemotherapy was introduced as an adjuvant treatment for breast cancer in the 1970s (73). The current standard of chemotherapy is anthracycline–taxane regimens (145, 191). Comparisons between different combinations of agents have shown that the most efficacious regimen is anthracycline-based therapy (epirubicin or doxirubicin plus cyclophosphamide) with the addition of a taxane (docetaxel or paclitaxel) for 4–6 months (73, 318-320). This approach has demonstrated a significant reduction in 10-year breast cancer-related mortality by approximately 40%, irrespective of tumor characteristics, and there are no apparent adverse effects on mortality not related to breast cancer (73, 318-320). Advancements in chemotherapy administration, such as dose-dense scheduling (increasing the rate of delivery without altering the overall dose), have further enhanced breast cancer survival rates with tolerable toxicity (321). The specific administration details of chemotherapy, including dose intensity and the number of treatments, are tailored

based on different risk factors and individual patient considerations in standard clinical practice (145, 191).

Presently, chemotherapy is administered in both neoadjuvant and adjuvant settings, yielding comparable outcomes (145, 273, 322). Neoadjuvant chemotherapy serves as the standard treatment for patients with an inoperable primary tumor (6, 191) and those with stage 2–3 breast cancer (145). In certain subtypes like HER2-positive breast cancers and TNBC, neoadjuvant treatment has evolved into the standard of care as the achievement of pCR is correlated to clinical outcome and influences the choice of adjuvant therapy (6, 145, 273). According to Swedish guidelines, adjuvant chemotherapy is currently recommended for patients under the age of 35 years (191). For patients with triple-negative disease and HER2-positive disease, adjuvant chemotherapy is recommended for tumors larger than 5 mm (191).

For TNBC, other chemotherapeutic agents have been introduced to improve outcomes in neoadjuvant and adjuvant settings. Carboplatin is a platinum-based crosslinking agent that causes DNA damage and is currently combined with taxanes in a neoadjuvant setting as it improves the pCR rate and event-free survival in stage 2–3 TNBC (323-326). In case of TNBC patients undergoing neoadjuvant treatment, who do not achieve pCR, the addition of adjuvant capecitabine, an oral prodrug of fluorouracil, further improves overall survival (327).

HER2-targeted treatment

In the 1990s, the monoclonal antibody trastuzumab was the first HER2-targeted treatment to be developed. Trastuzumab was initially approved for the treatment of HER2-positive metastatic breast cancer in 2001, and since 2005, at has been approved for use in an adjuvant setting (6, 328, 329). It binds HER2 in the extracellular domain, and its main mechanism of action is the enhancement of antibody-mediated cytotoxicity, although it can also contribute to the inhibition of the pro-survival signals of HER2 intracellular activation (6, 93, 328, 329). Trastuzumab is given as a combination treatment with chemotherapy and has revolutionized the field of breast cancer therapeutics given that since its introduction, it has completely changed the outcome of HER2-positive disease (93, 328). Data from RCTs have shown that after receiving adjuvant treatment with trastuzumab, the recurrence rate is reduced by approximately 50% among HER2-positive patients (272, 330-332). After the remarkable progress with trastuzumab, other HER2-targeted therapies have been developed (328). Among these are pertuzumab, another monoclonal antibody that binds to the dimerization domain and inhibits HER2 heterodimerization with other HER family receptors (328). In the adjuvant setting, adding pertuzumab to trastuzumab has been shown to confer a slight but significant improvement in clinical outcomes, especially for node-positive patients (333-335). The combination of trastuzumab, pertuzumab, and docetaxel in a neoadjuvant setting is the standard treatment regimen of choice for high-risk HER2-positive breast cancer worldwide.

Another HER2-targteted treatment is the antibody-drug conjugate trastuzumab-emtansine (T-DM1), in which a payload is carried as “cargo” and links the tubulin inhibitor emtansine to trastuzumab (328). When patients on dual-HER2 blockade do not achieve a pCR, adjuvant treatment with T-DM1 has been found to significantly and substantially improve outcomes with an reduction of distant metastasis risk by approximately 50% (336). Neratinib is another successful HER2-targeting agent that stands out because its mechanism of action involves being a tyrosine kinase inhibitor that binds irreversibly to HER1 (EGFR), HER2, and HER4 (328). One year of extended adjuvant therapy using neratinib after completion of chemotherapy and trastuzumab improves clinical outcomes in HER2-positive breast cancer (337-339). It is generally used for high-risk ER-positive HER2-positive breast cancer (328, 338, 340). Since the success of T-DM1, more antibody-drug conjugates have been developed, with the most successful one being trastuzumab-deruxtecan, which has revolutionized the treatment of metastatic breast cancer (341-343). Because of its effect on tumors with no HER2-amplification but some IHC expression of HER2, the term HER2-low tumors were recently coined (341-343).

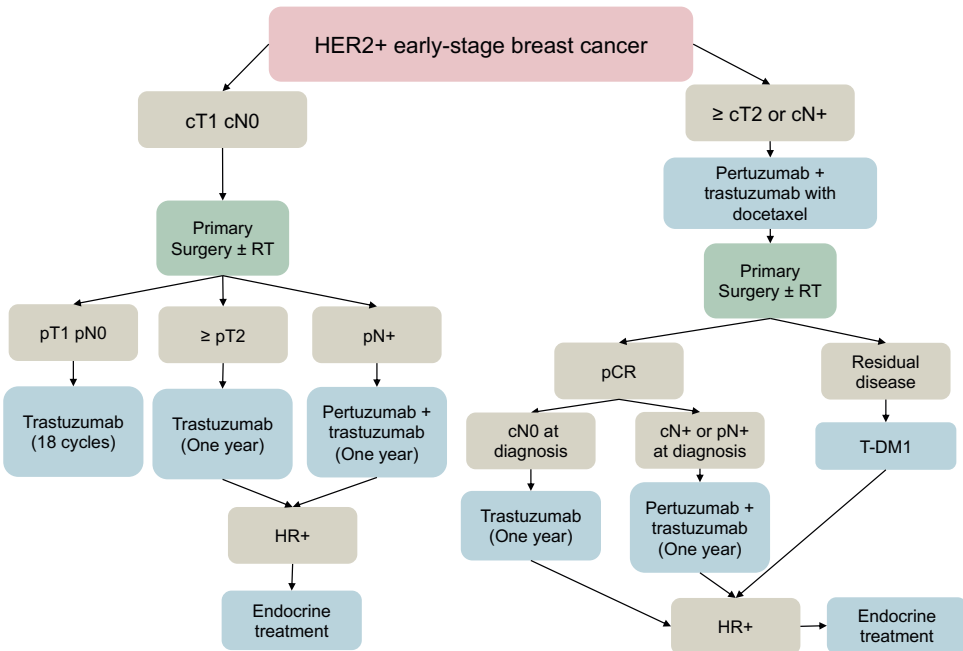


Figure 13.
A brief overview of the treatment for HER2-positive early breast cancer according to ESMO guidelines (196). Trastuzumab (with or without pertuzumab) is given in combination with chemotherapy. When the tumor is double positive for hormone receptors (ER and/or PR) and HER2, endocrine treatment is administred as described in Figure 11. HR: hormone receptor. RT: radiotherapy.

Poly (adenosine diphosphate–ribose) polymerase (PARP) Inhibitors

There has been considerable interest in poly (adenosine diphosphate–ribose) polymerase (PARP) inhibitors designed to disrupt PARP enzymes and eliminate tumor cells with homologous recombination repair deficiency (344). Breast cancers in individuals carrying germline *BRCA1/2* pathogenic variants are prone to exhibit homologous recombination repair deficiency (133, 134, 345, 346). The Olympia trial demonstrated that adjuvant olaparib in patients with pathogenic germline *BRCA1/2* variants resulted in improved disease-free and overall survival for cases of ER-positive/HER2-negative breast cancer and TNBC that are considered high-risk (347, 348).

Immune checkpoint inhibitors

A new development in the treatment of TNBC is the introduction of immune checkpoint inhibitors (ICIs) that target programmed death 1 (PD-1) or programmed death ligand 1 (PD-L1), which helps prevent the immune system from attacking (349). When PD-L1 engages with PD-1, the T-cell function is inhibited, which is co-opted by tumor cells to avoid destruction by the immune system (349). ICIs are monoclonal antibodies that block this PD-L1/PD-1 complex from forming enhancing T-cell-mediated killing (349). In stage II–III TNBC, the combination of ICIs (pembrolizumab or atezolizumab with standard chemotherapy) in the neoadjuvant setting improves pCR rates and event-free survival (350-353). Pembrolizumab in an adjuvant setting also improves event-free survival regardless of pCR status (351). The use of ICIs has now been incorporated into clinical practice internationally and in Sweden (145, 191).

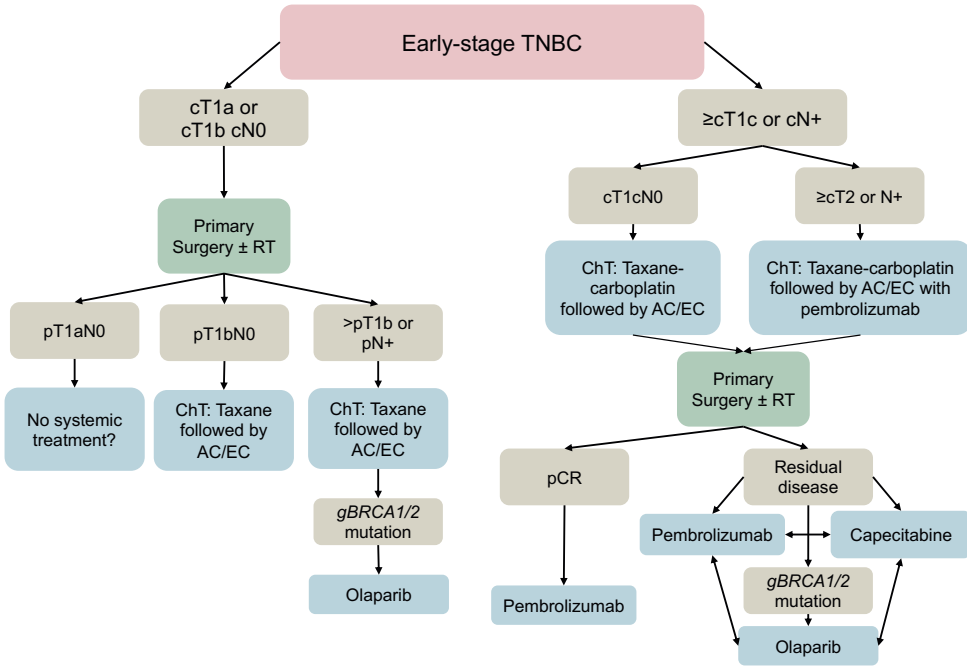


Figure 14.
A brief overview of TNBC treatment according to ESMO guidelines (196). ChT: chemotherapy. *g*: germline. AC/EC: doxorubicin-cyclophosphamide /epirubicin-cyclophosphamide. Currently, there is a lack of data directly comparing the efficacy of pembrolizumab, capecitabine, and olaparib in the TNBC setting for residual disease following neoadjuvant chemotherapy. Hence there is some degree of interchangability of treatments. G: germline

Despite advances in the treatment and care of breast cancer, there are a significant number of patients who still experience relapse despite optimal treatment (6, 73). This underscores the ongoing need for refining prognostication and treatment-prediction. Consequently, there is large interest in exploring new biomarkers for breast cancer (6). This thesis specifically focuses on two such biomarkers: caveolin-1 (CAV1) and insulin-like growth factor binding protein 7 (IGFBP7).

Caveolin-1 (CAV1)

CAV1 is a small oligomeric scaffolding protein that plays pivotal biological roles (354). Its significance lies in its strict requirement for the formation of caveolae, which are essential membrane structures found in virtually all tissues (354, 355). These caveolae serve as integral components for organizing signaling modules and regulating membrane internalization, making CAV1 a master regulator of cell signaling (354, 355). Beyond its structural role, CAV1 exhibits a remarkable versatility by binding to a myriad of proteins, thereby orchestrating a wide array of

cellular functions (354). Notably, it influences cholesterol homeostasis, endocytosis, receptor internalization, lipid accumulation, intracellular signaling, and proliferation (354-356). The regulatory mechanisms governing CAV1 protein levels are complex and multifaceted (354).

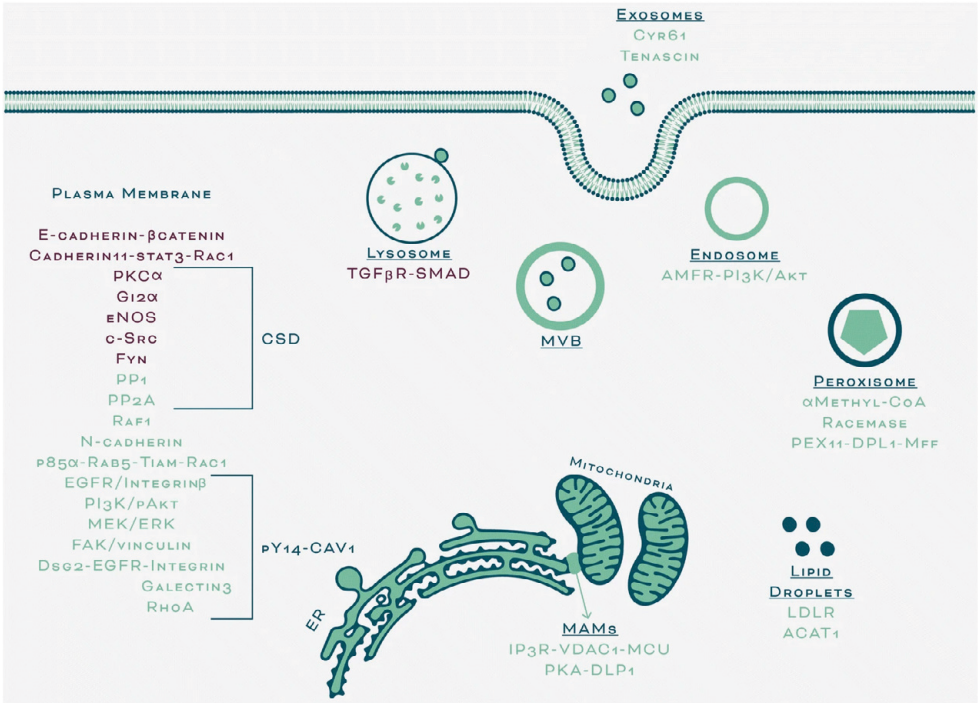


Figure 15.
Signaling events modulated by CAV1 at different cellular localizations. Proteins regulated by CAV1 functioning as a tumor suppressor are indicated in purple. Alternatively, those proteins modulated by CAV1 functioning as a tumor promoter are shown in green. Illustration from Simon et al. (354) © 2020 Simon et al. Open access.

Investigation of CAV1 expression in early-stage colorectal cancer has revealed a downregulation of protein levels without a corresponding reduction in mRNA levels (357). This observation hints at the involvement of post-transcriptional regulation (357). Epigenetic regulation is emerging as a pivotal factor that influences CAV1 expression in cancer (357). Hypermethylation of the CAV1 promoter is notably implicated and results in reduced protein levels in breast and prostate cancers (357). This epigenetic modulation extends to other malignancies like gastric adenocarcinoma, where CpG island hypermethylation correlates with decreased mRNA and protein levels and impacts patient survival (354, 358). Addressing these epigenetic changes, DNA-hypomethylating agents like 5-aza-2′-deoxycytidine have proven effective in restoring CAV1 expression across various cancer types (354).

The modulation of signaling events mediated by CAV1 holds significance in cancer (354). Notably, there is substantial interaction between CAV1 and Rho GTPases that promotes metastatic development through the stimulation of Src kinase-dependent activation of multiple pathways (359). The elevated expression of CAV1 correlates with improved cell survival, anchorage-independent growth, and activation of processes such as EMT, invasion, and resistance to anti-neoplastic drugs (360-362).

Numerous studies have demonstrated the crucial role of CAV1 in metabolic processes (363). For instance, *Cav1* null mice exhibit notable metabolic alterations and mitochondrial dysfunction in white adipose tissue coupled with compensatory gluconeogenesis (364). The indispensability of CAV1 for mitochondrial functionality in normal cells is evident, as shown by CAV1 knockdown leading to inhibited mitochondrial respiration and adenosine triphosphate (ATP) production owing to impaired cardiolipin biosynthesis (363). This disruption results in elevated expression levels of p53 and p21, which culminates in premature senescence (365). Conversely, cancer cells displaying mitochondrially localized CAV1 exhibit enhanced resistance to stress, more stable mitochondrial membrane potential, and increased mitochondrial biogenesis, which contributes to heightened cell survival (366). Moreover, CAV1 actively participates in cholesterol transport between the plasma membrane and the Golgi apparatus, and it facilitates the transport of Golgi-resident proteins from the cell surface back to the cell interior (354).

Additionally, CAV1 plays a role in forming and stabilizing lipid droplets, which are essential for the storage of neutral lipids (354, 363). These lipid droplets serve as major regulators of lipid metabolism, transport, and signaling (367). Regarding lipid droplet biogenesis, CAV1 promotes lipid and protein accumulation in the Golgi apparatus before their entry into these organelles (354, 363). Metastatic breast cancer cells express elevated levels of CAV1, low-density lipoprotein (LDL) receptors, and acetyl coenzyme A cholesterol acyltransferase 1 (ACAT1) enzymes, which facilitates their incorporation into LDL particles and promotes proliferation (368). Moreover, the migration potential of MDA-MB-231 cells depends on ACAT1 and correlates with increased lipid accumulation (369). Treatment with ACAT1 inhibitors not only reduces LDL receptor expression and LDL-enhanced proliferation, but also downregulates the CAV1/MAPK pathway (370). Hence, ACAT1 enhances the tumor-promoting function of CAV1 by favoring LDL uptake, contributing to the formation and stabilization of lipid droplets, and sustaining tumor-cell proliferation (354). A study of castration-resistant prostate cancer also highlighted the adjuvant effects of simvastatin, a statin that blocks cholesterol biosynthesis and regulates the expression of CAV1 leading to delayed progression (371). Lovastatin in combination with non-steroidal anti-inflammatory drugs reduces the expression and membrane localization of CAV1 (372). This results in the inhibition of CAV1-dependent cell-survival signals mediated by AKT

activation, along with other downstream signaling effectors, like signal transducer and activator of transcription 3 (STAT3) and MAPKs (372).

Elevated CAV1 expression, notably in pancreatic cancer, renders tumor cells highly responsive to conjugated or albumin-bound chemotherapeutic drugs (373). Higher CAV1 expression is also linked to taxane resistance in both preclinical and clinical studies (374, 375). A translational subtstudy within the GeparSepto trial recently reported that CAV1 expression predicted a worse response to paclitaxel and poorer clinical outcomes in these patients (376). Notably, CAV1 protein expression in stromal cells has been identified as a potential prognostic biomarker in breast cancer (377-380). Other studies found that sensitivity to treatment with trastuzumab and T-DM1 relied heavily on the vesicle-transport properties of tumor cells (381-383). Specifically, breast cancer cells expressing moderate CAV1 levels are at least five times more sensitive than CAV1-lacking cells (381, 382). CAV1's role in trastuzumab internalization via endocytosis was validated, and hypoxia-induced CAV1 redistribution hindered trastuzumab internalization and promoted resistance to T-DM1 treatment (384, 385). These findings suggest that CAV1 could serve as an effective prognostic marker for the outcomes of T-DM1-treated patients.

CAV1 is also involved in modulating glycolytic activities that are crucial for tumor survival (363). Elevated CAV1 levels promote glucose uptake and ATP production by stimulating glucose transporter 3 (GLUT3) transcription (386). Conversely, CAV1 knockdown reduces glucose uptake and lactate output, which is indicative of Warburg-effect suppression (386). Additionally, CAV1 interacts with the insulin receptor (InsR) and IGF-1R and activates AKT signaling, which enhances glucose uptake and lactate output (386). The correlation between CAV1 and growth factor response supports its role in tumor metabolism modulation, potentially through the InsR/IGF-1R pathway, thereby enhancing survival (386, 387). CAV1 appears to regulate the balance between the glucose-dependent mitochondrial respiration, aerobic glycolysis, and lipid-dependent energy metabolism that are crucial for tumor survival, likely via IGF-1R (363).

Insulin-like growth factor binding protein 7 (IGFBP7)

Another significant player in the InsR/IGF-1R pathway is IGFBP7, a 27-kD protein that is predominantly expressed in the vasculature, where its influence is particularly pronounced (114, 388). Initially identified in senescent mammary and meningeal cells, this protein shares around 20–25% homology with other members of the IGFBP family (IGFBPs 1–6) (116, 388, 389). Early studies demonstrated that IGFBP7 has the capability to bind to IGFs with reduced affinity compared to other IGFBPs (114). In cancer research, IGFBP7 has primarily been studied in connection with IGF-1R signaling (388), which is clearly implicated in breast cancer growth, proliferation, and survival. Its investigation in other biological aspects of cancer is comparatively limited (104, 105).

Notably, IGFBP7 stands out from its counterparts by its ability to bind to both IGF-1R and InsR (390) (as well as IGF-1 and IGF-2) (390). The binding of IGFBP7 to InsR hinders its activation and leads to the downregulation of mitogenic pathways (388). IGFBP7 has a distinct property compared to other IGFBPs of binding insulin with higher affinity than IGFs (390, 391). IGFBP7 binding reduces the activation and internalization of IGF-1R in response to IGF-1/2 while concurrently heightening the sensitivity of IGF-1R to insulin stimulation (390, 391). IGFBP7 has demonstrated the ability to extend the surface retention of IGF-1R during insulin/IGF1 stimulation, leading to prolonged IGF-1R signaling in leukemia (390, 391). Additionally, evidence indicates that IGFBP7 contributes to the sustained presence of IGF-1R on the cell surface, thereby extending insulin/IGF1 stimulation and amplifying AKT activation, which imparts mitogenic and pro-survival effects (390, 391).

Several studies have tried to delineate the role of IGFBP7 outside the context of IGF-1R signaling (388). IGFBP7 has the capability to interact with activin A, thereby affecting the growth-suppressing effects mediated by the transforming growth factor beta (TGF-β) superfamily (388, 392). Additionally, IGFBP7 binds to heparan sulfate on the cell surface, although the specific biological consequences of this interaction are unclear (393). The binding of IGFBP7 to type IV collagen was identified through observations of co-localization in the vascular basement membrane (394). This interaction was further confirmed by directly measuring the attachment of radiolabeled IGFBP7 to extracellular matrix proteins (394). It was demonstrated that the expression of IGFBP7 in tumor-associated endothelium is significantly higher than in healthy endothelial cells (395). However, despite these significant findings, many aspects of IGFBP7's functions in cancer are unclear and warrant further research to explore the intricacies of its role in cancer biology.

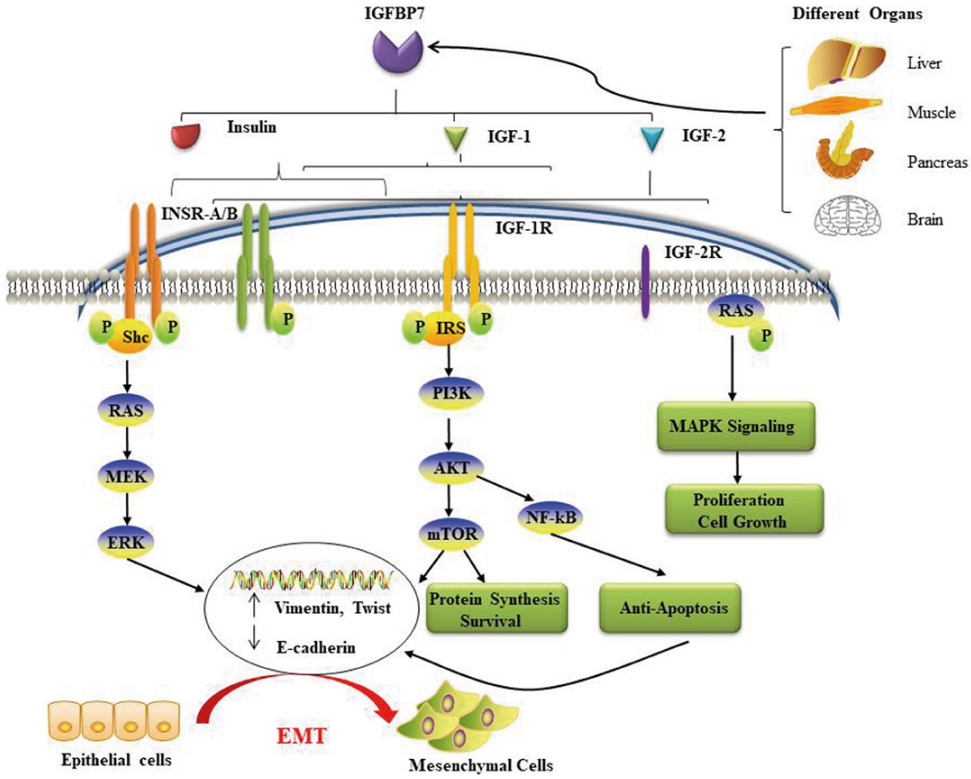


Figure 16.
Schematic overview of IGFBP7 and its role in the IGF/Insulin pathway. Illustration from Jin et al. (388).

Earlier studies have indicated that tumor IGFBP7 expression and circulating levels serve as prognostic factors (396-400). The prognostic relevance of circulating IGFBP7 levels is contingent upon the membrane status of tumor IGF-1R (396). Elevated circulating IGFBP7 levels have been associated with an increased risk of liver cancer (401). However, the molecular mechanisms underlying this observation are unclear. Nevertheless, it is indicated that IGFBP7 serves as an important biomarker in cancer. There are still knowledge gaps, and its role in cancer (including breast cancer) has not been extensively characterized.

The role of IGFBP7 is more thoroughly understood in cardiovascular disease than in cancer and is crucial in the development and progression of heart failure (402-404). IGFBP7 stands out as one of the most reliable biomarkers for heart failure identified to date and predicts cardiovascular events, myocardial infarction, and all-cause mortality (404-407). In the field of cardiology, IGFBP7 is regarded as a marker of senescence (374), which is a previously mentioned hallmark of cancer (20). Tissue senescence is characterized by permanent cell-cycle arrest coupled with

the loss of cellular homeostatic mechanisms that maintain tissue renewal (408, 409). Recent research has uncovered a group of molecules released by senescent cells known as the senescence-associated secretome (403, 404). These molecules exert autocrine, paracrine, and endocrine effects that collectively result in cell cycle arrest (403, 404, 408, 409). IGFBP7 is recognized as a constituent of the senescence secretome (404). In conditions marked by cellular injury and aberrant growth, IGFBP7 impedes cell proliferation by inducing G1-phase cell-cycle arrest, which diminishes the probability of propagating maladaptive cellular disruptions (408-411). Significantly, obesity contributes to senescence, and the expression of IGFBP7 rises proportionally with increases in body mass index, likely serving as a compensatory mechanism (412). Insulin resistance correlates with an elevated serum concentration of IGFBP7 (412). Moreover, IGFBP7 is recognized as one of the most reliable indicators of the effectiveness of sodium-glucose co-transporter-2 (SGLT2) inhibitors, which are medications known for their extensive cardiorenal and metabolic protective properties (413-415). IGFBP7 is also deemed a robust predictor of acute kidney injury and is generally a marker of renal disease (413-415).

Aims

Overall

- To investigate and characterize two new potential prognostic and/or treatment-predictive biomarkers, CAV1 and IGFBP7, in relation to metabolism, angiogenesis, and the tumor microenvironment on a genomic, transcriptomic, and proteomic level in breast cancer

Specific

Paper I

- To investigate the role of tumor-specific expression levels of CAV1 protein in different spatial localizations and *CAV1* gene expression in relation with clinicopathological factors, signaling pathways, and prognosis in breast cancer

Paper II

- To study whether *CAV1* polymorphisms could predict locoregional recurrent and/or contralateral breast cancer and whether tumor-specific CAV1 modifies the potential associations

Paper III

- To characterize *CAV1* gene expression in TNBC with regard to clinicopathological factors, molecular features, tumor microenvironment, and prognosis

Paper IV

- To investigate whether IGFBP7 protein and gene expression are associated with clinicopathological factors, Insulin/IGF signaling, and prognosis in breast cancer

Paper V

- To study whether *IGFBP7* gene expression could predict the efficacy of the IGF-1R targeting agent ganitumab and prognosis in breast cancer

Methods and methodological considerations

"A model is a lie that helps you see the truth"

— Siddhartha Mukherjee

Study design

In clinical research, there is a hierarchy of taxonomy and based on the underlying study types (416). This framework is generally applicable to clinical research, while translational research exhibits more variability due to factors such as the type of assays and analyses conducted. Clinical research can be categorized as observational and interventional studies. Also known as (clinical) epidemiology studies, observational studies involve investigators observing and describing patterns, associations, or characteristics without intervening (416). Descriptive studies simply describe observations, while analytical studies explore relationships between characteristics and make comparisons between groups (416). Most studies in this thesis (I-IV) utilize a cohort study design, which is a type of observational study design where a group with a common characteristic (e.g., breast cancer) is followed over time to study various outcomes (416). A translational or molecular epidemiological study is a specific type of observational study that investigates molecular markers in disease development and outcome and makes comparisons based on these molecular markers.

Interventional studies assess the effects of medical interventions or treatments on participants with or without a comparison group. RCTs randomly assign interventions between control and treatment groups and are considered the gold standard in medical research (416). The evidence hierarchy supports RCTs as providing the best evidence and having the most potential to influence clinical practice (416). These are followed by non-randomized controlled trials, observational studies, and lastly case reports. Although RCTs are considered the gold standard for scientific research, there are instances where it would be unethical to perform them (416). This ethical limitation arises when certain factors, such as smoking, cannot be subjected to randomization due to the potential associated

harms, or when randomization is simply not possible, as with molecular markers. Clinical trials are categorized into phases based on their objectives: Phase I (safety and dosage), Phase II (efficacy and side-effects), Phase III (large-scale efficacy, determining superiority or non-inferiority, and safety), and Phase IV (post-marketing surveillance) (416). The last paper (V) in this thesis makes use of data from a Phase II clinical trial (324).

Validity and reliability

Validity is a crucial concept in clinical research and can be either internal or external. Internal validity refers to the study's ability to measure what it intends to measure (416). Essentially, it represents "the degree to which a study is free from bias or systematic errors" (416). The next section delves into a more detailed discussion of bias and systematic errors. Generally, internal validity is deemed an indispensable condition, a *conditio sine qua non*. Without internal validity, external validity cannot be achieved, rendering the results and conclusions meaningless.

External validity (or generalizability as it is sometimes called) refers to the extent to which the results of a study can be applied or generalized to populations that did not participate in the study. Generally, there is a tradeoff between high internal validity (as in RCTs) and lower generalizability due to the more stringent patient selection and strict procedural requirements. While external validity is crucial for the potential implementation and implications of the results, it does not inherently affect the results or scientific value of the study. However, the applicability of the study can be significantly lower. External validity is desirable for implementing results but should not come at the expense of a study's internal validity. Applying non-valid results to other populations is not only futile but potentially hazardous.

Reliability is a crucial concept in research and measurement and denotes the consistency, stability, or dependability of a measurement instrument or assay. It refers to the extent to which results obtained from a specific method or tool are consistent, reproducible, and free from random error, meaning one can repeatedly measure the same thing and still obtain consistent results. This is crucial not only for internal study validity, but also for the practical implementation of the results. Similarly, random errors, if truly random, result from chance alone, perhaps due to random measurement errors, leading to imprecision. Some imprecision can be tolerated and expected without affecting the internal validity of the study. Hence, confidence intervals are often reported as a measure of random variation. A challenge lies in determining whether the error is genuinely random.

Types of bias (systematic errors)

Bias in research refers to systematic errors that affect the design, conduct, or interpretation of a study and lead to distorted results. There are three main types of bias: selection bias, information bias, and confounding. Selection bias is a distortion of association due to a sample selection that does not accurately reflect the target population. Selection bias can occur when investigators use improper procedures for selecting a sample population, but it can also occur as a result of factors that influence continued participation of subjects in a study. Selection bias can result from the procedure used to select study participants when the selection probabilities are differential and not proportional between exposed and unexposed cases and controls from the target population. This can occur when exposure or related factors influence selection and results in spurious associations in the study. Selection bias occurs in cohort studies if the rates of participation or the rates of loss to follow-up differ by both exposure and outcome status. We can seldom know for sure the exposure and outcome status of nonrespondents, or persons lost to follow-up. Observational studies are subject to both types of selection bias, while clinical trials assign the intervention (or exposure) upon entry. This minimizes errors in selecting a sample population but is suspectable to factors to influencing continued participation in a study.

Information bias is another key factor and refers to errors in the measurement or classification of variables. Information bias can happen even if measured errors are equal between the compared groups and between those that do or do not experience the outcome of interest. There are different types of information bias, including measurement bias, which entails inaccuracies in data-collection methods or tools. Non-differential misclassification or measurement between comparators results in bias in the association or estimate towards null and differential misclassification, or measurement differences between comparators results in bias in the associations either towards or away from the null. Recall bias involves differences in the accuracy of recall or reporting of information between groups. Due to their health concerns, case groups may have greater incentive to recall past exposures than controls. Exposed persons in a cohort study may be concerned about their exposure and may over-report or more accurately report the occurrence of symptoms or the health outcome. Finally, observer bias refers to the observer's knowledge or expectations that influence the outcome assessment, which is commonly handled by blinding.

Confounding (or blurring of effects) is a common type of bias and poses a significant challenge in various observational studies (and sometimes clinical trials). It occurs when an external third factor called a confounder distorts the relationship between an exposure and an outcome. Confounding can occur when the identified exposure is associated with both the outcome of interest and an additional factor that may independently affect the outcome. The exposure of interest is seldom the sole

differing factor between exposed and unexposed groups, and other factors influence both the health outcome and exposure. As a result, confounding is frequent in observational studies.

In RCTs, the intervention is randomly allocated, so it is the only factor that differs between the compared groups, thus minimizing confounding. Observational studies, however, often require strategies to control for potential confounders, which most commonly involve multivariable analysis. Multivariable analysis can be used to investigate the relationship between multiple independent variables (exposures) and a dependent variable (outcome) simultaneously. A multivariable analysis is conducted to adjust for potential confounders. The model examines whether the estimate or association between an independent variable (exposure) and outcome remains consistent across strata of other independent variables. The exposure is considered independent by the statistical model if this consistency is observed. Nevertheless, it is crucial to strike a balance and avoid "over-adjustment" by incorporating an excessive number of variables into multivariable models, especially in the presence of sparse data. Over-adjustment can lead to imprecision, introduce bias, and drive associations toward the null hypothesis.

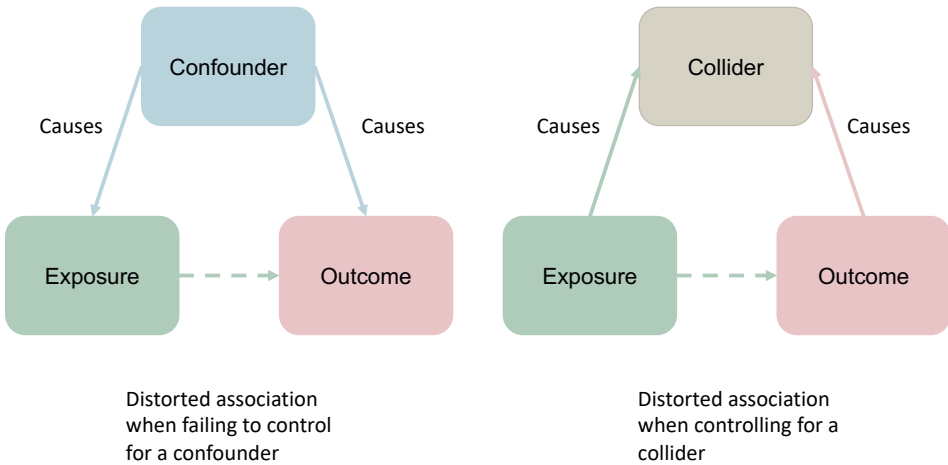


Figure 17.
Schematic overview of the difference between confounders and colliders

Spurious associations can also arise from the inclusion of colliders, which are influenced by two other variables in the model. Careful consideration of which variables to include in the analysis is essential to ensure accurate and meaningful results. Generally, the studies outlined in this thesis (papers I–V) all included confounders that were selected *a priori* based on established associations with the outcome. These confounders typically encompassed clinicopathological factors and treatments. The aim was to explore whether the examined markers provided

supplementary information beyond established prognostic and/or treatment-predictive factors.

To evaluate potential causality, the Bradford Hill criteria were developed with consideration of temporality, strength of association, consistency, biological gradient, specificity, biological plausibility, coherence, experimental evidence, and analogy (417). In the context of this thesis, which focuses on molecular markers, determining causality is challenging. The emphasis lies more on identifying consistent markers for prognosis and treatment prediction, where the critical role of a molecular marker in a process may not be necessary. For instance, Ki67 is a marker associated with cell cycle regulation and is upregulated without being essential for initiation or completion of the process.

Cohorts

Breast cancer and blood cohort

Papers I, II, and IV are primarily based on data from the Breast Cancer and Blood (BC-Blood) cohort, which is a prospective, population-based cohort initiated at Skåne University Hospital in Lund in 2002. The cohort's participant-inclusion phase was completed in 2019, and patient follow-up is ongoing. The primary purpose of this cohort is to explore the interplay of genetic and lifestyle factors in relation to prognosis in breast cancer. Subsequently, the cohort expanded its scope to include the secondary goal of identifying novel tumor markers.

Eligibility for participation in the BC-Blood cohort was restricted to female breast cancer patients with no prior history of breast cancer diagnosis and no other cancer diagnosis within the preceding 10 years. Recruitment occurred during a preoperative visit, where all participants provided written informed consent. During the visit, participants completed a comprehensive three-page questionnaire covering reproductive history, lifestyle factors (such as smoking and alcohol use), medication use in the past week, and exogenous hormone use. Additionally, research nurses took anthropometric measurements, including height, weight, breast volume, waist circumference, and hip circumference. Blood samples were also collected at this stage. The blood samples underwent centrifugation and were promptly frozen to −70°C within a 2-hour timeframe to preserve them for future use.

The cohort's follow-up structure involved four to five physical study visits at intervals of three to six months, (seven to nine months only for patients receiving both radiotherapy and chemotherapy), one year, two years, and three years postoperatively. During these visits, participants answered a one-page questionnaire (a short version of the preoperative questionnaire), and research nurses took

anthropometric measurements and blood samples. Subsequently, follow-up questionnaires were sent to participants biannually via mail.

Data regarding clinicopathological characteristics (including histological grade, histological type, pT category, pN category, ER, and PR), breast cancer events (such as distant metastasis), and death were extracted from patient charts and pathology reports. Routine HER2 assessment commenced in November 2005. A retrospective analysis of HER2 was conducted using a dual gene-protein assay on tumor tissue microarrays (TMAs) for patients included from 2002–2012 to establish HER2 status for cases with missing data (418). Similarly, Ki67 was routinely assessed only from March 2009, and due to its heterogeneous expression, it was deemed unsuitable for assessment on TMA. Consequently, Ki67 was not incorporated into any of the clinical prediction models.

Adjuvant breast cancer treatments were prescribed based on the discretion of the treating physician(s). Data on treatments were gathered from both follow-up questionnaires and patient charts. Only treatments administered before the occurrence of the first breast cancer event were considered adjuvant and were duly recorded. Otherwise, all treatments received before the last follow-up or death were documented.

For patients included between 2002 and 2012 (n = 1018), tumor TMAs were constructed for a total of 984 patients. This subcohort formed the foundation for papers I, II, and IV. Genotyping was conducted for patients included from 2002–2016, and all but one of the patients included between 2002 and 2012 had available genotype information.

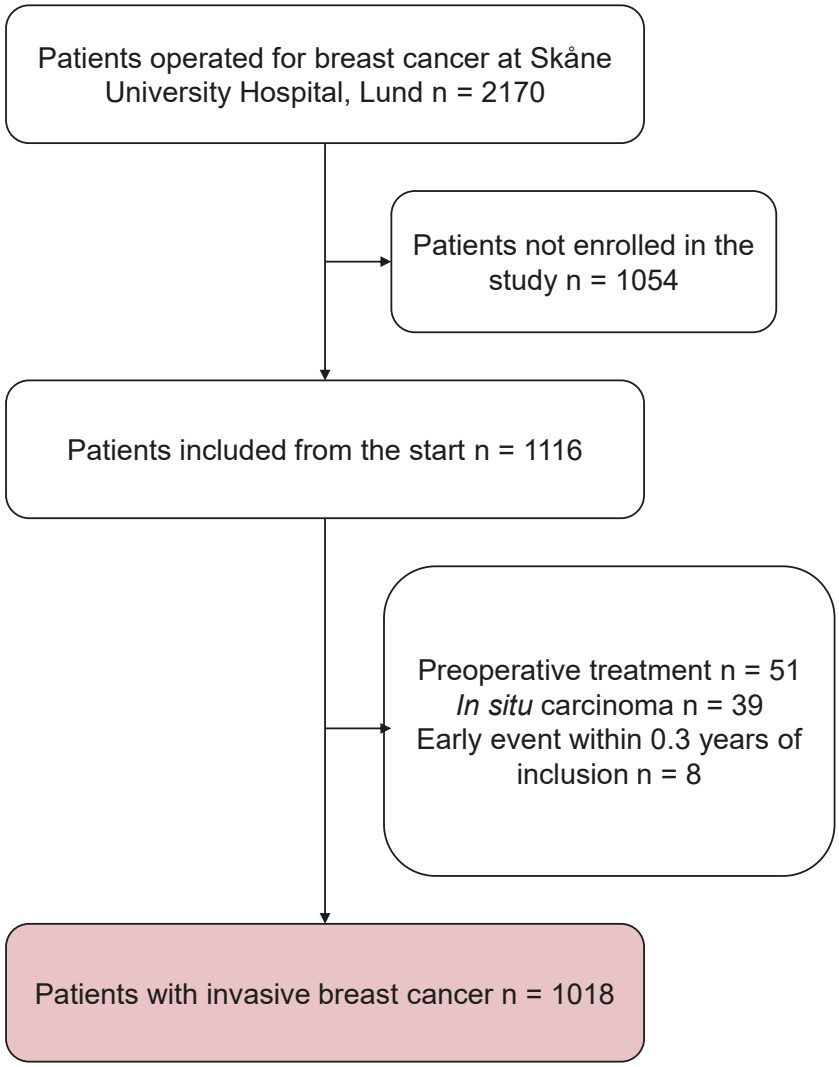


Figure 18.
Flowchart of included an excluded patients in the BC-Blood cohort that forms the basis (the red box) of paper I, II, and IV.

SCAN-B

The Sweden Cancerome Analysis Network – Breast (SCAN-B) study (ClinicalTrials.gov ID NCT02306096) is a cohort study of consecutively enrolled breast cancer patients from seven hospitals in the healthcare region of Southern Sweden, along with two additional Swedish hospitals (Jönköping and Uppsala) (419, 420). The primary goals of the SCAN-B study are to develop, validate, and integrate molecular tumor markers that have clinical utility into routine healthcare (420).

All newly diagnosed breast cancer patients were invited to participate, and enrollment was integrated into the clinical routine (419-421). The eligibility criteria encompassed a preoperative diagnosis of primary invasive breast cancer, and as of autumn 2012, patients with a preoperative suspicion for breast cancer or those undergoing neoadjuvant therapy (419-421). During routine preoperative/pre-biopsy blood work, additional blood samples were collected and stored (419-421). For patients proceeding directly to surgery, fresh tumor-cell-enriched specimens were extracted from the surgical sample after routine assessment by a pathologist and later preserved in RNAlater (419-421). Study samples were only taken from tumors if they did not interfere with routine clinical diagnostics (419-421). Patients undergoing preoperative biopsy had additional study biopsies taken and preserved in RNAlater, and samples were sent to the central research laboratory of SCAN-B in Lund for various molecular assays (419-421).

Clinicopathological and follow-up data along with information on adjuvant medical treatment were sourced from the Swedish National Breast Cancer Quality Register (NKBC) (421). Local pathology, surgery, and oncology departments report clinicopathological data, treatments, and outcomes to the NKBC. Data on adjuvant treatment for the SCAN-B cohort include endocrine treatment, chemotherapy, and HER2-targeted treatment (419, 421). The treatments are at the discretion of the treating physicians (419).

For papers III and V, patient enrollment occurred between September 1, 2010, and May 31, 2018, with follow-up extending until November 2021 (421). A subcohort of TNBC patients from SCAN-B was also utilized and consisted of patients diagnosed with TNBC between 2010 and 2015 at a Region Skåne Hospital (345). TMAs were constructed for these patients from obtained tumor tissue (345). Exclusion criteria for this cohort included inconsistencies in TNBC status after clinical chart review, insufficient tumor material, failure of quality filters for massively parallel sequencing of ribonucleic acid (RNA-seq), or unavailability of formalin-fixed paraffin-embedded (FFPE) tissue for TMA analysis (345).

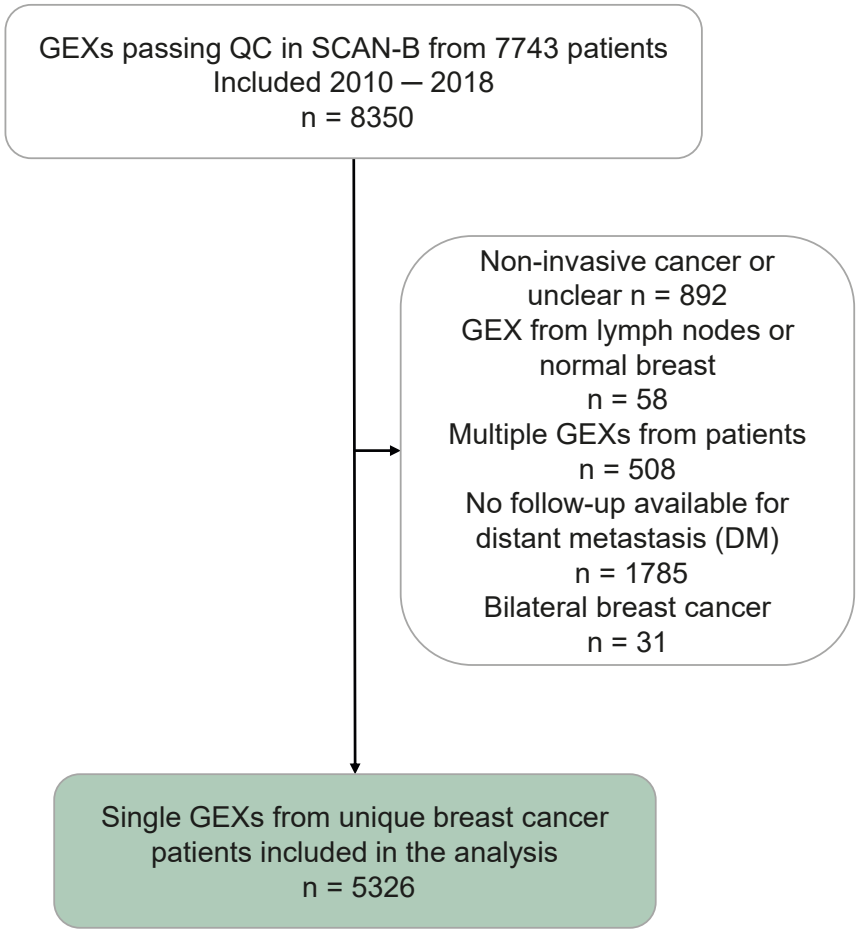


Figure 19.
Flowchart of included and excluded patients and GEXs that form the basis (the green box) for paper III and V

I-SPY2

“Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2” (I-SPY2) is an ongoing, open-label, adaptively randomized phase II multicenter trial focusing on neoadjuvant therapy for early-stage breast cancer with a high risk of recurrence (NCT01042379) (324, 339). It is a platform trial that is concurrently assessing multiple investigational arms (324, 339). Each arm comprises an investigational agent or combination added to a standard-of-care neoadjuvant chemotherapy backbone, which also functions as a shared control arm (324, 339). The randomization allocates patients with a preference for arms based on continuously updated Bayesian probabilities of pCR rates for each subtype, and 20% of patients are randomly assigned to the control arm.

In I-SPY2, an arm is deemed successful when it achieves the predefined efficacy threshold of an 85% probability of success in a hypothetical, subtype-specific 300-patient, 1:1 confirmatory phase III trial. An arm is discontinued for futility if the predicted probability of success in phase III is below 10% (324, 339). The maximum accrual for an agent across all subtypes is predefined at 120 patients. Subtypes are based on hormone receptor status, HER2 status, and MammaPrint® high-risk status (MP1 or MP2) (324, 339, 422).

The primary endpoint is the pCR evaluated at the time of surgery and defined as the absence of invasive disease in the breast and regional nodes (ypT0/is and ypN0) (324, 339, 350, 423-425). The primary analysis follows a modified intent-to-treat approach and includes all participants who receive the allocated therapy for evaluation (324, 339). Participants were categorized as having "non-pCR" status if they switched to non-protocol-assigned therapy, forwent surgery, or withdrew from the trial (324, 339).

Women were considered eligible for participation in the I-SPY2 trial if they were aged 18 years or older and diagnosed with clinical stage II or III breast cancer, a tumor diameter of at least 2.5 cm by clinical examination, and a minimum of 2 cm as assessed by imaging (324, 339). Exclusion criteria were an Eastern Cooperative Oncology Group performance status score exceeding one and a history of prior chemotherapy for this cancer (324, 339). Additionally, patients with hormone receptor-positive tumors and low-risk MammaPrint® scores were excluded due to the limited benefit from systemic chemotherapy (324, 339).

All participants were administered weekly intravenous paclitaxel (12 doses of 80 mg per m^2 of BSA) either alone in the control arm or in combination with the designated experimental regimen in the experiment arm (324, 339). This was followed by four doses of intravenous doxorubicin (60 mg per m^2 of BSA) and cyclophosphamide (600 mg per m^2 of BSA) every two to three weeks (324, 339). Patients with HER2-positive cancer also received trastuzumab for the initial 12 weeks with a loading dose of 8 mg per kilogram of body weight in week one,

followed by a maintenance dose of 6 mg/kg every three weeks in weeks four, seven, and ten (324, 339). After FDA approval, pertuzumab was incorporated into standard therapy for HER2-positive patients and involves a loading dose of 840 mg in week one, followed by a maintenance dose of 420 mg every three weeks in weeks four, seven, and ten (324, 339). Following neoadjuvant chemotherapy, definitive surgery (either lumpectomy or mastectomy) was performed based on the treating surgeon's discretion with sentinel or axillary node dissection according to the National Comprehensive Cancer Network and local guidelines (324, 339). Radiotherapy followed national and local guidelines, and adjuvant treatment, although not mandated by the study, was recommended per the National Comprehensive Cancer Network guidelines at the oncologist's discretion (324, 339).

Core needle biopsies were obtained from primary breast tumors prior to treatment (422). A 5-µm section was stained with hematoxylin and eosin (H&E), and pathologic evaluation was carried out to ensure the tissue contained a minimum of 30% tumor (422). If the tissue sample met this criterion for 30% or greater tumor cellularity, it was centrally sectioned at the I-SPY 2 laboratory to generate 10 to 30 5-µm sections for microarray profiling (422). These sections were processed at Agendia, for RNA extraction and gene expression profiling using Agilent 44K microarrays (422). For each array, the green-channel mean signal underwent log2 transformation and was centered within the array to its 75th quantile according to the manufacturer's data-processing recommendations (422). To avoid negative values, a fixed value of 9.5 was added. Probeset level data per array were mean collapsed to the gene level, and common genes across the two platforms were identified (422). The expression data from over 900 I-SPY2 patients were combined into a single gene-level dataset after batch adjustment using ComBat, resulting in a normalized expression dataset comprising 987 patients ×19,134 genes (422).

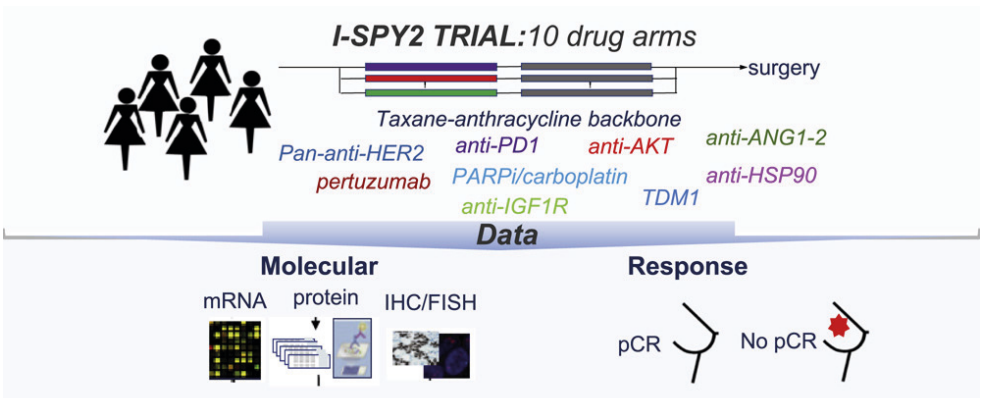


Figure 20.
Schematic overview of the I-SPY2 trial. Illustration from Wolf et al. (422). © 2022 Wolf et al. Open Access.

Publicly available databases

Beyond the most well-known and annotated public breast cancer cohorts and trials, there are a plethora of public databases containing genomic, transcriptomic, and clinical data from breast cancers. It is custom to have gene expression data in a public repository like the Gene Expression Omnibus (GEO) database, as well as other databases. This allows researchers to freely access any data that they would like for their research. For this thesis, two primary datasets were used: the Gene expression-based Outcome for Breast cancer Online (GOBO) database and the GSE31519 database, which are both compiled from a wide variety of breast cancers (426, 427). GOBO comprises 1881 tumors, from which gene expressions were profiled on Affymetrix microarrays with available follow-up for survival analysis, which can be stratified by molecular subtype (426). The GOBO platform is a versatile and user-friendly online tool designed for conducting various analyses, and its functionalities include rapid evaluation of gene-expression levels in different subgroups of breast tumors and examining the association between gene-expression levels of individual genes and outcome (426).

The GSE31519 database consists of TNBC cases (n = 579) extracted, normalized, and complied from 28 other public gene-expression datasets to ensure a large collection of TNBCs (427). Details regarding pooling, quality control, and analysis pipeline are available elsewhere (427). There were 327 TNBCs with available follow-up (427). Both GOBO and GSE31519 allow for the investigation individual genes and gene signatures in relation to clinicopathological characteristics, molecular features, and prognosis. The main drawback is that the data are already set, and there is no ability to add or alter data to better fit the research purpose. There is also an inherent disadvantage in that researchers that were not involved in data collection are not familiar with the intricacies of the data and are not familiar with certain considerations, which can limit the interpretability of results to some extent. Therefore, public datasets tend to be used to independently confirm the findings of a project and perhaps sometimes to better characterize features of biomarker-defined subsets of breast cancer tumors. This thesis also made use of data from two well-known public databases, which are described below.

METABRIC

METABRIC is a case series of patients who were diagnosed with early-stage breast cancer between 1977 and 2005 and was derived from five tumor banks in the UK and Canada (212, 234, 235). Primary fresh-frozen breast cancer specimens were collected from the tumor banks for further molecular profiling (212, 234, 235). The study was an observational case series, and tumor samples were all excised before systemic therapy (212, 234, 235). Pathological data were obtained from the original histopathology reports. Expert breast cancer pathologists conducted assessments on

FFPE sections stained with H&E and evaluated the presence of invasive tumors, pre-malignant or benign alterations, tumor cellularity, and lymphocytic infiltration. Tumor cellularity was visually scored using a semiquantitative approach. When available, IHC scoring was used to determine the ER and HER2 status of tumors. In instances where IHC markers were unavailable, gene expression data were utilized if available to establish the ER and HER2 status (212, 234, 235). The data underwent manual curation and basic quality control procedures (234). Additionally, other clinical information, including follow-up data, were obtained from the five tumor banks (212, 234, 235).

RNA was isolated from 10 sections with thickness of 30 µm or from 10 sections with thickness of 8 µm and were quantified using a NanoDrop spectrophotometer. RNA quality was further assessed using an Agilent 2100 Bioanalyser Nanochip (212). Tumor samples with an RNA integrity number (RIN) > 7 were subjected to expression array hybridization (212). Total RNA was utilized to produce biotin-labeled cRNA with the Illumina Totalprep RNA amplification kit and then hybridized onto Illumina Human HT-12 v3 Expression Beadchips according to the manufacturer's instructions. Subsequently, the arrays were scanned on an Illumina BeadArray Reader (212).

Upon completion of scanning and the availability of raw data, each BeadChip underwent processing using a custom script developed by the METABRIC consortium (212). To ensure data quality, potential outlier arrays were identified using bead-level quality-control scores derived from control probes, and only the arrays passing quality control were retained for subsequent normalization procedures (212). ER-positive and ER-negative samples underwent separate quantile normalization, and the results were averaged to establish the final target distribution (212). Subsequently, each array underwent normalization to the target by quantile normalizing probes belonging to the target distribution. To eliminate batch effects associated with the array's position on the Illumina BeadChip, a linear model was fitted using the Limma-Voom package (212, 428).

TCGA

The Cancer Genome Atlas (TCGA) is a case series of biospecimens collected from newly diagnosed patients with invasive breast adenocarcinoma undergoing surgical resection and no prior treatment for their disease (chemotherapy or radiotherapy) (207). Approximately 1,100 tumors were collected from cancer centers worldwide. Clinical data were extracted from patient records and pathology reports, with no central review of biomarkers (207). The clinical calls from each supplying site were employed to classify clinicopathological characteristics and led to some inconsistencies, particularly in ER and PR status, possibly due to historical reasons and local guidelines (207). Only HER2 status was re-assessed (207). Following the American Society of Clinical Oncology/College of American Pathologists

guidelines, a consistent new clinical status call was made, supplemented with fluorescence ISH calls, and further supplemented with amplification calls predicted by the copy-number data (207). Adjustments were made only to cases previously labeled as equivocal or NA in the preceding steps (207). For AJCC stages, the previous staging was converted to align with AJCC Edition 7 whenever possible. Follow-up was conducted locally at each site (207).

The tumor sections needed to contain an average of 60% tumor-cell nuclei with less than 20% necrosis for inclusion in the study per the TCGA protocol (207). Each H&E-stained case was reviewed by a board-certified pathologist to confirm that the tumor specimen was histologically consistent with breast adenocarcinoma and that the adjacent normal specimen contained no tumor cells (207). RNA and DNA were extracted from tumor tissue and adjacent normal tissue specimens using a modification of the DNA/RNA AllPrep kit. RNA integrity was assessed via the Agilent RNA6000 nano assay, and only cases with RIN >7.0 were included. By the time of the data freeze, the TCGA consortium had received 1,377 breast adenocarcinoma cases, and 72% passed quality control (207). Various platforms were employed for molecular profiling and included gene expression microarrays, RNA-seq, DNA methylation arrays, miRNA sequencing, Affymetrix SNP arrays, exome sequencing, and reverse phase protein arrays (207).

Biomarker studies

A biomarker is a cellular, biochemical, and/or molecular characteristic (including genetic and epigenetic characteristics) that can be objectively measured and evaluated as an indicator or characteristic of normal biological processes, pathogenic processes, or responses to a therapeutic intervention (429). Biomarkers can be found in various biological materials, such as blood, urine, tissues (including tumor), or imaging data. They are meant to provide valuable information about physiological, pathological, or pharmacological states (429). Examples of biomarkers include proteins, genes, hormones, metabolites, or specific imaging features (429).

Research on biomarkers revolves around discovering new biomarkers, validating their association with diseases, and implementing them in clinical practice (429). Similar to clinical trials, biomarker research consists of several phases (429). The discovery phase involves exploring various data types, such as genomic, proteomic, or metabolomic data, to identify potential biomarkers. Following this, the clinical and analytical validation phase aims to confirm and validate biomarker associations through larger-scale studies with diverse patient populations. The analytical part of this phase consists of assessment of sensitivity, specificity, determination of cutoff values, and determination of reproducibility across different laboratories and

platforms. This includes development of standardized assays for biomarker measurement. The subsequent clinical-utility phase assesses the practical application of biomarkers in clinical settings. This is often achieved through clinical trials to evaluate the impact of biomarkers on patient outcomes. The trials include both prospective analyses, which are preferred, and retrospective analyses. Prospective cases series or cohorts of the intended clinical population can also be used. The cost-effectiveness and feasibility for routine clinical use are also evaluated. Finally, the qualification phase involves regulatory approval and the assessment of biomarker test performance in the population. While these phases are not as distinct as clinical-trial phases, certain studies may encompass multiple phases. This thesis primarily focuses on the discovery phase, with some exploration of the validation phase as well.

Tissue microarray and immunohistochemistry

IHC is a laboratory technique in which primary antibodies specific for biomolecules are applied to a tissue sample to visualize and localize antigens with the help of secondary antibodies (430). The antigen is often a protein located in one or more compartments of the cell (e.g., the membrane, the cytoplasm, or the nucleus). The antibody used to bind the antigen (the primary antibody) is usually of the IgG class and can be either monoclonal or polyclonal (430). Monoclonal antibodies bind only to one epitope of the antigen, making them specific, whereas polyclonal antibodies can bind to several epitopes, suggesting higher sensitivity for the antigen. (430) Small changes in the epitope can impair the binding ability of a monoclonal antibody, while the binding capacity of a polyclonal antibody is less affected by changes in one epitope. Monoclonal antibodies are produced in hybridomas, making the availability reliable once the hybrid cell line is in place (430). Polyclonal antibodies can differ over time since they are generated in different animals, and the availability of polyclonal antibodies depends on the size and lifespan of the animal used for its generation (430). The primary antibodies used in paper I, III, and IV were all polyclonal and well validated (385, 431). Primary antibody binding was then visualized using a secondary antibody conjugated to a peroxidase that catalyzes the production of a dark brown (or red) color from the substrate 3,3'-diaminobenzidine. IHC has been used for decades, but it was only in the late 1990s that the concept of TMAs was first described (432). A TMA is a collection of cylindrical tissue cores embedded in a single paraffin block (432). The possibility of simultaneous staining of several different samples increases the throughput of IHC and decreases the technical variation between staining of different samples (432).

The end result of this technique is an image of a stained TMA section. Although image analysis software that can quantify pixel intensity of scanned images, most are still evaluated manually. Potential sources of non-biological variation in an IHC

experiment can be technical, such as unspecific binding of antibodies, or incomplete blocking of endogenous peroxidase activity, or they can be related to evaluation bias (430). The most common situation is that the extent and intensity of the staining are evaluated on a semi-quantitative scale, where cutoff points are determined for each cohort (433). Within the experiment, a higher value represents a stronger staining. The data can then be analyzed much in the same way as a microarray experiment, only with fewer and discrete variables. A significant consideration for TMA analysis is the representativeness of the cores in relation to whole-section slides (430, 433). The cores are extracted from representative non-necrotic tumor regions, and TMAs are deemed suitable for assessment of markers with more homogeneous expression, but they may not accurately reflect markers with heterogeneous expression, such as Ki67 (430, 433). To enhance the comprehensiveness of this thesis, it would be advantageous to acquire whole-section slides for a select group of tumors in addition to TMA samples for certain studies. By comparing marker scoring between these two sample types, the level of agreement, and the scoring of the markers on a TMA can be validated.

The evaluation is a subjective process, which is a major drawback (433). When evaluating IHC staining, it is important to be blinded to group data since knowledge such as the patient outcome can lead to unintentional bias (433). It is also important to be aware of the so-called “diagnostic drift” that may occur when one person evaluates a large cohort or when the evaluation is performed during an extended period of time, leading to a gradual change in the assessment of the IHC expression over time. In all papers, two persons have independently evaluated the IHC expression without knowledge of the clinical data (433). Differences in the scoring were discussed to reach consensus in order to minimize inter-observer discrepancies. A senior investigator was consulted when consensus was not reached. The advent of more multiplex IHC techniques allows for staining of a tissue using several antibodies at once allows for more complex subtyping of cells within the tissue, as well as potential expression of a certain protein only in a particular subset of tumor and stromal cells associated with specific clinical outcomes. A certain subset of tumor or stromal cells may also confer independent prognostic information.

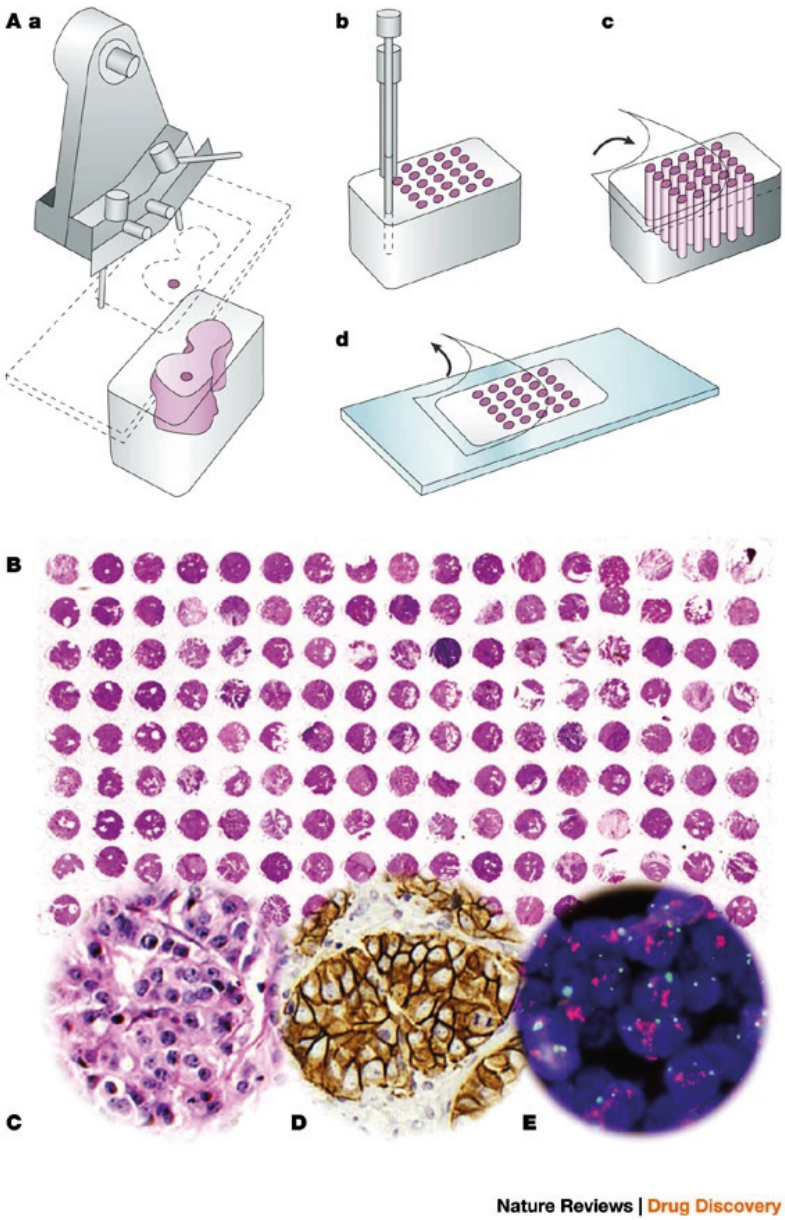


Figure 21.
The TMA method followed by different stainings (C) H&E, (D) IHC, and (E) immunofluroesence as decribed by Sauter et al. (434). © 2003 Nature Publishing Group. Reprinted with permission from Nature Publishing Group.

High-throughput approaches

Gene expression microarrays are used to simultaneously measure the relative levels of all the mRNA molecules in a sample. In short, the mRNA molecules are converted to cDNA and fluorescently labeled. The intensity detected at each spot on the array corresponds to the relative abundance of the mRNA transcript that the probe sequence was directed towards. For single-channel platforms, each sample is labeled with only one dye, and only one sample is hybridized to each microarray. Two-channel platforms use two dyes to label samples, enabling simultaneous profiling of two samples, which eases comparisons.

There are several strengths and weaknesses of gene expression microarrays. For example, some target–probe interactions are more efficient than others due to technical bias (variation in array production, as well as mRNA amplification, labeling, and hybridization efficiency). This means that a higher raw intensity from one spot does not mean the transcript is more abundant in the sample. For this reason, this type of microarray experiment must include many samples. Only then can transcript levels be compared between samples under the assumption that each target–probe interaction is subjected to the same technical bias in all samples. Before this is possible, several data-processing steps need to be performed, including data normalization. Quantile normalization means that each probe on the array is ranked by intensity and then reassigned a value based on the probe intensity distribution on all arrays. This process removes the effect of global differences between samples/arrays and diminishes differences due to technical bias. When a dataset is quantile normalized, it is assumed that all samples have the same distribution of transcript levels. This assumption may not hold true if the samples are biologically very different. In the case of a cohort of tumor samples, the assumption is usually made. The main strength of gene-expression microarrays is the wealth of data and powerful analysis methods that can be used on data.

The other major pitfalls of array-based gene-expression analysis are batch effect problems. A batch effect is a type of technical bias that cannot be corrected by standard normalization methods. The problem occurs when samples are treated batch wise, which usually happens on several occasions from retrieval of the tumor sample to hybridization. In other words, they describe a setting where variation between samples can be better explained by technical factors than by true biological variation. To combat batch effects, it is essential to know about all the sources and to use proper randomization in the experimental design. The usual way to correct batch effects is linear mixed models with confounding factors included as random intercepts or using empirical Bayesian methods (435, 436), although these techniques can unintentionally remove actual biological variation.

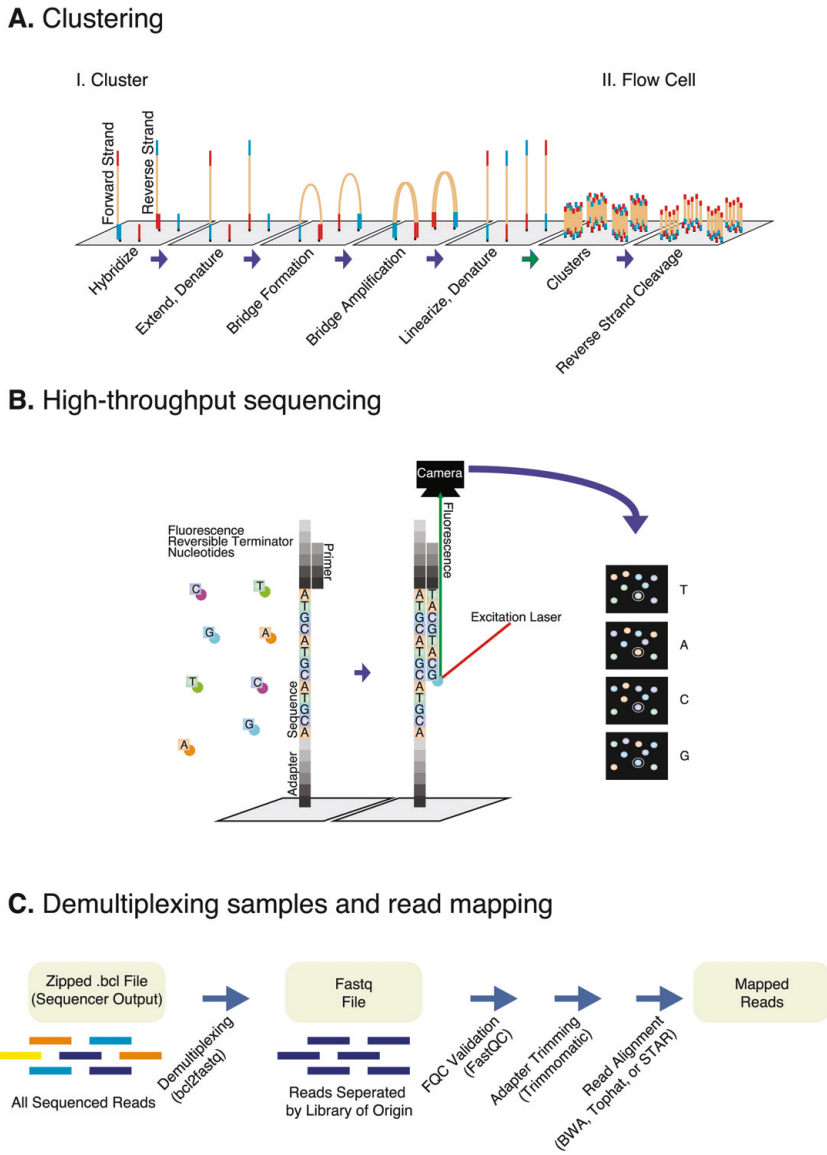


Figure 22.
Schematic overiew of next generation sequencing. Illustration from Chaitankar et al. (437) © 2016 Chaitankar et al. Open Access.

RNA-seq has emerged as the leading methodology for transcriptome profiling. In recent years, it has effectively replaced microarrays as the principal method for transcriptome profiling since it offers many advantages over previous methods. RNA-seq has a greater dynamic range and reproducibility and allows for detection

of *de novo* transcripts, quantification of known transcripts, and single-base resolution. These capabilities enable a multitude of applications, such as the possibility to detect fusion genes, copy-number aberrations, and structural variants, as well as the analysis of splicing and isoform switching. While the name suggests direct sequencing of RNA molecules, it is instead typically performed by sequencing cDNA resulting from RNA reverse-transcription.

To prepare a sample for high-throughput short-read sequencing, input RNA must be transformed into a sequencing library. In brief, library preparation consists of isolating and extracting mRNA from the tissue followed by quality control methods such as using agarose gel electrophoresis and spectrophotometry (e.g., Nanodrop). Then, the sample is enriched for mRNA by extracting poly(A) mRNA from total RNA. RNA is then fragmented, typically enzymatically using RNase enzymes to generate fragments of a suitable size range for sequencing. Reverse transcription is performed using reverse transcriptase enzymes and primers to synthesize complementary cDNA. If mRNA enrichment is performed, the primers are usually oligo(dT) or random hexamers. The cDNA fragments then undergo end repair to blunt the ends and add a single "A" base overhang at the 3' ends. Sequencing adapters are then ligated to both ends of the cDNA fragments. The adapters contain sequences necessary for attachment to the sequencing platform and priming during sequencing.

Following adapter ligation, size selection is often performed to remove undesired fragments (e.g., adapter dimers or very short fragments). Polymerase chain reaction amplification of the cDNA library is then performed to increase the amount of DNA material for sequencing. Before the final library undergoes sequencing, quality control is performed to assess its concentration, fragment size distribution, and purity by various methods such as bioanalyzer or fragment analyzer methods. SCAN-B used a variety of protocols (dUTP/TruSeq, NeoPrep/TruSeq, and dUTP/NeoPrep) for library preparation (420).

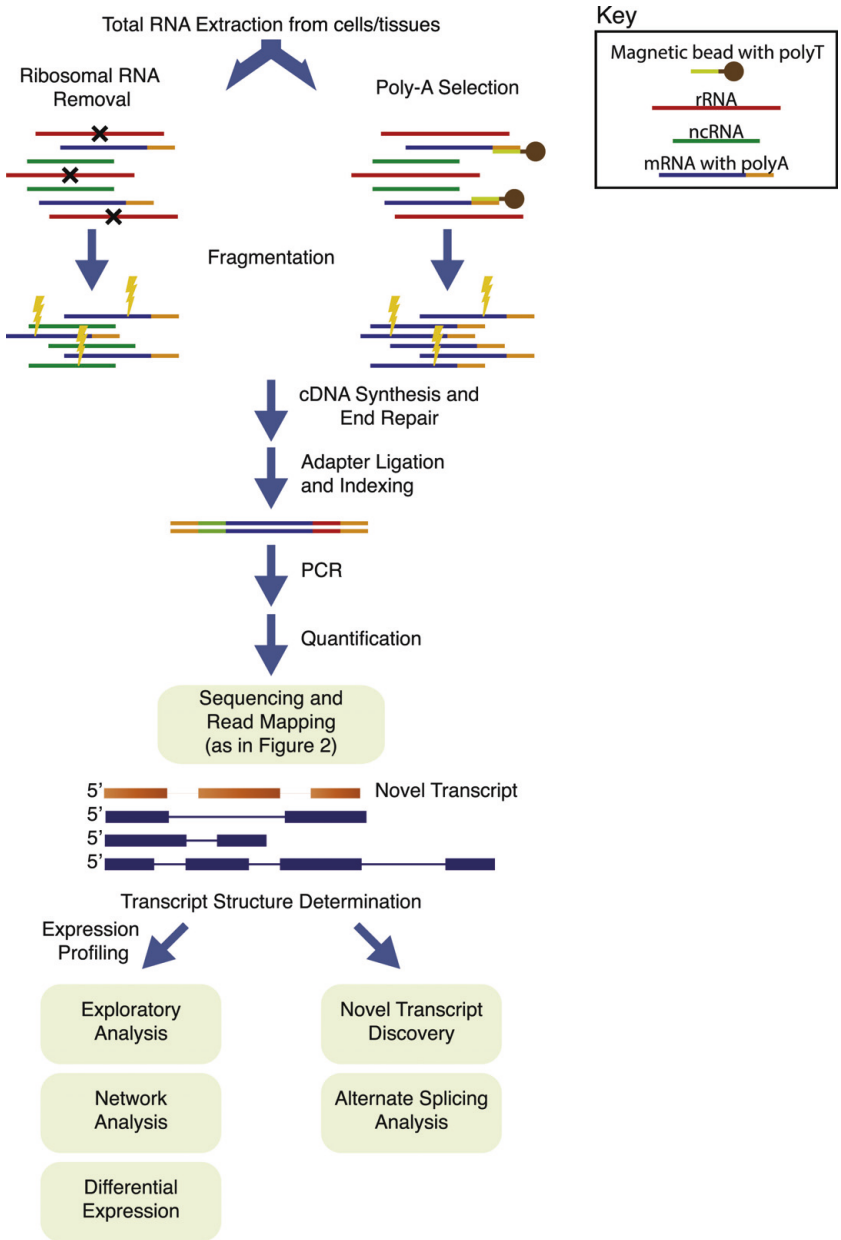


Figure 23.
Schematic overiew of library preparation for next generation sequencing. Illustration from Chaitankar et al. (437). © 2016 Chaitankar et al. Open Access.

Sequencing can be single-ended or paired-end, where a template molecule is only sequenced from one end or both ends, respectively, and the library can preserve

information about the strand from which a transcript originated. An important parameter in a sequencing setup is the average depth of coverage or number of reads to target for each sample. RNA-seq differs from DNA sequencing in that reads are distributed approximately proportionally to their expression level in the input sample, meaning that the average sequencing depth across an RNA-seq dataset is not a useful metric. Instead, the total number of sequencing reads is used to express sequencing depth.

In the SCAN-B RNAseq processing pipeline, sequencing was performed using Illumina HiSeq, NextSeq, and NovaSeq sequencers, and base-calling was done using Illumina's on-instrument CASAVA software, which converts fluorescence signals from read clusters into nucleotide base calls (420, 421). Sequencing reads are demultiplexed into samples FASTQ files (a widely used format for further analysis) in the Picard suite using IlluminaBasecallsToFastq (421). After sample demultiplexing, reads were trimmed to remove adapters and poor-quality base reads using Trimmomatic, all of which may complicate read alignment and bias expression estimation (421). Adapter contamination occurs when the cDNA template being sequenced is shorter than the requested read length, and thus, sequencing continues into the adapter (421). Low-quality base read can occur at the 5' ends due quality model calibration and at the 3' end due to imperfect sequencing (421). Afterwards, reads that match selected genomic sequences are removed, such as ribosomal RNA using bowtie. Despite library preparation procedures with selection of poly(A)-tailed RNA, ribosomal RNA may still be sequenced (421). Removing these sequences saves computation time and space and removes a source of analysis errors.

Using this information, reads were aligned to or the GRCh38/hg38 version of the human reference assembly using Hisat2, and the transcriptome target is defined by GENCODE (421). During read alignment (also called mapping), individual reads are placed into the correct position along a reference genome. For RNA-seq this is typically done with the help of a transcriptome annotation that provides information about splice junctions and transcript isoforms. RNA-seq aligners are splicing-aware and can take this extra information into account during alignment. Using the aligned reads, transcript expression was estimated using Stringtie with the help of a transcript annotation that describes introns and exons (GENCODE release 27 with using protein coding transcripts as the transcriptome model) (421). The number of reads per transcript is counted and summed at the gene level. Novel transcripts are discarded. Raw counts are biased by transcript length and number of reads per sample, so counts need to be normalized to enable within-sample and between-sample comparison. The measures of fragments per kilobase of exon model per million mapped reads (FPKM, used for paired-end data) were introduced as a measure for expression that is normalized within samples for library size and transcript length and are widely used. A later measure is transcripts per million reads (TPM), which accounts for the same factors but reverses the order of normalization

operations to enable better comparability between samples. Expression was estimated in FPKM. To reduce skewing of the data, FPKM values were log2 transformed, and a constant was added to avoid zeros since log2 (0) is undefined. A comprehensive quality analysis of early SCAN-B RNA-seq datasets has been performed (420). An important confounding problem in RNA-seq is (still) batch effects. Within SCAN-B, laboratory and sequencing processes have been optimized to minimize batch effects (421).

Genetics

To enhance the understanding of the genetic underpinnings of diseases, scientific research has focused on identifying variations in the human genome to discover genetic risk factors and develop genetic tests for diagnosis, prognosis, and personalized treatment plans. There are over 1.4 million known SNPs, and although many are not localized within genes, they encompass the majority of genetic variation in humans (149, 156). SNPs are variations in the genome where one nucleotide of the DNA chain has been exchanged with another (149, 156). Genetic population studies are conducted to identify associations between SNPs and diseases or individual traits, for which SNPs can serve as risk markers.

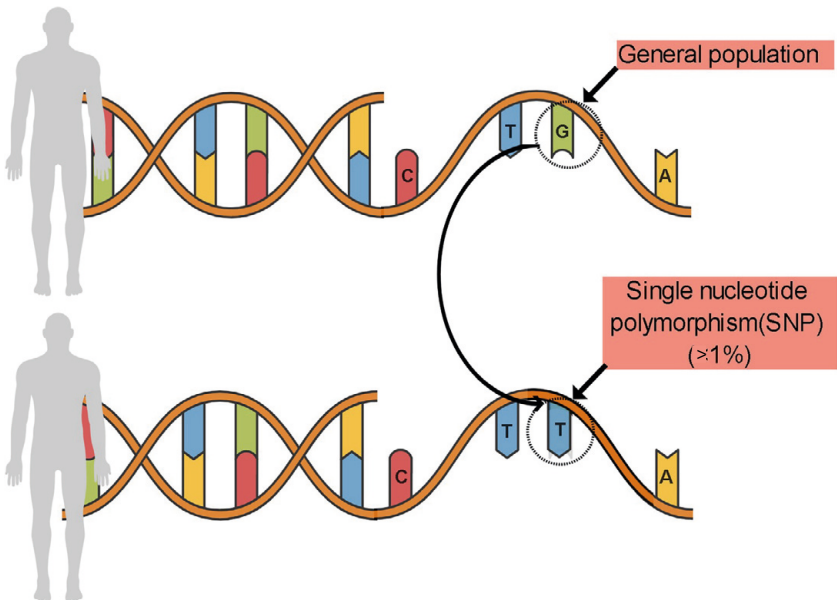


Figure 24.
An illustration of a SNP from Lima et al. (438). A SNP refers to a genetic variation where a single nucleotide (an allele) differs between individuals within a population. This variation results in alternative alleles at the specific genomic position. When the frequency of a particular allele surpasses 1% within a population, it signifies a SNP or common variant.

Another crucial concept is alleles, which are alternative forms of a gene at a specific chromosome locus (149, 156). Genes can have multiple versions or alleles that cause slight variations in the DNA sequence and influence diverse traits. Alleles are inherited from parents and contribute to genetic diversity within populations (149, 156). Diploid organisms like humans possess two alleles for each gene that are inherited from both parents (149, 156). An individual's genotype for a particular gene is determined by the combination of alleles they possess, which is classified as homozygous (if both alleles are identical) or heterozygous (if the alleles differ) (149, 156). Genotype refers to an individual's genetic makeup, specifically the alleles present at specific chromosome loci. In contrast, a haplotype denotes a specific combination of alleles along a stretch of DNA that tends to be inherited together (149, 156). Haplotypes consist of genetic variants (alleles) located on the same chromosome (149, 156). Haplotypes gain significance in the context of linkage disequilibrium (LD), where alleles at different loci are more likely to be inherited together than expected by chance (149, 156). LD alleles are physically linked on a chromosome and lead to non-random coinheritance and correlated frequencies in a population (149, 156).

The Hardy-Weinberg equilibrium (HWE) represents an ideal state of Mendelian inheritance. In this theoretical equilibrium, there is no genetic drift in the population due to factors such as mutations, fertility differences, natural selection, or migration (149, 156). If a population were in a true HWE, the distribution of alleles would remain consistent across generations (149, 156). Although this equilibrium is not observed in natural populations, it serves as a benchmark for assessing genetic drift. HWE is commonly utilized as a quality-control measure in genetic studies (149, 156). Proximity to the expected ideal value suggests robust genotyping quality. Conversely, significant deviations in the distribution of a genetic variant raise concerns about the reliability of genotyping (149, 156).

In paper III, genotyping was conducted in the BC-Blood cohort using a DNA microarray. DNA extraction from the leukocyte portion of whole blood was carried out using the DNeasy® Blood and Tissue kit and processed with QiaCube according to the manufacturer's instructions. The Centre for Translational Genomics at Lund University performed genotyping using the OncoAarray and initial quality control. The OncoArray used in this thesis comprises approximately 500,000 successfully manufactured markers or probes based on GRCh37/hg19 human genome assembly (439). Nearly 50% of these markers, which are part of the GWAS backbone (Illumina HumanCore), were selected to tag the vast majority of known common variants, thereby capturing most of the variation in the human genome (439). The remaining markers were chosen from seven lists, including those from disease consortia representing major cancer sites (the breasts, ovaries, lungs, colon, and prostate), potential modifiers of cancer risk in *BRCA1* or *BRCA2* pathogenic germline variant carriers, and a “common” list encompassing variants of general interest, such as pharmacogenomic markers and variants relevant to cancer-

associated traits (439). OncoArray configuration allows flexibility for cancers not initially involved in the array design by enabling the addition of custom content (439). SNP allocation principles were established by consensus among members of the OncoArray Consortium designed for research into the genetic basis of cancer, detailed elsewhere (439).

Standard quality control procedures were applied to all scans with the exclusion of samples with low call rates ($< 1 \times 10–5$) and SNPs with minor allele frequency $< 1\%$, or call rate $< 99\%$ (439). For *CAV1* SNPs, genotype-intensity cluster plots underwent manual examination for reliability assessment (148). Out of six *CAV1* SNPs (rs10256914, rs959173, rs3807989, rs3815412, and rs8713), five passed quality control and were in HWE, while the excluded SNP had a minor allele frequency $< 1\%$. *CAV1* haplotypes were constructed by cross-tabulating the five *CAV1* SNPs among the 1017 genotyped patients. The most likely SNP combinations were used to construct the haplotypes, and the analysis was limited to those present in more than 10% of patients.

Statistical and bioinformatic analyses

> *"The main purpose of a significance test is to inhibit the natural enthusiasm of the investigator"*
>
> — Frederick Mosteller

Statistical inference and type I and II errors

The main reason why researchers use statistical analysis in their studies is to be able to draw conclusions about a population based on a smaller sample of that population, which is called statistical inference. There are two main approaches to statistical inference: Frequentist and Bayesian statistics, which differ fundamentally in their interpretation and application of probability. Frequentist inference is also known as classical inference and focuses on the long-run frequency or probability of events occurring. In this framework, probability is interpreted as the limit of the relative frequency of an event occurring over repeated observations. Frequentist methods rely on data-driven estimation and hypothesis testing, where parameters are considered fixed and unknown but estimable. This is the approach used in all the papers in this thesis. Bayesian statistics, on the other hand, is based on Bayes' theorem, which updates prior beliefs about parameters using observed data to produce posterior beliefs. In Bayesian inference, probability represents a degree of belief or uncertainty about an event, incorporating both prior knowledge and new evidence.

Generally, there is a null hypothesis defined by the investigators, which typically assumes that there is no difference when comparing groups. Different statistical methods are used in order to test if the null hypothesis can be rejected. In this context, three statistical concepts are important: effect/point estimate, confidence intervals, and the p-value. A p-value is a value between 0 and 1 and measures how consistent the observed difference in the comparison is with the null-hypothesis. The p-value is a measure of the probability of observing a difference as large or larger as the one found in the sample if every model assumption was correct, including the null hypothesis. The degree of acceptable certainty is the level of significance and is often set at a p-value less than 0.05.

A problem arises in the dichotomization as a statistically significant result or not and when the conclusion of "no difference/association" is based on a p-value being arbitrarily set at a certain threshold (e.g., 0.05). In reality, p-values of 0.04999 and 0.05001 are not that different, but one denotes that there is a significance, while the other denotes that there is no difference, which is a very simplistic conclusion. There has been a shift towards confidence intervals and estimation to improve interpretation and give more meaningful information. This is especially important in the clinical context since the p-value is not an estimate of the effect, it is the amount of evidence in the sampled data against the null hypothesis. Also, it is more common to consider a p-value as the level of significance without setting an arbitrary limit, which is an improvement, although more careful interpretation is warranted from readers. Perhaps most importantly, statistical significance is not the same as clinical significance, and in many epidemiological, translational, and clinical studies, the question is whether the effect is clinically meaningful and whether the estimates are precise enough to draw firm conclusions. An exception is the RCT, where a strict p-value cut-off is necessary to decide whether the investigational treatment is superior or non-inferior to the standard of care in a head-to-head comparison, essentially determining whether to continue with standard care or switch to the investigational treatment.

When performing hypothesis testing, there is always a risk of two errors occurring: types I and II errors. Type I error is related to the level of significance (α) and happens when one wrongly rejects a true null hypothesis. Simply put, a type I error is a false positive; the test shows a statistically significant difference even though there is no difference in the underlying population from which the sample was drawn. Type II error, on the other hand, is a false negative and happens when one wrongly accepts a false null hypothesis. The probability of a Type II error is denoted by β and depends on the power of the study, where β equals 1 minus the power (β = 1 - power). The probability of having a Type II error is commonly set at 0.20. Statistical power refers to the probability that a statistical test will correctly reject a false null hypothesis. In simpler terms, it measures the likelihood that a study will detect an effect or difference when one truly exists. A study with high statistical power has a better chance of detecting true effects, while a study with low power is

less likely to do so. The statistical power is influenced by several factors. Larger sample sizes generally result in higher statistical power because they provide more information, the magnitude of the difference or effect being studied also affects statistical power, and larger effect sizes are easier to detect and result in higher power. The chosen alpha level (usually set at 0.05) determines the threshold for rejecting the null hypothesis. A lower significance level reduces the chance of Type I errors but decrease statistical power. The amount of variability or spread in the data can impact statistical power with greater variability reduces power, while less variability increases power depending on the type of statistical test used. A homogenous study population and standardized data collection could actually increase power. In essence, statistical power reflects the ability of a study to detect true effects or differences. However, statistical power is much harder to estimate and use. The reason is that most studies besides clinical trials do not use pre-specified sample size and most likely use all available patients/participants or biospecimens for analysis, and there are other limiting practical factors such as cost, sample availability, and recruitment rates.

Estimated type I and II error is actually valid if one test is performed. In simple terms, for type I, there is a one in twenty chance of a false positive for one test, but if two tests are performed, this chance increases to approximately 9.8% (if the two null hypotheses are true) if the alpha level is kept at 0.05. If five tests are performed (and all null hypotheses are true) the chance is 22.6%. The mathematical formula is as follows:

$$p = 1 - (1 - \alpha)^n$$

Where the p stands for probability of a at least one Type I error, alpha is the significance level of an individual test, and n is the number of statistical tests. When these numbers are calculated, they highlight the problem with multiple testing, which arises when researchers conduct multiple statistical tests on the same dataset or set of hypotheses without adjusting for the increased risk of false positives. This refers to the problem with multiple testing, which arises when researchers conduct multiple statistical tests on the same dataset or set of hypotheses without adjusting for the increased risk of false positives. In other words, as more tests are performed, the likelihood of obtaining at least one significant result by chance alone increases, leading to inflated Type I error rates.

Combined with a very static view (which should not be used) that a p-value under 0.05 implies an association or cause and that only then can a study be interesting, which leads to “data dredging”. This quite common practice refers to when researchers conduct numerous statistical tests on a dataset until statistically significant results are found. Often, the *a priori* hypotheses or theoretical basis are weak or nonexistent. While it may seem like an exploratory approach to uncover hidden patterns or relationships, data dredging poses several problems as it leads to

an inflated rate of Type I error, and without appropriate correction methods, the risk of false positives becomes unacceptably high. This leads to false discoveries due to the appearance of spurious correlations or associations that do not reflect true underlying relationships. These false discoveries can mislead subsequent research efforts and lead to erroneous conclusions. This contributes to the reproducibility crisis in science as results obtained through data dredging are less likely to be replicated in independent datasets or studies. Since the observed associations are often driven by chance rather than genuine effects, they may not withstand scrutiny or hold in different contexts. This is mainly due to the overfitting of statistical models to the data, which captures noise or random fluctuations instead of true patterns. While exploratory data analysis can be a valuable precursor to hypothesis generation, data dredging blurs the distinction between hypothesis generation and hypothesis testing. Without clear hypotheses guiding the analysis, researchers may miss important relationships or overlook relevant variables. This is often aggravated by selective reporting, and in many cases researchers, fail to disclose the extent of the exploration conducted. This can introduce publication bias and distort the scientific literature. Taken together, this has contributed to a large discussion about metascience driven by Professor Ioannidis, who very famously published the highly cited article "Why most published research findings are false" estimated that as much as 70% of publications are false (440). This was followed up by, "Why Most Discovered True Associations Are Inflated" (441).

To mitigate the problems associated with data dredging and replication, more stringent guidelines have been adopted in many disciplines. Researchers should adopt transparent and principled approaches to data analysis, including preregistration of hypotheses, the use of appropriate correction methods for multiple testing, and validation of findings in independent datasets. Clear reporting of all analyses conducted, including nonsignificant results and data exploration procedures, could enhance the credibility and reproducibility of research findings. Moreover, researchers can use methods to correct for multiple testing and then control p-value inflation. Multiple testing correction is a statistical method used to adjust the significance threshold for hypothesis testing when multiple comparisons are performed simultaneously. In many scientific studies, researchers test multiple hypotheses or perform numerous statistical tests, which increase the likelihood of falsely declaring a result as significant by chance alone.

There are several approaches to multiple testing correction, and the choice depends on factors such as the structure of the data and the assumptions made. The two most common ones are Bonferroni correction and the false discovery rate (FDR) or Benjamini-Hochberg procedure. Bonferroni correction is one of the simplest and most conservative methods and adjusts the significance threshold (usually α, commonly set at 0.05) by dividing it by the number of tests being performed. For example, if conducting 10 tests, Bonferroni correction would set the significance threshold at $0.05/10 = 0.005$ for each test. While effective at controlling the family-

wise error rate, it can be overly stringent and lead to reduced power. The family-wise error rate denotes the probability of making at least one Type I error (false positive) in a set of hypothesis tests or comparisons. Unlike Bonferroni correction, which controls the family-wise error rate, FDR correction controls the expected proportion of false positives among the rejected hypotheses rather than the family-wise error rate. This method ranks the p-values from smallest to largest, compares them to a critical threshold determined based on the desired FDR level, and then determines which null hypotheses to reject. Due to the exploratory nature of the papers in this thesis, the Benjamini-Hochberg procedure was chosen since it is less conservative. There are also other less common methods, such as permutation testing, bootstrapping, and the Holm-Bonferroni method.

Descriptive statistics

In all papers in this thesis, descriptive statistics were used to report the basic characteristics of the study participants. Descriptive tables are often used to reduce a large amount of data into a simpler summary. This will help the readers evaluate the generalizability of the findings of the study. Hence, descriptive statistics is used in all papers in this thesis. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines recommend avoiding using significance tests because they can lead to misinterpretation and inappropriate comparisons, which diverts focus from the main research question, and the scientific objectives of the study are not directly addressed (442). Perhaps most importantly, statistical tests are used to drive inference about the underlying population and not the study sample, which means that there is no need to use statistical tests to evaluate differences between groups within the sample, as that is not their intended purpose. Nonetheless, most of the studies in the thesis use different statistical tests to make comparisons between groups. It can of course be valuable when evaluating new biomarkers to see what type of clinicopathological and molecular features they are associated with to gain a better understanding of biology and clinical aspects. In hindsight, I would have avoided many of them and used them only to test whether certain characteristics differed between groups. Also, many statistical tests were performed, and most likely, there was a high rate of false positives, which lead to problems with drawing conclusions from these tests.

For comparisons between two independent groups of continuous variables the Student's t-test (a parametric test) or Mann-Whitney U-test (a non-parametric test) are mainly used. Data that conform to a normal distribution can be tested using parametric tests, while for non-normal, non-parametric tests can be used, but they are less informative. For example, they cannot provide point estimates or confidence intervals. "Parametric" means that the test makes assumptions about data used in the analysis. Non-normally distributed variables can be transformed in a number of ways to approximate the normal distribution. Logarithmic transformation (used in

papers III and V) can sometimes be used to approximate the normal distribution. Comparisons between multiple groups can be done using an analysis of variance (ANOVA), or the Kruskal-Wallis test, which is non-parametric When comparing categorical variables (frequencies) between groups the chi-square test or the Fisher's exact test can be used. However, all of these methods besides ANOVA (which can extend to multivariate ANOVA, among others) have a drawback in that they are univariable, and adjustment for confounding is not possible. Since comparing different clinicopathological characteristics between groups was never the main focus, in most papers, multivariable analysis was not performed.

Linear regression

Regression is a flexible statistical framework with the aim of providing understanding about how an outcome of interest varies as one or multiple variables vary. This framework can be adapted to different natures and distributions of interest and can be used for inference and prediction. The most common and simple modality of regression analysis is linear regression, which is a statistical method used to model the relationship between a continuous dependent variable and one or more independent variables. It is commonly used to understand and predict the behavior of a continuous outcome variable based on one or more predictor variables. The basic form of linear regression is known as simple linear regression and involves fitting a straight line to a set of data points. The equation of the line is represented as:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

y is the dependent variable (the variable being predicted or explained), x_1 is the independent variable (the predictor variable), β_0 is the intercept (the value of y when x_1 is zero), β_1 is the slope (the change in y for a one-unit change in x_1), and ε is the error term (the difference between the observed and predicted values of y).

The goal of linear regression is to estimate the values of β_0 and β_1 that best fit the data. This is typically done using the least squares method, which minimizes the sum of the squared differences between the observed and predicted values of y. The main assumptions of the linear regression are that the relationship between y and x is linear, the variance of the residual is the same for any value of x, observations are independent of each other, there is absence of multicollinearity (independent variables are not highly correlated), x is normally distributed, and fixed values of x with no variability. Linear regression analysis allows researchers to quantify the strength and direction of the relationship between variables, make predictions about the dependent variable based on the values of the independent variables, and test hypotheses about the statistical significance of the relationships between variables.

Logistic regression

Logistic regression is an extension of linear regression and is used to model the relationship between one or more independent variables (predictors) and a binary outcome variable. Binary means it has only two possible outcomes (yes and no). Unlike linear regression, which models continuous outcomes, logistic regression models the probability of the outcome using a logit function. The logit function is applied to the odds, which is the probability of success divided by the probability of failure, resulting in the natural logarithm of odds.

The main objective of logistic regression is to estimate the probability that a particular outcome occurs based on the values of the independent variables. The logistic function transforms the linear combination of the independent variables into a value between 0 and 1, which represents the probability of the outcome occurring. In logistic regression, the coefficients are estimated using maximum likelihood estimation. The likelihood function measures the probability of observing the data given the parameters of the statistical model. In other words, it quantifies how likely the observed data are under the assumption that the model is true. The maximum likelihood estimation adds together the log-likelihood of all observations/individuals, and by maximizing the log-likelihood, it finds the values of parameters that make the observed data most probable under the assumed model. Once the maximum likelihood estimates of the parameters are obtained, the log-likelihood value itself can serve as a measure of model fit. A higher log-likelihood indicates that the model provides a better fit to the data, while a lower log-likelihood suggests poorer fit.

Once the logistic regression model is fitted, it can be used to make predictions about the probability of the outcome for new observations based on their values of the independent variables. Additionally, the model coefficients can provide insights into the strength and direction of the relationship between the independent variables and the log-odds of the outcome. This was the primary reason for using logistic regression in papers I and V.

The main assumptions of logistic regression are independence of observations, linearity between continuous variables and logit-transformed outcomes, absence of multicollinearity, and lack of strongly influential outliers. Logistic regression yields reliable, robust, and valid results when a larger sample size of the dataset is considered. A rule of thumb is to have a minimum of 10 cases considering the least frequent outcome for each independent variable to reduce overfitting. Logistic regression can be extended to accommodate more than two possible discrete outcomes, whether categorical or ordinal. In such cases, it is referred to as multinomial logistic regression for categorical outcomes and ordinal regression for ordinal outcomes. The goal of multinomial logistic regression is to estimate the probability of each category of the dependent variable relative to a reference category, given the values of the independent variables.

Survival analysis

In cancer research, comparison of survival between different groups is often essential. Therefore, survival analysis, also called time-to-event analysis, is widely used. In all types of survival analyses regardless of which exact method is being used certain requirements must be fulfilled, generally related to follow-up. Censored patients lost to follow-up should have the same incidence of event as patients that remain in the study. There should also be adequate length of follow-up to capture enough events for sufficient power and similar completeness of follow-up between the compared groups. This best achieved by having good and complete follow-up, which is true for the datasets used in this thesis. Also, very importantly, the survival probabilities for patients included early should be the same for patients recruited later in the study. This presents a problem for the BC-blood cohort and METABRIC, which both had very long inclusion periods, and treatments have changed over time. This could then bias the results. A good way to handle this problem is to use a stratified Cox model, which allows for groups to have different baseline hazards, and then patients could be grouped by year(s) of inclusion.

The standard methods used are the Kaplan-Meier (KM) method combined with the log-rank test to illustrate and compare differences in survival probabilities. These were used in almost all papers in this thesis. The KM estimate uses the exact failure and censoring time and considers the number of individuals at risk for an event to estimate survival probability changes at the time of each event. The KM method is a nonparametric method, and if individuals are lost to follow-up, this, *per se*, does not affect the estimate of survival probability. However, as more individuals are censored, each individual event becomes more influential. Therefore, when few individuals remain at risk, the KM estimate should be interpreted with caution. The KM method can be used to estimate survival at different timepoints during the follow-up, or more commonly, KM estimates can be visualised as KM curves. In all papers, KM estimates and log-rank tests have been used to examine survival differences between groups. One can argue that the follow-up time in some of the KM curves should have been reduced since many patients had been censored. Therefore, to minimize the risk of misinterpretation, the KM curves were combined with a life table over the number of individuals at risk at certain time points.

The log-rank test is a nonparametric test with the null hypothesis that there is no difference in survival between the groups. It is the most widely used method for comparing two or more groups in terms of survival. This method compares the observed number of events to the expected number under the null hypothesis of no survival differences between the groups. Then, the method compares a test statistic based on these numbers to a $\chi 2$ distribution with degrees of freedom equal to the number of groups minus one to determine the p-value. The major drawbacks of the log-rank test are that it does not provide information regarding the effect size and

that it is a univariable analysis method, thus it is impossible to adjust for potential confounders.

Cox proportional hazards regression

The most common method used to analyze survival data is Cox regression, which is also known as the Cox proportional hazards regression model. It is used for understanding the relationship between the time until an outcome of interest occurs (death or disease recurrence) and one or more predictor variables. This thesis focuses mainly on prognostic factors; hence, Cox regression was used in all papers. The hazard is the incidence rate of an event in an infinitesimal short time period, (e.g. the probability of an event within the next second). For all practical purposes, hazards can be thought of as incidence rates, meaning that Cox regression compares the incidence of different groups. Incidence is defined as the number of events per unit of time. For example, if there are 50 deaths recorded in 100 person-years of follow-up, the incidence estimation is 0.5 deaths per person-year. The cumulative sum of incidence over a set time is known as risk (the probability of experiencing a particular event within a specified time frame) and expressed as a percentage, like the 5-year mortality. Incidence and risk are closely related concepts in survival analysis. Incidence contributes to the overall risk of experiencing an event during a time period and points towards the direction of risk. However, it is important to note that Cox regression does not estimate relative risk directly, and therefore, caution should be exercised when interpreting its results as measures of risk. Despite this, in many studies, including the present studies, Cox regression results have been interpreted as risk, which may not accurately reflect the nature of the analysis.

In Cox regression, the hazard function represents the instantaneous risk of experiencing the event at any given time and is modeled as a function of the predictor variables. The Cox regression model estimates the hazard ratios (HRs) associated with each predictor variable while controlling for other variables in the model. These HRs indicate the relative change in the hazard rate for one unit change in the predictor variable while holding all other variables constant. The HR can be roughly interpreted as the incidence rate ratio.

The Cox proportional hazards regression model is written as follows:

$$h(t) = h_0 \times \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

where $x_1 \cdots x_p$ represents the predictor variables, and $h_0(t)$ is the baseline hazard at time t, which is the hazard of an individual having all the predictors set to zero. The β coefficients represent the effect estimate, and when exponentiated, they are transformed into the HR for each level of the variables. The HR is a point estimate and does not express the statistical variation (or random error) around the estimate. The CI helps to quantify the precision of this sample estimate and is highly related to the p-value.

Cox proportional hazard regression is a semiparametric method because it makes no assumption about the distribution of survival times. However, it does assume that the HR between compared groups remains constant over time, which is known as the proportionality assumption; hence its name. The proportional hazards assumption can be checked graphically and using statistical tests based on the scaled Schoenfeld residuals. In principle, the Schoenfeld residuals are independent of time if the proportionality assumption is fulfilled. A plot that shows a non-random pattern against time is evidence of violation of the proportional hazards assumption. This approach was used in most papers. In reality this can hard to truly fulfill. When smaller violations occur, the HR can be seen as the mean HR across the time-period. Other methods to handle violations of the proportional hazard assumption include stratifying the model by survival time in other words fitting and using separate Cox regression models for different follow-up time points (e.g., 0-3 years, 3-6 years, 6-10 years, etc.), thus estimating separate hazard ratios (HRs) and independent (time-varying) coefficients for each time period. However, this approach reduces power as the data are split into smaller pieces.

Cox regression can also be stratified by a variable suspected to have a time-varying effect, which requires this variable to be categorical or categorized. Each stratum k possesses a distinct baseline hazard but shares common values for the coefficient vector β. Stratification assumes that other covariates behave similarly across all strata, implying that HRs remain consistent across strata. Although stratification effectively addresses the issue of non-proportionality and is straightforward to implement, it does have drawbacks. Notably, stratification by a non-proportional variable precludes the estimation of its strength and its testing within the Cox model. Therefore, this approach should be chosen when the direct quantification of the effect of the stratification variable is not a primary concern. The final way is to include time-varying effects (coefficients) in the Cox regression model. This is done by including the variable's interaction with some function of time. If the function f(t) by which the effect varies with time is known, the effect is modelled, although this is sometimes more easily said than done. Other assumptions of the Cox regression are the independence of observations, absence of multicollinearity, linearity between continuous variables and the hazard functions (commonly tested with Martingale residuals), and lack of strongly influential outliers, which are very similar to those of other regression models.

Similar to logistic regression, the Cox proportional hazard regression uses the log-likelihood to estimate model parameters using uses the partial likelihood function. The function is called "partial" because it focuses only on the individuals who experience an event and ignores those who did not. The partial likelihood function orders all the individuals in the dataset by the time it takes for the event to occur, from shortest to longest survival time. Then, it calculates the likelihood that the ordering of events happened the way it did, based on the differences observed

between individuals who had the event at different times. Finally, the model parameters (coefficients) to maximize the likelihood that the observed ordering of events occurred are estimated. In the event of a tie, where two individuals experience the event at the same time, various methods like Breslow's method can be employed to address this situation.

Competing risks

In survival analyses, there is often a possibility that more than one type of event could occur as competing risks. The definition of a competing risk is an event that hinders the observation of an event of interest or alters its probability. In other words, a competing risk competes with the event of interest to remove individuals from the population at risk. In the presence of competing risks, one option is to fit a cause-specific Cox proportional hazards model, in which competing events constitute censoring. The corresponding hazard ratios should then be interpreted in a hypothetical scenario where all the competing risks have been eliminated. This can be when the competing events are independent; i.e., the variables of interest are not associated with the cause of censoring, so censoring is more or less equal across groups.

The second frequently used analysis strategy is a competing risk analysis to estimate the cumulative incidence function (CIF) for each event of interest while accounting for the presence of competing events. The CIF represents the probability of experiencing a particular event over time, which is conditional on the probability of surviving (experiencing no competing events) over time. When estimating CIFs, the competing events do not have to be independent. In competing risk analysis, the CIF describes the instantaneous rate at which an event occurs at a given time, given that an individual has survived up to that time without experiencing any competing events. The CIF can be estimated separately for each event of interest, which allows researchers to assess the risk of each event while considering the presence of competing risks. The analysis is usually performed using the Fine and Gray model for subhazards (443). This model extends the Cox proportional hazards model used in standard survival analysis to accommodate competing risks (443). To test the robustness of our results in paper III, competing risk analysis was performed.

The KM estimate is based on the assumption that the event of interest is the only possible event. Since the life expectancy of humans is limited, analyses of cause-specific mortality or recurrence introduces competing risks and a bias in the estimate. Instead of using the one minus the KM estimate, a cumulative incidence curves based on both the event of interest and competing risk events would introduce less bias to the curves and survival estimate. However, the KM estimate is considered the gold standard, and many clinicians are more familiar with interpreting it, so it is still widely used.

Endpoints

Different endpoints were used in the various studies included in the thesis. This was due to a combination of slightly different aims and the availability of data; certain endpoints were only available for specific datasets. The primary endpoint in papers I and IV was any breast cancer event. Secondary analyses also included distant metastasis and overall survival. For paper I, data-driven exploratory analysis was conducted for locoregional recurrence and contralateral breast cancer. Paper II built upon paper I and used all five previous endpoints with the primary endpoints being locoregional recurrence and contralateral breast cancer. Paper III had a different set of endpoints due to the use of different cohorts, recurrence, distant metastasis, overall, and breast cancer-specific survival. The main endpoint was distant metastasis. Paper V was somewhat different since the main endpoint was pCR with secondary endpoints of recurrence and distant metastasis. The main problem with endpoints is that they are related but not exactly the same, and certain factors can have varying influence on each specific one. This is very clearly showcased in paper I for CAV1 in malignant cells. Often, there are somewhat different definitions of what constitutes an event and an endpoint, which can influence the results, which is why it is important for a publication or reference to clearly define what constitutes an endpoint, regardless of what it is called. Preferably, the primary endpoint should be clearly defined in the statistical analysis plan prior to the start of the study. The reason for choosing the endpoint for each study can vary and generally the main endpoints in this thesis have been a type of "event-free" interval implicitly stating that death is censored and not considered an event of interest. Of course, survival (meaning time to death) is by far the most important clinical endpoint, but in biomarker studies, one would generally assume that only way a biomarker affects survival is through disease progression (e.g., recurrence) and that only the time to the event (an interval) is relevant. By using a composite endpoint, the signal might be weakened since it may only be associated with a single component.

Interaction analysis

Interaction analysis is a statistical method used to examine whether the relationship between two variables changes depending on the level of a third variable. In other words, it investigates whether the effect of one independent variable on the dependent variable varies across different levels of another independent variable. Interaction effects are particularly relevant when studying complex relationships between variables and when considering the possibility that the effect of one variable may be contingent upon the presence or absence of another variable. Interaction analysis enables researchers to identify whether the relationship between two variables is conditional on the level of a third variable, determine the direction and strength of the interaction effect, and understand how the effect of one variable may differ across different subgroups of the population defined by another variable.

Interaction is commonly tested by including the interaction term in the regression model (regardless of which type of regression model it is).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \times x_2) + \varepsilon$$

In this linear regression model, β_1 represents the effect of the first variable x_1 in the reference group of x_2, β_2 represents the effect of the second variable x_2 in the reference group of x_1, β_3 represents the interaction term, and ε is the error term.

There are two main types of interactions, additive and multiplicative. In an additive interaction, the assessment examines whether the combined effect of two variables on the outcome is larger than the sum of their individual effects. If not, and the sum of their individual effects is smaller than their combined effects, then the additional effect represents their additive interaction. It is commonly quantified as the relative risk due to interaction (RERI). In this thesis, the mainly multiplicative interactions were tested, but it may have been of interest to also explore additive interactions, especially for prognostic biomarkers.

Alternatively, multiplicative interaction instead assesses whether the combined effect is larger (or smaller) than the product of the individual effects of the variables. Multiplicative interaction can be assessed for more than two variables. If the multiplicative interaction term provides evidence of an interaction, it indicates that the effect of variable x_1 on outcome y is different across different strata of variable x_2. In other words, the relationship between x_1 and y is moderated by x_2 or vice-versa.

The evidence can be evaluated by including a product term in a regression model. Then, the Wald test uses the parameter's sample estimate and an estimate of variability to obtain evidence of whether the included factor improves the regression model (by adding independent information). The Wald test can be used to simultaneously test many parameters. Otherwise, the likelihood ratio test can be used to compare models with or without the interaction term. This test provides evidence on how much additional prognostic information the interaction term gives the model in addition to the already included factors. In practical terms, both tests give very similar results.

Multiple imputation and missing data

Missing data present a major problem in epidemiological and clinical research. Missing data (or missing values) are defined as the data values that are not stored for a variable in an observation of interest. The problem of missing data is relatively common in almost all research and can have a significant effect on the conclusions that can be drawn from the data. Missing data present various problems. The absence of data reduces statistical power, cause bias in the estimation of parameters, reduce the representativeness of the samples, and complicate the analysis. Each of these distortions may threaten the validity of trials and can lead to invalid conclusions. The best way to deal with missing data is to retrospectively re-examine and collect data to minimize the missing values if possible. However, in many cases, this may not be possible, and in reality, the best way to deal with missing data is to plan a careful study design that minimizes it. Regardless, in practical terms, it is something that many researchers have to deal with. In many cases, the chosen method is to simply run a complete case analysis that includes only observations (or patients) that have data available for all variables examined in the statistical analysis. This can be sound if the missingness is low.

There are three main types of missing data. The first type is "missing completely at random" (MCAR), which occurs when the probability of a value being missing is unrelated to both observed and unobserved data (444). In other words, the missingness is completely random and occurs independently of any other variables in the dataset. For example, data may be missing due to equipment failure or administrative errors. The second type are "missing at random" (MAR) and occurs when the probability of a value missing depends only on observed data and not unobserved data (444). In this case, the missingness can be systematically related to other variables in the dataset, but conditional on the observed variables, it is random. For example, if men are less likely to report their weight than women, once gender is known, the probability of missing weight values is random. The third type is "missing not at random" (MNAR), which occurs when the probability of a value missing depends on unobserved data or the missing values themselves (444). In this case, the missingness is non-random and can be related to the value of the missing data. One example is if individuals with higher income are less likely to report their income, and income is missing for those individuals. Understanding the type of missing data is important because it influences the choice of imputation method and the validity of statistical analyses. For example, if data are MCAR, any imputation method can be used without introducing bias. However, if data are MNAR, imputation methods may introduce bias and are generally not suitable. In reality, classifying missing data patterns may not be straightforward, and many cases exhibit patterns that lie somewhere between MAR and MNAR. This nuanced understanding is crucial when interpreting results derived from imputation methods. In the sensitivity analysis in papers I and II, we choose to handle missing data with multiple imputation by chained equations (MICE). This statistical method is used to

address missing data by imputing multiple sets of plausible values for the missing observations (444). It is a flexible and widely used approach that can handle various types of missing data patterns and can be applied to both categorical and continuous variables (444). In the MICE approach, the process begins by imputing the mean for every missing value in the dataset, which serves as placeholders. (444). Then, the mean imputations for a specific variable ("var") are reset to missing (444). The observed values of "var" are then regressed on the other variables in the imputation model. In this regression model, "var" serves as the dependent variable, while all other variables are the independent variables, both the observed and newly imputed values. These regression models function as they would in contexts outside of imputing missing data. The missing values for "var" are subsequently replaced with predictions (imputations) derived from the regression model. This process is repeated for each variable with missing data, with one complete cycle through all variables constituting one iteration. After one iteration, all missing values are replaced with imputed values. Typically, this procedure involves ten iterations. Once the designated number of cycles is complete, the entire imputation process is repeated to generate multiple imputed datasets. Each dataset is then analyzed using standard statistical methods for complete data, yielding multiple analysis results. Finally, by combining these results using Rubin's rules, a single overall result is produced (444).

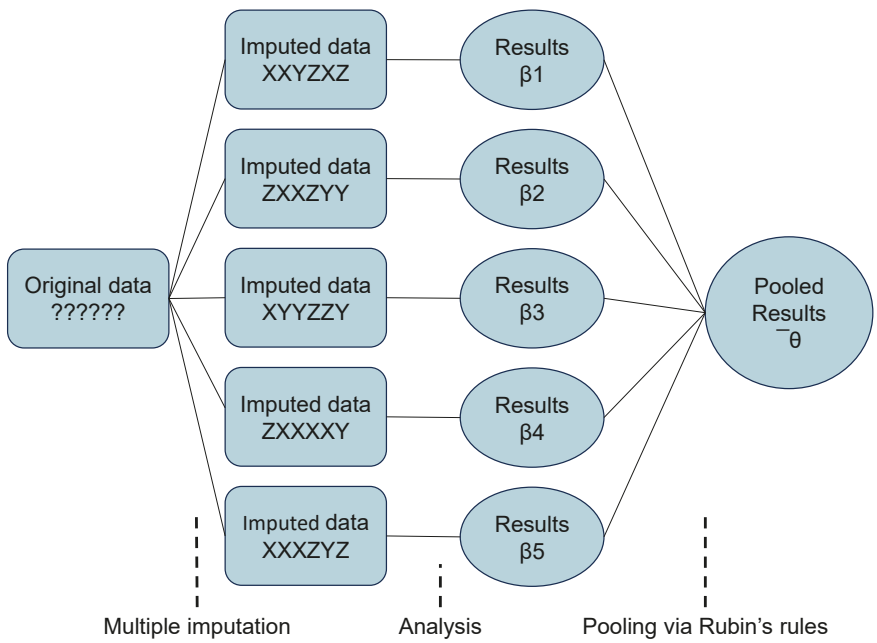


Figure 25.
Schematic overview of MICE.

The benefit of the MICE is that in addition to restoring the natural variability of the missing values, it incorporates the uncertainty due to the missing data, which results in valid statistical inference (444). Restoring the natural variability of missing data can be achieved by replacing it with imputed values drawn from a normal distribution of likely values, which are predicted using variables correlated with the missing data. Incorporating uncertainty is made by producing different versions of the missing data and observing the variability between the imputed data sets. This allows MICE to handle complex missing data patterns. It also allows for the inclusion of auxiliary variables in the imputation models, which can improve imputation accuracy. MICE has been shown to produce valid statistical inference that reflects the uncertainty associated with the estimation of the missing data (444). Furthermore, multiple imputation turns out to be robust to the violation of the normality assumptions and produces appropriate results even in the presence of a small sample size or a high fraction of missing data (444).

The other studies generally exhibited a high degree of data completeness, and we opted not to utilize imputation methods to address missing data. Instead, we relied on complete case analysis. However, it would have been beneficial to conduct a sensitivity analysis to validate the robustness of the results.

Differential gene expression analysis

Differential gene expression (DGE) analysis is a fundamental technique in systems biology that aims to identify genes with expression levels that are significantly altered between different experimental or biological conditions, such as diseased versus healthy tissues and drug-treated versus untreated samples (435, 436). This analysis provides insights into the molecular mechanisms underlying biological processes and diseases. This method is relatively new and was developed after the introduction of the gene expression microarray, which introduces new challenges when evaluating large-scale data. Thus, there are a number of methods for differential expression analysis for microarray/RNA-Seq data (435). However, there is no consensus about the most appropriate pipeline or method for identifying differentially expressed genes from RNA-Seq data as each one has its own strengths and limitations (435, 436). In papers III and V, the Limma-Voom package was used for DGE analysis, which has good sensitivity and validity for detecting differentially expressed genes (428, 435, 436). The Limma-Voom package can also be used to analyze data from gene expression microarrays, which many newer DGE methods cannot do because they have their own normalization procedures and require raw counts (428, 435, 436). This can be relevant when comparing results from different datasets, as done in paper V. The actual analysis involves several steps after data processing (as described above). The expression matrices are subject to statistical analysis by various methods to compare gene expression levels between the groups. At their core, each method uses an extension of linear regression (a generalized

linear model) with each gene being the dependent variable combined with an empirical Bayes method to handle dispersion (428, 435, 436). Differential expression analysis generates a list of genes ranked by their statistical significance and fold-change values (436). Then, to account for the large number of hypotheses tested simultaneously (i.e., thousands of genes), multiple testing correction methods such as FDR are applied to control for the false-positive rate (436). Finally, and most importantly, the results are interpreted, and differentially expressed genes are further analyzed to elucidate their potential roles in the studied biological.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) is a computational method used in systems biology to determine whether predefined sets of genes exhibit statistically significant differences in expression between two or more biological states or conditions (445). Unlike traditional differential gene-expression analysis, which focuses on individual genes, GSEA evaluates the coordinated expression changes of groups of functionally related genes known as gene sets or pathways (445). GSEA is particularly useful for analyzing high-throughput gene expression data, such as microarray or RNA-seq, and for identifying biological pathways or processes that are dysregulated in disease states, drug treatments, or experimental conditions. It provides a more holistic view of gene expression changes compared to traditional gene-wise analysis and can reveal coordinated changes in gene expression that may be missed by focusing solely on individual genes. The main steps of GSEA are obtaining curated gene sets, which are collections of genes sharing common biological functions, pathways, or regulatory mechanisms. These are obtained from databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), or the Human Molecular Signatures Database (445-448).

Genes are then ranked based on their differential expression between experimental conditions using a metric such as fold change or the t-statistic (445). GSEA calculates an enrichment score for each gene set to assess whether genes in the set are enriched at the top or bottom of the ranked gene list (445). This score reflects the degree of correlation between the gene set and biological group. The scores are then normalized by gene set size, which yields a normalized gene-enrichment score (445). The significance of gene set enrichment is assessed using permutation-based methods (445). Multiple testing correction is applied to control for false-positive findings using FDR. This results in enriched gene sets that are interpreted to gain insights into the underlying biological processes, pathways, or molecular mechanisms that are dysregulated between experimental conditions. GSEA provides a systems-level view of gene expression changes and can reveal coordinated changes in pathways or biological processes that may be missed by traditional DGE analysis.

Ethical considerations

> *"Ethics is knowing the difference between what you have a right to do and what is right to do"*
>
> — Potter Stewart

All participants in the involved cohorts provided written informed consent prior to their inclusion. Patients have the option to withdraw at any time for any reason and are not obligated to provide justification for their withdrawal. The studies presented in this thesis adhere to the principles outlined in the Declaration of Helsinki. The fundamental tenets of this declaration emphasize that the pursuit of new knowledge must never supersede the rights and interests of individual research subjects. This underscores that research should always prioritize the well-being of patients and ensure that they are not subject to exploitation. Another crucial aspect of the declaration emphasizes the significance of the research objective and requires that the potential benefits outweigh the risks and burdens imposed on the research subjects. This implies that a study's importance and the potential benefits for participating patients should justify the acceptance of higher risks and burdens, especially in cases where the research aims to make significant advancements, such as discovering new cures for diseases. The objective of the studies in this thesis is to evaluate the prognostic and treatment-predictive capabilities of two biomarkers, CAV1 and IGFBP7. The hope is that the results will help to improve treatment and care, so the benefits should outweigh risks. The participants were not offered any compensation for study participation.

The majority of patient cohorts in this thesis were observational and involves no interventions, thus posing minimal health risks to participants. An exception is the I-SPY2 clinical trial, where a significant number of participants received an experimental drug, which elevated the potential risks compared to those in observational studies. Participants may experience unforeseen and sometimes severe side-effects due to the experimental treatment, and the treatment itself might prove unsuccessful, exposing them to downsides without benefits. However, the potential reward is substantial as the experimental treatment holds promise as a potential cure that may not be otherwise achievable. Clinical trials represent a unique opportunity for research to directly impact individual patients. It is noteworthy that I-SPY2 participants face a high risk of recurrence, which makes the exploration of treatments to reduce recurrence-risk particularly valuable. Additionally, the I-SPY2 consortium has a history of identifying highly successful experimental drugs, such as olaparib, durvalumab, and pembrolizumab, which are now integrated into clinical practice.

All of the involved cohort studies managed sensitive patient data obtained from diverse sources, like epidemiological questionnaires, patient charts, registries, and

biological samples. Handling personal data requires the utmost care to ensure privacy protection and confidentiality. The datasets are pseudonymized to minimize potential risks. Nevertheless, even if personal data have undergone de-identification, encryption, or pseudonymization, the potential to re-identify an individual means the data remain classified as personal data. Data are only considered non-personal when the individual is not identifiable, which is achieved through irreversible anonymization. When conducting biomarker studies, it is crucial to exercise caution to avoid depleting biological material that may be required for future clinical prognostication. For instance, samples for TMA and RNA-seq are only obtained when the clinical pathologist deems that there is a sufficient amount of spare tumor material. The current study used some tumor material that had already been collected, but no additional tumor material was gathered. It is once again crucial to meticulously plan the research to avoid unnecessary depletion of collected biological samples.

A unique challenge in preserving patient privacy arises with the concept of open data, which is a crucial element in this thesis. Open data serve as a valuable resource for researchers and facilitate swift access to high-quality information, but they necessitate responsible use. In an era where research demands significant resources and time, open data stand out as an efficient means to harness the contributions of patients who volunteer for studies. Although it offers great potential, open data require careful consideration, and demand both methodological and subject-specific expertise for identification and analysis. Researchers bear a heightened responsibility to uphold ethical standards and refrain from any misuse or attempts to identify individual patients. Furthermore, there is a risk of neglecting patient material when the original principal investigator is no longer involved, which emphasizes the need for sustained attention to this valuable resource.

Prior to commencement, any research involving human participants or animals must obtain approval from an ethics committee or institutional review board. This approval is meant to ensure the protection of individual rights and verify that the studies adhere to ethical, legal, and regulatory norms and standards.

The BC-Blood cohort was approved by the Lund University Ethics Committee (Dnr 75-02, Dnr 37-08, Dnr 658-09, Dnr 58-12, Dnr 379-12, Dnr 227-13, Dnr 277-15, and Dnr 458-15).

SCAN-B was approved by the Lund University Ethics Committee with the following applicable approvals (Dnr 658-09, Dnr 277-15, Dnr 58-12, and Dnr 01252-19).

TCGA was approved by each of the respective institutional review board at each tissue-source site (207, 449). The institutional review boards also approved submission of cases to TCGA (207, 449).

The METABRIC study was approved by the ethics committees at the University of Cambridge and the British Columbia Cancer Research Centre (212, 234, 235).

The I-SPY2 trial was approved by each respective institutional review board of the participating sites. In this trial, patients signed consent forms when screened for the trial and provided additional consent to continue after treatment allocation but before the start of the treatment. The treatment was open-label, so the patients knew which treatment they were receiving when they consented for the second time.

Results and discussion

"Success is stumbling from failure to failure with no loss of enthusiasm"

— Winston Churchill

Caveolin-1 as a biomarker: Results

Paper I

Positive cytoplasmic staining of CAV1 in malignant cells was associated with unfavorable metabolic profiles, including large breast volumes and younger age. Conversely, strong CAV1 staining in stromal cells was linked to younger age. Positive cytoplasmic staining of CAV1 in malignant cells was associated with unfavorable tumor characteristics, whereas strong CAV1 staining in stromal cells was associated with favorable tumor characteristics, such as ER-positivity and low histological grade. Combined CAV1 status showed associations similar to positive CAV1 staining in malignant cells. In the TCGA dataset, CAV1 gene and protein expression were positively correlated and showed associations with non-luminal subtypes and pathways related to cell cycle control, inflammation, and the IGF/insulin system.

The localization of CAV1 was associated with neither the incidence of any breast cancer event nor distant metastasis. However, positive cytoplasmic staining of CAV1 in malignant cells was associated with increased incidence of contralateral breast cancer, and stromal CAV1 was associated with increased incidence of locoregional recurrence. The combined CAV1 status did not provide additional prognostic information beyond the individual localizations of CAV1.

The association between strong staining of CAV1 in stromal cells and the incidence of any breast cancer event was modified by several prognostic factors, including BMI and invasive tumor size. Strong staining of CAV1 in stromal cells was associated with increased incidence of any breast cancer events in normal-weight patients, but not in overweight patients. Also, strong staining of CAV1 in stromal cells was associated with an increased incidence of breast cancer events in patients with small tumors (pT1), but not in patients with larger tumors (pT2/3/4). The

results highlight that CAV1 in stromal cells is associated with the incidence of any breast cancer event, even among supposedly low-risk patients.

Paper II

Database searches revealed that all five *CAV1* SNPs were correlated with other genetic variants in *CAV1* and regulate its expression in adipocytes. Specifically, the rs3807989 A-allele and rs3815412 C-allele were genotypes linked to lower CAV1 gene expression. Tumor-specific CAV1 in stromal cells and in malignant cells were stable across *CAV1* genotypes and haplotypes.

Carriers of the rs3815412 CC genotype exhibited an elevated incidence of contralateral breast cancer. No interaction was observed between the rs3815412 SNP and tumor-specific CAV1 status in malignant cells regarding the incidence of contralateral breast cancer. Among the five common haplotypes, only the TTACA haplotype was associated with outcomes. Possessing at least one copy of the TTACA haplotype was associated with increased incidence of locoregional recurrence. This association was more pronounced in patients not treated with radiotherapy compared to those treated with radiotherapy. Furthermore, no interaction was observed between tumor-specific CAV1 in stromal cells and the TTACA haplotype regarding the risk of locoregional recurrence. After additional adjustment for BMI, HER2 status, and CAV1 status, the effect estimates are essentially the same in both complete-case and multiple-imputation models.

Paper III

In all databases investigated, *CAV1* gene expression was most highly expressed in normal-like subtypes, followed by luminal A PAM50 subtypes. There was also an inverse association with the ROR category in all cohorts. The distribution of *CAV1* gene expression was similar across the TNBC (Lehmann) subtypes in all cohorts, with the highest *CAV1* expression in the MSL subtype, followed by the M subtype.

Depending on spatial localization, strong CAV1 protein staining was associated with different clinicopathological and molecular characteristics. Strong CAV1 protein staining in malignant cells was associated with higher histological grade but no axillary lymph-node involvement. The opposite was seen for strong CAV1 protein staining in stromal cells. Strong CAV1 staining in malignant cells was positively associated with the M subtype and negatively associated with the IM subtype. Strong CAV1 staining in stromal cells was positively associated with the LAR. CAV1 in either malignant nor stromal cells were not strongly correlated with CAV1 gene expression in the tumors or clinical outcome. The combined CAV1 status was also not associated with CAV1 gene expression or clinical outcomes.

In *CAV1*-high tumors, genes related to cellular lipid metabolism, endothelial cells, platelet activation, and vascular homeostasis were upregulated. Hallmark signatures enriched in *CAV1*-high tumors included EMT, adipogenesis, coagulation, angiogenesis, and hypoxia. In *CAV1*-low tumors, the G2M checkpoint, E2F targets, interferon alpha and beta response, and MYC targets were enriched, which suggests increased proliferation and immune response in these tumors. *CAV1* was most highly expressed in stromal cells in endothelial cells, followed by perivascular-like cells and CAFs, while it was weakly expressed in malignant cells and barely expressed at all in immune cells. *CAV1*-high tumors had higher relative abundance of endothelial and stromal cells compared to CAV1-low tumors. Additionally, *CAV1*-high tumors were associated with dominance of carcinoma ecotype (CE)6, followed by CE1. This indicates that *CAV1*-high tumors have an enriched microenvironment for stromal cells while being deficient in immune cells. *CAV1* gene expression was correlated to several cell states of fibroblast and endothelial cells, supporting a potential role in an active stromal component in TNBC that promotes vascularization and EMT, as well as suppressing the immune response.

Further analysis of GOBO revealed that patients with ER-negative tumors with high *CAV1* expression had shorter distant metastasis-free survival in univariable and multivariable analyses. The difference in the distant metastasis-free interval was especially apparent in the subset of tumors classified as basal, which implied that *CAV1* expression is a potential prognostic marker in TNBC. In the multivariable analyses, *CAV1*-high tumors in SCAN-B conferred an increased incidence of recurrence, distant metastasis, and mortality. Likewise, in the GSE31915 cohort, *CAV1*-high tumors had shorter event-free survival.

Caveolin-1 as a biomarker: Discussion

CAV1 has been found to influence prognosis at the genomic, transcriptomic, and proteomic levels, although the associations varied somewhat. These findings indicate that the prognostic impact of CAV1 is highly reliant on context, and its specific association with prognosis is contingent upon host factors and tumor characteristics. Notably, recent research conducted by our group uncovered that elevated *CAV1* gene expression is linked to a significantly unfavorable prognosis among patients with PAM50 ROR high tumors (450). Furthermore, *CAV1* mRNA expression levels substantially modify the prognostic significance provided by PAM50 ROR, which highlights the context-dependent nature of CAV1 (450). Studies within this thesis also support the involvement of CAV1 in hypoxia, inflammation, lipid metabolism, and EMT across various contexts. A pivotal aspect linking these diverse (patho)physiological processes is the TME (61, 62). Consistent with this, CAV1 exhibited heightened expression in stromal cells, which is consistent with previous research (354, 356). The correlation observed between

normal-like, mesenchymal stem-like, and stroma modules with *CAV1* gene expression further emphasizes its strong association with stromal cells and an active TME (354, 356). Findings from papers I and III corroborate the notion that CAV1 is highly expressed in the TME, with CAV1 staining generally being stronger in stromal cells compared to malignant cells. Moreover, mRNA expression of CAV1 in the single-cell atlas of breast cancer also indicated stronger expression in stromal/endothelial cells compared to malignant cells.

The consistent correlation observed across all three studies underscores the key relationship between CAV1 and lipid metabolism in breast cancer cohorts. It is widely recognized that CAV1 actively participates in cholesterol transport and the formation of lipid droplets (354, 363). These lipid droplets serve as major regulators of lipid metabolism (367). Loss of CAV1 in adipose tissue results in improper fat storage, which leads to lipodystrophy, insulin resistance, hypertriglyceridemia, and metabolic syndrome without a corresponding increase in adiposity (451, 452). CAV1 deficiency in adipose tissue also triggers the recruitment of M2 macrophages, which foster obesity-related inflammation that promotes tumorigenesis. It is evident that CAV1 plays a crucial role in regulating adipose tissue, which is central to the development of metabolic syndrome and obesity (354, 356, 451), as indicated in all three papers.

In particular, the two genotypes associated with a heightened risk for non-distant events correlate with reduced *CAV1* gene expression in adipocytes due to LD with expression quantitative trait loci of the CAV1 gene. Decreased CAV1 levels lead to increased aromatase expression, thereby elevating estrogen levels in surrounding tissues and promoting breast cancer tumorigenesis (453). Furthermore, CAV1 deficiency impairs insulin receptor stabilization, which results in insulin resistance and inflammation (454). Additionally, CAV1 has been implicated in promoting tumorigenesis by facilitating LDL uptake and contributing to the formation and stabilization of lipid droplets, thus sustaining tumor-cell proliferation under adverse conditions (354).

CAV1 plays a crucial role in maintaining the membrane integrity of tumor cells and regulates lipid metabolism and fatty acid oxidation (354, 356). Loss of CAV1 impairs lipid storage and metabolic processes, including the Warburg effect, which is essential for tumor survival (363). Elevated CAV1 expression stimulates glucose uptake and ATP production, whereas its knockdown suppresses the Warburg effect (363). Moreover, CAV1 interacts with insulin and IGF-1 receptors, thereby enhancing glucose uptake and lactate output through AKT signaling (386). Furthermore, CAV1 regulates the switch between glucose-dependent mitochondrial respiration and aerobic glycolysis, as well as lipid-dependent energy metabolism, which are crucial for tumor survival (363, 455). These various processes, such as inflammation, hyperinsulinemia, and altered metabolism (354, 356, 363), all contribute to an increased metastasis rate and may partially explain why CAV1 serves as a prognostic factor.

Our findings in two of the studies suggest that CAV1 is highly expressed in endothelial cells, which is consistent with prior research. Endothelial-tumor crosstalk is crucial for both tumor growth and intravasation, which is a pivotal step that is necessary for metastasis (19, 26, 61). The tumor vasculature plays a vital role in promoting metastasis, and the intravasation of malignant cells is a critical event that is required for metastatic dissemination (19, 26, 61). It is perhaps not surprising that CAV1 is correlated with an increased incidence of distant metastasis. However, the intricacies of this process remain incompletely characterized, and it is unclear how CAV1 is specifically related to it, which warrants further investigation.

There is evidence closely related to this that CAV1 plays a significant role in angiogenesis. It has been proposed that CAV1 modulates angiogenesis and neovascularization in response to ischemia through the regulation of vascular endothelial growth factor (VEGF)-dependent endothelial nitric oxide synthase (eNOS) activation in endothelial cells (354, 356, 456). Ischemia induces hypoxia, a condition in which CAV1 is evidently implicated (354, 356, 456). HIF1α and HIF2α directly target CAV1 as a transcriptional target, leading to metabolic reprogramming through the attenuation of MYC expression (457). Specifically, HIF1α is also recognized as an adverse prognostic indicator in breast cancer cases (458, 459). Therefore, exploring the relationship between CAV1 and HIF1α could provide valuable insights. Since (neo)vascularization is a hallmark of cancer and essential for tumor survival, it provides an additional potential explanation for why CAV1 serves as a negative prognostic factor (15).

There was also compelling evidence across the studies indicating that CAV1 is closely associated with EMT, which is consistent with the literature (354, 356). Previous research has implicated CAV1 with TGF-β in that it reprograms the TGF-β signaling pathway, which has effects raging from suppressing tumor formation to promoting oncogenesis (387, 457). Our findings revealed high expression of CAV1 in breast cancers classified as normal-like and mesenchymal, which are characterized by elevated expression of EMT-related genes and TGF-β signaling, thus supporting this notion (207, 244). Moreover, in the single-cell atlas of human breast tissue, CAFs enriched in EMT features and myogenesis exhibited a correlation with CAV1 expression, further suggesting a strong association with EMT (61). This consistent pattern aligns with CAV1 being recognized as a marker of metastasis. Furthermore, it is well-documented that CAV1 expression induces extracellular matrix remodeling, which facilitates metastasis (460). Notably, CAV1 expression was associated with CE2, which is implicated in extracellular matrix-related remodeling and fibrosis (461). In general, it can be inferred that CAV1 is highly correlated with an ecosystem characteristic of cancer marked by the enrichment of stromal features and cells, which is associated with a worse prognosis in breast cancer.

A particularly intriguing finding was the association between CAV1 polymorphisms and CAV1 protein expression with the same endpoint, such as locoregional

recurrence and contralateral breast cancer. Surprisingly, these associations were found to be independent of each other. However, further validation is required to confirm these findings, and additional studies are needed to thoroughly investigate the relationship between genomic CAV1 and CAV1 expression in breast cancer tumors. Equally compelling is the discovery that CAV1 protein expression levels in malignant cells were predictive of the development of contralateral breast cancer after diagnosis, which warrants further investigation. Potentially, the reported negative correlation with *BRCA1/2* might provide insights.

As previously discussed, the regulation of CAV1 protein levels primarily occurs through epigenetic mechanisms, particularly hypermethylation of the CAV1 promoter region in cancer (357). It has been noted previously that mRNA and protein levels of CAV1 do not correlate well (357). This observation may help explain the findings in paper III, but not those in paper I. However, it is important to note that CAV1 protein expression was not assessed in endothelial cells, where it is abundantly expressed (354, 387). Therefore, one should interpret the results in paper III within the context of the phenotype related to tumors with high *CAV1* gene expression.

The primary limitations of these studies lie in the absence of additional validation in independent cohorts to substantiate the findings. Although these studies are hypothesis-driven and largely consistent with one another, it is essential to recognize that they entail retrospective analyses of prospective study cohorts, which necessitates prospective confirmation. Additionally, due to the observational nature of these studies, residual confounding factors cannot be entirely ruled out. It would be valuable to compare different predictive models employing various validated biomarkers and prognostic tools to assess how CAV1 fares relative to them and to determine the optimal approach.

IGFBP7 as a biomarker: Results

Paper IV

This study examined the levels of tumor-specific IGFBP7 and its gene expression in relation to clinicopathological factors and prognosis in breast cancer. Low tumor-specific IGFBP7 protein levels were associated with prior menopausal hormone therapy and less aggressive tumor characteristics, while higher levels were linked to more aggressive features such as ER negativity, PR negativity, and higher histological grade. *IGFBP7* gene expression showed moderate positive correlations with other IGFBPs and *IGF1* while weakly negatively correlating with age and *ESR1*.

Low tumor-specific IGFBP7 levels were associated with a lower incidence of any breast cancer event compared to intermediate and high IGFBP7 levels. Similar patterns were observed concerning the incidence of distant metastasis, but not overall survival. Interactions were found between high tumor-specific IGFBP7 and alcohol abstinence, ER status, and tamoxifen treatment, which suggests varying prognostic implications based on these factors. In alcohol abstainers, elevated tumor-specific IGFBP7 protein levels were linked to a higher 10-year incidence of any breast cancer event. Conversely, among alcohol drinkers, higher IGFBP7 levels were associated with a decreased incidence of any breast cancer event. High IGFBP7 levels conferred with a somewhat lower 10-year incidence of distant metastasis in patients with ER+ tumors. In patients with ER+ tumors, elevated tumor-specific IGFBP7 levels were correlated with a reduced 10-year incidence of any breast cancer event in those treated with tamoxifen, but not in those who did not receive tamoxifen treatment.

Paper V

In both ISPY-2 and SCAN-B, *IGFBP7* gene expression was correlated with *IGFBP3-6* and *IGF1* and *IGF2* expression. *IGFBP7* gene expression was highest in the normal-like subtype, followed by the luminal A subtype in both the ISPY-2 trial and SCAN-B cohort. Likewise, *IGFBP7* expression was positively correlated with stroma, lipid, and early response to growth signaling and negatively correlated with mitotic checkpoint and progression gene modules. *IGFBP7* expression was stable across IHC breast cancer subtypes.

In all patients (across all treatment arms), *IGFBP7* expression was not associated with the odds of achieving pCR. There was, however, an interaction between *IGFBP7* expression and efficacy of ganitumab/metformin plus chemotherapy treatment in achieving pCR. Higher *IGFBP7* gene expression conferred lower odds of achieving pCR in the arm receiving ganitumab/metformin plus chemotherapy, but not in the control arm receiving chemotherapy alone. When divided by breast cancer subtype (high-risk HR-positive/HER2-negative versus. TNBC), the ability of *IGFBP7* expression to identify breast cancers more likely to respond to ganitumab/metformin plus chemotherapy than to chemotherapy alone was more apparent in TNBC. The improved efficacy of ganitumab/metformin plus chemotherapy treatment compared to standard chemotherapy in achieving pCR was confined to the approximately 25% of patients in the lowest quartile of *IGFB7* expression.

In SCAN-B, after adjustment for age, clinicopathological factors, and treatment in the multivariable models, high *IGFBP7* expression was associated with increased incidence of recurrence and distant metastasis. Notably, higher expression of several genes coding for proteins involved in endothelial cell regulation and extracellular matrix remodeling were seen in tumors with the highest *IGFBP7* expression, which

supports a potential association with a tumor-promoting TME. Significantly, hallmarks in these tumors included EMT, TGF-β signaling, coagulation, and angiogenesis, and downregulated hallmarks included MYC response.

IGFBP7 as a biomarker: Discussion

In summary, the results suggest that IGFBP7 may serve as a marker of poor prognosis in breast cancer. Additionally, *IGFBP7* gene expression exhibited the potential to predict treatment response to IGF-1R targeting monoclonal antibody. The associations with prognosis were generally consistent across the two studies. *IGFBP7* expression displayed stable associations with clinicopathological factors and molecular features across SCAN-B and ISPY-2, which are typically associated with somewhat favorable clinicopathological factors, although no strong associations were evident. Conversely, IGFBP7 protein levels, as opposed to *IGFBP7* gene expression, were strongly associated with unfavorable clinicopathological factors. This contrast between protein and gene expression levels of *IGFBP7* suggests that at the proteomic and transcriptomic levels, IGFBP7 represents different biomarkers to some extent. However, no analyses were conducted to confirm whether mRNA and protein levels of IGFBP7 are correlated, which would be crucial for future studies to explore.

The molecular characteristics of tumors exhibiting high IGFBP7 expression of both mRNA and protein suggest an aggressive TME that is conducive to metastasis. This aligns with the findings that both tumor-specific IGFBP7 protein levels and *IGFBP7* gene expression serve as poor prognostic markers in breast cancer. The results from paper V strongly indicate that *IGFBP7* expression can effectively classify distinct subtypes of the breast cancer microenvironment. However, there remain significant gaps in our understanding of how IGFBP7 modulates signaling by the IGF-1R family and potentially impacts the efficacy of anti-IGF-1R antibodies. Our findings underscore the need for further experimental investigations to elucidate how IGFBP7 influences the effectiveness of ganitumab, its biological effects on the TME, and how it may be used as a treatment-predictive marker for IGF-1R targeting agents if our findings are confirmed. Such endeavors are crucial for advancing our understanding of the role of IGFBP7 in breast cancer progression and its potential as a treatment-predictive and therapeutic target.

In cardiovascular disease, IGFBP7 has been identified as capable of inducing senescence and inflammation through the IGF-1R/IRS/AKT signaling axis (402). Cellular senescence is recognized as a new cancer hallmark ("enabling characteristic"), while inflammation is a long-established hallmark, and both play pivotal roles in this process (15, 20). Cells undergoing senescence develop a senescence-associated secretory phenotype, which becomes more abundant in

various organs during aging (38, 462). Moreover, it is well-documented that cardiovascular disease can modulate tumor immunity and inflammation (463, 464).

Traditionally, cellular senescence has been viewed as protective against neoplasia (20), but mounting evidence suggests otherwise (38). Senescent cells can paradoxically promote tumors through paracrine signaling, which contributes to proliferative signaling, evading apoptosis, inflammation, inducing angiogenesis, stimulating invasion and metastasis, and suppressing tumor immunity (38). Notably, the senescent state is extensively documented as a marker of therapy resistance (462). Senescence in CAFs has been shown to be tumor-promoting by conferring hallmark capabilities to cancer cells in the TME (462). Our previous findings have demonstrated a link between *IGFBP7* mRNA expression, therapeutic resistance, inflammation, and CAF activation in the TME, which are strongly associated with a senescent phenotype (38, 462). Further studies elucidating the interplay between senescence, IGFBP7, cardiovascular disease, and cancer are warranted to comprehensively understand their complex relationships and implications.

The strengths of the studies lie in the utilization of data from two prospective population-based cohorts, BC-blood, and SCAN-B, which enables the evaluation of biomarkers in a contemporary real-world setting (421). Furthermore, the I-SPY2 trial assessed ganitumab in a randomized controlled setting that offers an ideal platform for investigating treatment-specific biomarkers for ganitumab. The results from I-SPY2 also suggest that IGFBP7 warrants further investigation as a treatment-predictive biomarker in other cancer types, such as colon, ovarian, and prostate cancer, as well as sarcomas, where IGF-1R targeting agents have been explored (465, 466). Independent validation in another clinical trial evaluating IGF-1R targeting agents is also crucial.

However, many of the same limitations present in papers I-III also apply here. The absence of additional validation in independent cohorts to corroborate the findings is notable. While these studies are hypothesis-driven and generally consistent with one another, it is imperative to acknowledge that they involve retrospective analyses of prospective study cohorts, necessitating prospective confirmation. Additionally, due to the observational nature of these studies, residual confounding cannot be entirely ruled out.

Strengths and limitations

Paper	Strengths	Limitations
I Levels of CAV1 protein expression	Prospective population-based cohort study with relatively long follow-up. Body measurements were obtained from research nurses. Use of a well validated antibody.	Residual confounding due to the observational nature of the study. Staining was evaluated on TMAs and the final scoring was not confirmed on whole section slides.
II *CAV1* polymorphisms	Prospective population-based cohort study with relatively long follow-up. Data on CAV1 on both a genomic and proteomic level.	Residual confounding due to the observational nature of the study. No independent validation.
III CAV1 in TNBC	Consistent associations with molecular features and clinicopathological features across datasets. *CAV1* gene expression as a prognostic marker is validated in another independent cohort.	Residual confounding due to the observational nature of the study. The protein and gene expression of CAV1 was not strongly correlated.
IV Levels of IGFBP7 protein expression	Prospective population-based cohort study with relatively long follow-up. Use of a well-validated antibody. Comparatively large set of tumors. Use of multivariable analysis to adjust for potential confounding.	Residual confounding due to the observational nature of the study. Staining was evaluated with TMAs, and the final scoring was not confirmed on whole section slides.
V *IGFBP7* gene expression as a predictor for IGF-1R targeting agents	The use of RCT to evaluate a treatment-predictive marker for ganitumab. Consistent associations with molecular features and clinicopathological features across datasets. Consistent association with prognosis.	Residual confounding due to the observational nature of the study. No validation of the treatment-predictive ability of *IGFBP7* gene expression.

Conclusions

"It is never too late to be what you might have been"

— George Eliot

Paper I

- The prognostic impact of CAV1 protein expression depended on its localization, anthropometric, and tumor factors
- CAV1 in malignant cells predicted high recurrence risk in a group of patients with small body size and tumors that supposedly had "low risk" based on current clinical criteria
- CAV1 in malignant cells predicted metachronous contralateral disease

Paper II

- *CAV1* polymorphisms were associated with an increased risk for locoregional recurrence and contralateral breast cancer
- The association between *CAV1* polymorphisms and clinical outcome was not modified by CAV1 protein expression in the tumor

Paper III

- High *CAV1* gene expression was an independent prognostic factor in TNBC
- The molecular features of *CAV1* gene expression suggest a role in chemoresistance and a tumor promoting TME

Paper IV

- Both IGFBP7 protein levels and gene expression showed similar associations with clinicopathological factors
- Low levels of tumor-specific IGFBP7 were a potential marker of good prognosis
- The association between high levels of tumor-specific IGFBP7 and prognosis was dependent on host factors and treatment

Paper V

- A subset of breast cancer patients that have a good response to ganitumab can be identified by low *IGFBP7* gene expression
- High *IGFBP7* expression was predictive of poor outcome in breast cancer

Future perspectives

"Life can only be understood backwards; but it must be lived forwards"

— Søren Kierkegaard

This thesis investigated and to some extent validated two potential biomarkers, CAV1 and IGFBP7, which both seem to have relevance in the context of breast cancer, especially for prognosis. However, the findings are still premature, and there is still a long way before the findings (if true) could be implemented in the clinic. Both CAV1 and IGFBP7 could be proposed as potentially interesting biomarkers, especially for TNBC, which merits further study. This opens the door to several essential steps in the biomarker discovery pipeline.

Firstly, additional molecular and genomic characterization of these biomarkers is warranted. Understanding somatic mutations in key oncogenes, copy-number aberrations, methylation patterns, and metabolic processes related to CAV1 and IGFBP7 could provide valuable insights. Furthermore, further prognostic validation in other cohorts and datasets is necessary to confirm the relationship between CAV1, IGFBP7, and prognosis, particularly at the genomic and proteomic levels. Additionally, retrospective analysis of clinical trials would be valuable to investigate or validate the potential clinical implications of these biomarkers. It is also essential to explore the biological mechanisms underlying the association of IGFBP7 and CAV1 with an increased incidence of distant metastasis, as well as their role in the tumor microenvironment. Further investigation into why IGFBP7 is treatment predictive for IGF-1R targeting agents is also warranted. The potential of CAV1 as a treatment target, particularly for statins, should be explored through clinical cohorts, preferably in clinical trials, as well as mechanistic studies. This thesis has also provided evidence supporting the idea that biomarkers should be examined within the context of the host, and this interaction warrants further investigation. Both biomarkers had previously been associated with the TME, and this thesis contributed additional evidence to this concept. Perhaps, the TME can be viewed as a mediator between malignant cells and the host as it represents an intermediate state where cells are phenotypically different but lack the genomic characteristics of cancer cells. In summary, this thesis has laid the groundwork for the continued development of two promising biomarkers for clinical use. Through further characterization, refinement, and validation, CAV1 and IGFBP7 have the potential to enhance personalized medicine in the context of breast cancer treatment.

Acknowledgements

"When eating fruit, remember the one who planted the tree"

— A Vietnamese proverb

Thank you to everyone who has read this thesis—I am grateful for your interest in my work. Pursuing a PhD is challenging and not something one can do alone. I am deeply grateful for the support I received throughout my studies. It has been a long and sometimes difficult journey, at times seeming insurmountable. I would particularly like to acknowledge:

First and foremost, my main supervisor, *Helena Jernström*. *Helena*, thank you for believing in me when I was just a young medical student and recognizing my potential. I now realize how little I knew back then, but you saw something in me and introduced me to the world of science. For that, I am forever grateful. Your great knowledge of epidemiology, your critical questions on each project, and our impromptu scientific discussions have been invaluable. Thank you also for your commitment and encouragement to my education, which have provided me with the opportunity to explore new fields while keeping me focused and on track.

To my co-supervisors, *Ana Bosch,* and *Karolin Isaksson—Karolin*, your support and encouragement throughout my PhD years is much appreciated, always reminding me to maintain a clinical perspective in my projects. *Ana*, your good ideas have significantly improved my PhD project. I admire your dedication and enthusiasm for improving breast cancer care and am grateful for your willingness to share your vast knowledge, spanning both clinical and pre-clinical domains.

I would like to extend a heartfelt thank you to both past and present members of the Jernström Group: *Somayeh, Helga, Linn, Annelie, Jasmine, Linnea, Per Ola, Alexandra, Louise, Floor, Helén, and Linda*. A special thanks to *Somayeh*, for being a great colleague and friend. To *Linn*, a fellow PhD student, with whom I've shared the ups and downs of this journey. I am also grateful to *Linnea* and *Annelie* for the many enjoyable conversations during long workdays.

Additionally, I owe a big thank you to all my co-authors for their invaluable assistance throughout my projects.

Mattias Belting, thank you for initiating the Caveolin-1 projects. Your enthusiasm, interest, and assistance have been incredibly helpful. Your great passion for science is truly inspiring.

Michael N. Pollak, thank you for your invaluable input on Paper V and for recognizing its potential. Your insights have significantly enhanced the work. I have thoroughly enjoyed our many discussions over Zoom, and I am constantly in awe of your vast knowledge and expertise, especially how you always manage to find intriguing new articles about IGFBP7.

Karin Jirström, I am grateful for your prompt responses to my pathology/histology-related questions, your honest feedback on my papers, and your sound advice.

Johan Vallon-Christersson, thank you for all your assistance with the SCAN-B and for helping me understand the intricacies of the database. I also appreciate your valuable input on the papers concerning molecular subtypes and bioinformatics analyses.

Björn Nodin, thank you for making the many hours spent in front of the microscope enjoyable.

Edward Visse, thank you for your help with the TCGA database. *Signe Borgquist*, I am grateful for your valuable input. *Vineesh Indira Chandran* and *Magdalena Barbachowska*, thank you for laying the groundwork with Caveolin-1.

Pär-Ola Bendahl, I appreciate your assistance in answering my statistical questions. Your informative and pedagogical answers have taught me a lot about statistics, and I am grateful for your review of the statistical sections in this thesis and Paper III.

Ann Rosendahl and the BCPI group, thank you for engaging discussions and insightful journal club meetings. *Öyku*, thank you for the stimulating discussions on science and various other topics. *Maria*, I am grateful for our pleasant conversations and your valuable advice on navigating academia.

My colleagues at the Kamprad Building—*Maria, Johanna, Anders, Hugo, Valeria, Hampus H, Myriam, Anna D, Frederica, Hampus dR, Sara B, Marton, Axel, Sten, Sara, Aarya, Ann-Sofie, Maja, Kelin, Hedda, Anna BR, Alexandra, Wahed, Gabriel, and Sara W*—thank you for making it such a nice and inspiring place to work. I would also like to thank my colleagues in the wider Division of Oncology for the enjoyable poster evenings and retreats.

Carina and Marie, thank you for helping with the database updates and for organizing all the social activities at Kamprad.

Sohail Tavazoie, thank you for inviting me to visit your lab, which granted me new perspectives on science and a great collaboration. I have learned a lot from it and look forward to seeing the end results. *Wenbin Mei*, thank you for the fruitful

collaboration and discussions. I have learned a great deal from you, and I appreciate your care during my stay in the US.

The BCLU network, including *Sophia Zackrisson*, *Magnus Dustler*, and *Ylva Bengtsson*, thank you for the productive meetings and scientific discussions. *Ylva*, thank you for the many enjoyable chats about both science and other topics.

Linda Lindström and *Jonas Manjer*, my half-time reviewers, thank you for your valuable suggestions that helped improve my thesis.

Susanne André, thank you for all the administrative support and for the enjoyable impromptu conversations in the kitchen.

Anette Saltin, thank you for help and quick responses regarding research education.

I would like to extend my gratitude to everyone involved in the BC-blood study, especially *Christian Ingvar, Carsten Rose, Åke Borg,* and *Mårten Fernö* who helped initiate this project many years ago. Without your contributions, this thesis would not have been possible.

I would like to thank everyone at MDR for giving me new perspectives, who knew that doctoral student representation could be this fun. In particular, I would like to thank, *Sakshi;* for being an excellent work partner and someone who I could confide with. I am grateful for our friendship. *Divya* and *Esther*, thank you for being such wonderful friends. I truly appreciate all of our engaging discussions, both about academia and other topics. I would also like to extend my gratitude to the many wonderful people who I have meet through MDR, *Vasiliki, Kreema*, *Vibha, Juliane, Sanjay, Nika, Jack, Ana, Rafsan, Radhika* among others, who has made my time enjoyable at MDR.

I want to extend my thanks to everyone who has contributed in various ways to my PhD studies. While it might not have been possible to mention each of you individually in the formal acknowledgements, your support has not gone unnoticed.

Finally, I would like to thank my family and friends for reminding me that there is life beyond my PhD studies. I am especially grateful to my mom, *Monika*, and dad, *Laszlo*, who have been my biggest supporters from day one.

References

1. Odes EJ, Randolph-Quinney PS, Steyn M, Throckmorton Z, Smilg JS, Zipfel B, et al. Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa. S Afr J Sci. 2016;112(7-8):1-5.
2. Faguet GB. A brief history of cancer: age-old milestones underlying our current knowledge database. Int J Cancer. 2015;136(9):2022-36.
3. Hajdu SI. A note from history: Landmarks in history of cancer, part 1. Cancer. 2011;117(5):1097-102.
4. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-49.
5. The Swedish Board of Social Affairs and Health (Socialstyrelsen). Cancer i Siffror 2023. www.socialstyrelsen.se2023.
6. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Nat Rev Dis Primers. 2019;5(1):66.
7. Harbeck N, Gnant M. Breast cancer. Lancet. 2017;389(10074):1134-50.
8. Plevritis SK, Munoz D, Kurian AW, Stout NK, Alagoz O, Near AM, et al. Association of screening and treatment with breast cancer mortality by molecular subtype in US women, 2000-2012. JAMA. 2018;319(2):154-64.
9. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005;353(17):1784-92.
10. Duffy SW, Tabár L, Yen AMF, Dean PB, Smith RA, Jonsson H, et al. Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women. Cancer. 2020;126(13):2971-9.
11. Bosman FT. Integrative Molecular Tumor Classification: A Pathologist's View. In: Boffetta P, Hainaut P, editors. Encyclopedia of Cancer (Third Edition). Oxford: Academic Press; 2019. p. 279-85.
12. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149(5):994-1007.
13. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546-58.
14. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature. 2020;578(7793):122-8.
15. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.

16. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144(5):646-74.
17. Fearon K, Strasser F, Anker SD, Bosaeus I, Bruera E, Fainsinger RL, et al. Definition and classification of cancer cachexia: An international consensus. Lancet Oncol. 2011;12(5):489-95.
18. Massagué J, Ganesh K. Metastasis-initiating cells and ecosystems. Cancer Discov. 2021;11(4):971-94.
19. Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: A hallmark of cancer revisited. Signal Transduct Target Ther. 2020;5(1):28.
20. Hanahan D. Hallmarks of Cancer: New dimensions. Cancer Discov. 2022;12(1):31-46.
21. Butt AJ, Firth SM, Baxter RC. The IGF axis and programmed cell death. Immunol Cell Biol. 1999;77(3):256-62.
22. Lotem J, Sachs L. Control of apoptosis in hematopoiesis and leukemia by cytokines, tumor suppressor and oncogenes. Leukemia. 1996;10(6):925-31.
23. Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. Eur J Cancer. 1997;33(5):787-91.
24. Hanahan D, Folkman J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell. 1996;86(3):353-64.
25. Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. Nat Med. 2006;12(8):895-904.
26. Gerstberger S, Jiang Q, Ganesh K. Metastasis. Cell. 2023;186(8):1564-79.
27. Pavlova NN, Zhu J, Thompson CB. The hallmarks of cancer metabolism: Still emerging. Cell Metab. 2022;34(3):355-77.
28. Warburg O. On respiratory impairment in cancer cells. Science. 1956;124(3215):269-70.
29. Kim R, Emi M, Tanabe K. Cancer immunoediting from immune surveillance to immune escape. Immunology. 2007;121(1):1-14.
30. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability — an evolving hallmark of cancer. Nature Rev Mol Cell Biol. 2010;11(3):220-8.
31. Bakhoum SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and its microenvironment. Cell. 2018;174(6):1347-60.
32. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;140(6):883-99.
33. Qian B-Z, Pollard JW. Macrophage diversity enhances tumor progression and metastasis. Cell. 2010;141(1):39-51.
34. Yuan S, Norgard RJ, Stanger BZ. Cellular plasticity in cancer. Cancer Discov. 2019;9(7):837-51.

35. Kopp JL, von Figura G, Mayes E, Liu F-F, Dubois CL, Morris JP, et al. Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. Cancer Cell. 2012;22(6):737-50.
36. Saghafinia S, Homicsko K, Di Domenico A, Wullschleger S, Perren A, Marinoni I, et al. Cancer cells retrace a stepwise differentiation program during malignant progression. Cancer Discov. 2021;11(10):2638-57.
37. Faget DV, Ren Q, Stewart SA. Unmasking senescence: context-dependent effects of SASP in cancer. Nat Rev Cancer. 2019;19(8):439-53.
38. Lee S, Schmitt CA. The dynamic nature of senescence in cancer. Nat Cell Biol. 2019;21(1):94-101.
39. Serrano-Gomez SJ, Maziveyi M, Alahari SK. Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. Mol Cancer. 2016;15:18.
40. Thienpont B, Van Dyck L, Lambrechts D. Tumors smother their epigenome. Mol Cell Oncol. 2016;3(6):e1240549.
41. Lambert AW, Weinberg RA. Linking EMT programmes to normal and neoplastic epithelial stem cells. Nat Rev Cancer. 2021;21(5):325-38.
42. Okumura S, Konishi Y, Narukawa M, Sugiura Y, Yoshimoto S, Arai Y, et al. Gut bacteria identified in colorectal cancer patients promote tumourigenesis via butyrate secretion. Nat Commun. 2021;12(1):5674.
43. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. Nature. 2020;580(7802):269-73.
44. Helmink BA, Khan MW, Hermann A, Gopalakrishnan V, Wargo JA. The microbiome, cancer, and cancer therapy. Nat Med. 2019;25(3):377-88.
45. Swanton C, Bernard E, Abbosh C, André F, Auwerx J, Balmain A, et al. Embracing cancer complexity: Hallmarks of systemic disease. Cell. 2024;187(7):1589-616.
46. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, et al. Multistep nature of metastatic inefficiency: Dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am J Pathol. 1998;153(3):865-73.
47. Rahrmann EP, Shorthouse D, Jassim A, Hu LP, Ortiz M, Mahler-Araujo B, et al. The NALCN channel regulates metastasis and nonmalignant cell dissemination. Nat Genet. 2022;54(12):1827-38.
48. Ganesh K. Uncoupling metastasis from tumorigenesis. N Engl J Med. 2023;388(7):657-9.
49. Ye X, Weinberg RA. Epithelial–mesenchymal plasticity: A central regulator of cancer progression. Trends Cell Biol. 2015;25(11):675-86.
50. Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014;158(5):1110-22.

51. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. 2013;339(6119):580-4.
52. Placke T, Örgel M, Schaller M, Jung G, Rammensee H-G, Kopp H-G, et al. Platelet-derived MHC class I confers a pseudonormal phenotype to cancer cells that subverts the antitumor reactivity of natural killer immune cells. Cancer Res. 2012;72(2):440-8.
53. Paget S. The distribution of secondary growths in cancer of the breast. Lancet. 1889;133(3421):571-3.
54. Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, et al. Pre-metastatic niches: Organ-specific homes for metastases. Nat Rev Cancer. 2017;17(5):302-17.
55. Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, et al. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. N Engl J Med. 2017;377(19):1836-46.
56. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15(2):81-94.
57. Beeghly GF, Shimpi AA, Riter RN, Fischbach C. Measuring and modelling tumour heterogeneity across scales. Nat Rev Bioeng. 2023;1(10):712-30.
58. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: Past, present, and the future. Cell. 2017;168(4):613-28.
59. 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(6797):747-52.
60. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869-74.
61. de Visser KE, Joyce JA. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. Cancer Cell. 2023;41(3):374-403.
62. Bejarano L, Jordão MJC, Joyce JA. Therapeutic targeting of the tumor microenvironment. Cancer Discov. 2021;11(4):933-59.
63. Yuan S, Almagro J, Fuchs E. Beyond genetics: Driving cancer with the tumour microenvironment behind the wheel. Nat Rev Cancer. 2024.
64. Hart JR, Zhang Y, Liao L, Ueno L, Du L, Jonkers M, et al. The butterfly effect in cancer: A single base mutation can remodel the cell. Proc Natl Acad Sci U S A. 2015;112(4):1131-6.
65. Alonso-Curbelo D, Ho Y-J, Burdziak C, Maag JL, Morris IV JP, Chandwani R, et al. A gene–environment-induced epigenetic program initiates tumorigenesis. Nature. 2021;590(7847):642-8.
66. Yuan S, Stewart KS, Yang Y, Abdusselamoglu MD, Parigi SM, Feinberg TY, et al. Ras drives malignancy through stem cell crosstalk with the microenvironment. Nature. 2022;612(7940):555-63.

67. Klemm F, Maas RR, Bowman RL, Kornete M, Soukup K, Nassiri S, et al. Interrogation of the Microenvironmental Landscape in Brain Tumors Reveals Disease-Specific Alterations of Immune Cells. Cell. 2020;181(7):1643-60.e17.
68. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, et al. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol. 2010;49(5):725-36.
69. Taylor C, McGale P, Probert J, Broggio J, Charman J, Darby SC, et al. Breast cancer mortality in 500 000 women with early invasive breast cancer diagnosed in England, 1993-2015: population based observational cohort study. BMJ. 2023;381:e074684.
70. van der Meer DJ, Kramer I, van Maaren MC, van Diest PJ, C. Linn S, Maduro JH, et al. Comprehensive trends in incidence, treatment, survival and mortality of first primary invasive breast cancer stratified by age, stage and receptor subtype in the Netherlands between 1989 and 2017. Int J Cancer. 2021;148(9):2289-303.
71. Nordenskjöld AE, Fohlin H, Arnesson LG, Einbeigi Z, Holmberg E, Albertsson P, et al. Breast cancer survival trends in different stages and age groups–a population-based study 1989–2013. Acta Oncol. 2019;58(1):45-51.
72. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): An analysis of population-based cancer registry data. Lancet. 2011;377(9760):127-38.
73. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. Lancet. 2005;365(9472):1687-717.
74. Sopik V, Narod SA. The relationship between tumour size, nodal status and distant metastases: on the origins of breast cancer. Breast Cancer Res Treat. 2018;170(3):647-56.
75. van Maaren MC, Strobbe LJA, Smidt ML, Moossdorff M, Poortmans PMP, Siesling S. Ten-year conditional recurrence risks and overall and relative survival for breast cancer patients in the Netherlands: Taking account of event-free years. Eur J Cancer. 2018;102:82-94.
76. Gusterson BA, Stein T, editors. Human breast development. Semin Cell Dev Biol; 2012: Elsevier.
77. Inman JL, Robertson C, Mott JD, Bissell MJ. Mammary gland development: cell fate specification, stem cells and the microenvironment. Development. 2015;142(6):1028-42.
78. Paine IS, Lewis MT. The terminal end bud: the little engine that could. J Mammary Gland Biol Neoplasia. 2017;22(2):93-108.
79. Russo J, Russo IH. Toward a physiological approach to breast cancer prevention. Cancer Epidemiol Biomarkers Prev. 1994;3(4):353-64.
80. Bazira PJ, Ellis H, Mahadevan V. Anatomy and physiology of the breast. Surgery (Oxford). 2022;40(2):79-83.
81. Kumar T, Nee K, Wei R, He S, Nguyen QH, Bai S, et al. A spatially resolved single-cell genomic atlas of the adult human breast. Nature. 2023;620(7972):181-91.

82. Gray GK, Li CM-C, Rosenbluth JM, Selfors LM, Girnius N, Lin J-R, et al. A human breast atlas integrating single-cell proteomics and transcriptomics. Dev Cell. 2022;57(11):1400-20.e7.
83. Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell. 2010;7(3):403-17.
84. Yue W, Yager JD, Wang J-P, Jupe ER, Santen RJ. Estrogen receptor-dependent and independent mechanisms of breast cancer carcinogenesis. Steroids. 2013;78(2):161-70.
85. Sternlicht MD. Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. Breast Cancer Res. 2005;8(1):1-11.
86. Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. N Engl J Med. 2006;354(3):270-82.
87. Hall JM, Couse JF, Korach KS. The multifaceted mechanisms of estradiol and estrogen receptor signaling. J Biol Chem. 2001;276(40):36869-72.
88. Deroo BJ, Korach KS. Estrogen receptors and human disease. J Clin Invest. 2006;116(3):561-70.
89. Hua H, Zhang H, Kong Q, Jiang Y. Mechanisms for estrogen receptor expression in human cancer. Exp Hematol Oncol. 2018;7(1):24.
90. Björnström L, Sjöberg M. Mechanisms of estrogen receptor signaling: Convergence of genomic and nongenomic actions on target genes. Mol Endocrinol. 2005;19(4):833-42.
91. Conneely OM, Jericevic BM, Lydon JP. Progesterone receptors in mammary gland development and tumorigenesis. J Mammary Gland Biol Neoplasia. 2003;8(2):205-14.
92. Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli GA, et al. Progesterone receptor modulates ERα action in breast cancer. Nature. 2015;523(7560):313-7.
93. Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene. 2007;26(45):6469-87.
94. Iqbal N, Iqbal N. Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and therapeutic implications. Mol Biol Int. 2014;2014:852748.
95. 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-M r tumour antigen. Nature. 1984;312(5994):513-6.
96. 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(4785):177-82.
97. Bhardwaj P, Au CC, Benito-Martin A, Ladumor H, Oshchepkova S, Moges R, et al. Estrogens and breast cancer: Mechanisms involved in obesity-related development, growth and progression. J Steroid Biochem Mol Biol. 2019;189:161-70.

98. Felty Q, Xiong W-C, Sun D, Sarkar S, Singh KP, Parkash J, et al. Estrogen-induced mitochondrial reactive oxygen species as signal-transducing messengers. Biochemistry. 2005;44(18):6900-9.
99. Cavalieri E, Stack D, Devanesan P, Todorovic R, Dwivedy I, Higginbotham S, et al. Molecular origin of cancer: catechol estrogen-3, 4-quinones as endogenous tumor initiators. Proc Natl Acad Sci U S A. 1997;94(20):10937-42.
100. Fernandez SV, Russo IH, Russo J. Estradiol and its metabolites 4-hydroxyestradiol and 2-hydroxyestradiol induce mutations in human breast epithelial cells. Int J Cancer. 2006;118(8):1862-8.
101. Thomas C, Gustafsson J-Å. The different roles of ER subtypes in cancer biology and therapy. Nat Rev Cancer. 2011;11(8):597-608.
102. Lu F, Zahid M, Saeed M, Cavalieri EL, Rogan EG. Estrogen metabolism and formation of estrogen-DNA adducts in estradiol-treated MCF-10F cells: The effects of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin induction and catechol-O-methyltransferase inhibition. J Steroid Biochem Mol Biol. 2007;105(1-5):150-8.
103. Savage KI, Matchett KB, Barros EM, Cooper KM, Irwin GW, Gorski JJ, et al. BRCA1 deficiency exacerbates estrogen-induced DNA damage and genomic instability. Cancer Res. 2014;74(10):2773-84.
104. Sachdev D, Yee D. Disrupting insulin-like growth factor signaling as a potential cancer therapy. Mol Cancer Ther. 2007;6(1):1-12.
105. Pollak M. The insulin and insulin-like growth factor receptor family in neoplasia: an update. Nat Rev Cancer. 2012;12(3):159-69.
106. Christopoulos PF, Msaouel P, Koutsilieris M. The role of the insulin-like growth factor-1 system in breast cancer. Mol Cancer. 2015;14:43.
107. Ruan W, Kleinberg DL. Insulin-like growth factor I is essential for terminal end bud formation and ductal morphogenesis during mammary development. Endocrinology. 1999;140(11):5075-81.
108. Macias H, Hinck L. Mammary gland development. Wiley Interdiscip Rev Dev Biol. 2012;1(4):533-57.
109. Key T, Appleby P, Reeves G, Roddam A. Endogenous Hormones and Breast Cancer Collaborative Group. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. Lancet Oncol. 2010;11(6):530-42.
110. Kaaks R, Johnson T, Tikk K, Sookthai D, Tjønneland A, Roswall N, et al. Insulin-like growth factor I and risk of breast cancer by age and hormone receptor status—A prospective study within the EPIC cohort. Int J Cancer. 2014;134(11):2683-90.
111. Murphy N, Knuppel A, Papadimitriou N, Martin RM, Tsilidis KK, Smith-Byrne K, et al. Insulin-like growth factor-1, insulin-like growth factor-binding protein-3, and breast cancer risk: observational and Mendelian randomization analyses with ∼430 000 women. Ann Oncol. 2020;31(5):641-9.
112. Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B, et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. Lancet. 1998;351(9113):1393-6.

113. LeRoith D, Roberts CT, Jr. The insulin-like growth factor system and cancer. Cancer Lett. 2003;195(2):127-37.
114. Oh Y, Nagalla SR, Yamanaka Y, Kim H-S, Wilson E, Rosenfeld RG. Synthesis and characterization of insulin-like growth factor-binding protein (IGFBP)-7: recombinant human mac25 protein specifically binds IGF-I and-II. J Biol Chem. 1996;271(48):30322-5.
115. Pollak M. Insulin and insulin-like growth factor signalling in neoplasia. Nat Rev Cancer. 2008;8(12):915-28.
116. Hwa V, Oh Y, Rosenfeld RG. The insulin-like growth factor-binding protein (IGFBP) superfamily. Endocr Rev. 1999;20(6):761-87.
117. Karey KP, Sirbasku DA. Differential responsiveness of human breast cancer cell lines MCF-7 and T47D to growth factors and 17β-estradiol. Cancer Res. 1988;48(14):4083-92.
118. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. BMJ. 2000;321(7261):624-8.
119. Thomson CA, McCullough ML, Wertheim BC, Chlebowski RT, Martinez ME, Stefanick ML, et al. Nutrition and physical activity cancer prevention guidelines, cancer risk, and mortality in the women's health initiative. Cancer Prev Res (Phila). 2014;7(1):42-53.
120. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356(3):227-36.
121. Bodewes FTH, van Asselt AA, Dorrius MD, Greuter MJW, de Bock GH. Mammographic breast density and the risk of breast cancer: A systematic review and meta-analysis. Breast. 2022;66:62-8.
122. Santen RJ, Mansel R. Benign Breast Disorders. N Engl J Med. 2005;353(3):275-85.
123. Tamimi RM, Rosner B, Colditz GA. Evaluation of a breast cancer risk prediction model expanded to include category of prior benign breast disease lesion. Cancer. 2010;116(21):4944-53.
124. Frank C, Sundquist J, Yu H, Hemminki A, Hemminki K. Concordant and discordant familial cancer: Familial risks, proportions and population impact. Int J Cancer. 2017;140(7):1510-6.
125. Ahern TP, Sprague BL, Bissell MCS, Miglioretti DL, Buist DSM, Braithwaite D, et al. Family history of breast cancer, breast density, and breast cancer risk in a U.S. breast cancer screening population. Cancer Epidemiol Biomarkers Prev. 2017;26(6):938-44.
126. 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(9291):1389-99.
127. Foulkes WD. Inherited susceptibility to common cancers. N Engl J Med. 2008;359(20):2143-53.
128. Hemminki K, Granström C, Czene K. Attributable risks for familial breast cancer by proband status and morphology: a nationwide epidemiologic study from Sweden. Int J Cancer. 2002;100(2):214-9.

129. Lalloo F, Evans DG. Familial breast cancer. Clin Genet. 2012;82(2):105-14.
130. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990;250(4988):1684-9.
131. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 1994;266(5182):66-71.
132. Futreal PA, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S, et al. BRCA1 mutations in primary breast and ovarian carcinomas. Science. 1994;266(5182):120-2.
133. Lord CJ, Ashworth A. The DNA damage response and cancer therapy. Nature. 2012;481(7381):287-94.
134. Venkitaraman AR. Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. Science. 2014;343(6178):1470-5.
135. Breast Cancer Risk Genes — Association Analysis in More than 113,000 Women. N Engl J Med. 2021;384(5):428-39.
136. 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(23):2402-16.
137. Mavaddat N, Peock S, Frost D, Ellis S, Platte R, Fineberg E, et al. Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. J Natl Cancer Inst. 2013;105(11):812-22.
138. King MC, Marks JH, Mandell JB. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 2003;302(5645):643-6.
139. Mersch J, Jackson MA, Park M, Nebgen D, Peterson SK, Singletary C, et al. Cancers associated with BRCA1 and BRCA2 mutations other than breast and ovarian. Cancer. 2015;121(2):269-75.
140. Nyberg T, Frost D, Barrowdale D, Evans DG, Bancroft E, Adlard J, et al. Prostate cancer risks for male BRCA1 and BRCA2 mutation carriers: A prospective cohort study. Eur Urol. 2020;77(1):24-35.
141. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet. 2020;52(6):572-81.
142. Ahearn TU, Zhang H, Michailidou K, Milne RL, Bolla MK, Dennis J, et al. Common variants in breast cancer risk loci predispose to distinct tumor subtypes. Breast Cancer Res. 2022;24(1):2.
143. Milne RL, Kuchenbaecker KB, Michailidou K, Beesley J, Kar S, Lindström S, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet. 2017;49(12):1767-78.
144. Guzmán-Arocho YD, Rosenberg SM, Garber JE, Vardeh H, Poorvu PD, Ruddy KJ, et al. Clinicopathological features and BRCA1 and BRCA2 mutation status in a prospective cohort of young women with breast cancer. Br J Cancer. 2022;126(2):302-9.

145. 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. Ann Oncol. 2023;34(11):970-86.
146. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al. A population-based study of genes previously implicated in breast cancer. N Engl J Med. 2021;384(5):440-51.
147. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92-4.
148. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013;45(4):353-61, 61e1-2.
149. Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, et al. Genome-wide association studies in ancestrally diverse populations: Opportunities, methods, pitfalls, and recommendations. Cell. 2019;179(3):589-603.
150. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet. 2019;104(1):21-34.
151. Kramer I, Hooning MJ, Mavaddat N, Hauptmann M, Keeman R, Steyerberg EW, et al. Breast cancer polygenic risk score and contralateral breast cancer risk. Am J Hum Genet. 2020;107(5):837-48.
152. Yiangou K, Mavaddat N, Dennis J, Zanti M, Wang Q, Bolla MK, et al. Differences in polygenic score distributions in European ancestry populations: implications for breast cancer risk prediction. medRxiv. 2024.
153. Ho WK, Tan MM, Mavaddat N, Tai MC, Mariapun S, Li J, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. Nat Commun. 2020;11(1):3833.
154. Du Z, Gao G, Adedokun B, Ahearn T, Lunetta KL, Zirpoli G, et al. Evaluating polygenic risk scores for breast cancer in women of african ancestry. J Natl Cancer Inst. 2021;113(9):1168-76.
155. Lopes Cardozo JMN, Andrulis IL, Bojesen SE, Dörk T, Eccles DM, Fasching PA, et al. Associations of a breast cancer polygenic risk score with tumor characteristics and survival. J Clin Oncol. 2023;41(10):1849-63.
156. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Primers. 2021;1(1):59.
157. Prajzendanc K, Domagała P, Hybiak J, Ryś J, Huzarski T, Szwiec M, et al. BRCA1 promoter methylation in peripheral blood is associated with the risk of triple-negative breast cancer. Int J Cancer. 2020;146(5):1293-8.
158. Lønning PE, Berge EO, Bjørnslett M, Minsaas L, Chrisanthar R, Høberg-Vetti H, et al. White blood cell BRCA1 promoter methylation status and ovarian cancer risk. Ann Intern Med. 2018;168(5):326-34.

159. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol. 2012;13(11):1141-51.
160. Lambertini M, Santoro L, Del Mastro L, Nguyen B, Livraghi L, Ugolini D, et al. Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. Cancer Treat Rev. 2016;49:65-76.
161. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. Lancet. 2002;360(9328):187-95.
162. Manjer J, Johansson R, Berglund G, Janzon L, Kaaks R, Agren A, et al. Postmenopausal breast cancer risk in relation to sex steroid hormones, prolactin and SHBG (Sweden). Cancer Causes Control. 2003;14(7):599-607.
163. Hormones TE, Group BCC. Endogenous Sex Hormones and Breast Cancer in Postmenopausal Women: Reanalysis of Nine Prospective Studies. J Natl Cancer Inst. 2002;94(8):606-16.
164. Key TJ, Appleby PN, Reeves GK, Travis RC, Alberg AJ, Barricarte A, 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(10):1009-19.
165. Collaborative Group on Hormonal Factors in Breast Cancer. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. Lancet. 2019;394(10204):1159-68.
166. Morch LS, Skovlund CW, Hannaford PC, Iversen L, Fielding S, Lidegaard O. Contemporary hormonal contraception and the risk of breast cancer. N Engl J Med. 2017;377(23):2228-39.
167. Hunter DJ, Colditz GA, Hankinson SE, Malspeis S, Spiegelman D, Chen W, et al. Oral contraceptive use and breast cancer: a prospective study of young women. Cancer Epidemiol Biomarkers Prev. 2010;19(10):2496-502.
168. Breast cancer and hormonal contraceptives: Collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Lancet. 1996;347(9017):1713-27.
169. Jernström H, Loman N, Johannsson OT, Borg A, Olsson H. Impact of teenage oral contraceptive use in a population-based series of early-onset breast cancer cases who have undergone BRCA mutation testing. Eur J Cancer. 2005;41(15):2312-20.
170. World Health Organization (WHO). Waist circumference and waist-hip ratio: report of a WHO expert consultation, Geneva, 8-11 December 2008 2011 [Available from: https://www.who.int/publications/i/item/9789241501491

171. Chen Y, Liu L, Zhou Q, Imam MU, Cai J, Wang Y, et al. Body mass index had different effects on premenopausal and postmenopausal breast cancer risks: A dose-response meta-analysis with 3,318,796 subjects from 31 cohort studies. BMC Public Health. 2017;17(1):1-11.
172. Neuhouser ML, Aragaki AK, Prentice RL, Manson JE, Chlebowski R, Carty CL, et al. Overweight, obesity, and postmenopausal invasive breast cancer risk: A secondary analysis of the Women's Health Initiative randomized clinical trials. JAMA Oncol. 2015;1(5):611-21.
173. Borgquist S, Jirström K, Anagnostaki L, Manjer J, Landberg G. Anthropometric factors in relation to different tumor biological subgroups of postmenopausal breast cancer. Int J Cancer. 2009;124(2):402-11.
174. Chen L, Cook LS, Tang M-TC, Porter PL, Hill DA, Wiggins CL, et al. Body mass index and risk of luminal, HER2-overexpressing, and triple negative breast cancer. Breast Cancer Res Treat. 2016;157(3):545-54.
175. Ritte R, Lukanova A, Berrino F, Dossus L, Tjønneland A, Olsen A, et al. Adiposity, hormone replacement therapy use and breast cancer risk by age and hormone receptor status: A large prospective cohort study. Breast Cancer Res. 2012;14(3):R76.
176. van den Brandt PA, Ziegler RG, Wang M, Hou T, Li R, Adami H-O, et al. Body size and weight change over adulthood and risk of breast cancer by menopausal and hormone receptor status: A pooled analysis of 20 prospective cohort studies. Eur J Epidemiol. 2021;36(1):37-55.
177. Ewertz M, Jensen M-B, Gunnarsdóttir KÁ, Højris I, Jakobsen EH, Nielsen D, et al. Effect of obesity on prognosis after early-stage breast cancer. J Clin Oncol. 2011;29(1):25-31.
178. Petrelli F, Cortellini A, Indini A, Tomasello G, Ghidini M, Nigro O, et al. Association of obesity with survival outcomes in patients with cancer: a systematic review and meta-analysis. JAMA Netw Open. 2021;4(3):e213520-e.
179. Esposito K, Chiodini P, Colao A, Lenzi A, Giugliano D. Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. Diabetes Care. 2012;35(11):2402-11.
180. Pearson-Stuttard J, Papadimitriou N, Markozannes G, Cividini S, Kakourou A, Gill D, et al. Type 2 Diabetes and Cancer: An Umbrella Review of Observational and Mendelian Randomization Studies. Cancer Epidemiol Biomarkers Prev. 2021;30(6):1218-28.
181. Iyengar NM, Gucalp A, Dannenberg AJ, Hudis CA. Obesity and Cancer Mechanisms: Tumor Microenvironment and Inflammation. J Clin Oncol. 2016;34(35):4270-6.
182. Hopkins BD, Goncalves MD, Cantley LC. Obesity and Cancer Mechanisms: Cancer Metabolism. J Clin Oncol. 2016;34(35):4277-83.
183. Xu M, Cao FL, Li N, Gao X, Su X, Jiang X. Leptin induces epithelial-to-mesenchymal transition via activation of the ERK signaling pathway in lung cancer cells. Oncol Lett. 2018;16(4):4782-8.

184. Juárez-Cruz JC, Okoniewski M, Ramírez M, Ortuño-Pineda C, Navarro-Tito N, Castañeda-Saucedo E. Chronic leptin treatment induces epithelial-mesenchymal transition in MCF10A mammary epithelial cells. J Mammary Gland Biol Neoplasia. 2022;27(1):19-36.
185. Lee-Rueckert M, Canyelles M, Tondo M, Rotllan N, Kovanen PT, Llorente-Cortes V, et al. Obesity-induced changes in cancer cells and their microenvironment: Mechanisms and therapeutic perspectives to manage dysregulated lipid metabolism. Semin Cancer Biol. 2023;93:36-51.
186. Jung S, Wang M, Anderson K, Baglietto L, Bergkvist L, Bernstein L, et al. Alcohol consumption and breast cancer risk by estrogen receptor status: In a pooled analysis of 20 studies. Int J Epidemiol. 2015;45(3):916-28.
187. Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. JAMA. 2001;286(17):2143-51.
188. Macacu A, Autier P, Boniol M, Boyle P. Active and passive smoking and risk of breast cancer: a meta-analysis. Breast Cancer Res Treat. 2015;154(2):213-24.
189. Lundqvist A, Andersson E, Ahlberg I, Nilbert M, Gerdtham U. Socioeconomic inequalities in breast cancer incidence and mortality in Europe-a systematic review and meta-analysis. Eur J Public Health. 2016;26(5):804-13.
190. Menvielle G, Kunst AE, van Gils CH, Peeters PH, Boshuizen H, Overvad K, et al. The contribution of risk factors to the higher incidence of invasive and in situ breast cancers in women with higher levels of education in the European prospective investigation into cancer and nutrition. Am J Epidemiol. 2011;173(1):26-37.
191. Svenska Bröstcancergruppen/ Swedish Breast Cancer Group. Nationellt vårdprogram Bröstcancer 2024 [Internet]. Regionala Cancercentrum i Samverkan/Confederation of Regional Cancer Centres in Sweden; 2024 [updated 2022-02-07; cited 2024 2024-02-25]. Available from: https://kunskapsbanken.cancercentrum.se/diagnoser/brostcancer/vardprogram/.
192. Sinn HP, Kreipe H. A brief overview of the WHO classification of breast tumors, 4th Edition, focusing on issues and updates from the 3rd Edition. Breast Care (Basel). 2013;8(2):149-54.
193. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. Histopathology. 1991;19(5):403-10.
194. Hortobagyi GN, Edge SB, Giuliano A. New and important changes in the TNM staging system for breast cancer. Am Soc Clin Oncol Educ Book. 2018;38:457-67.
195. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin. 2017;67(2):93-9.
196. Loibl S, André F, Bachelot T, Barrios CH, Bergh J, Burstein HJ, et al. Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2024;35(2):159-82.

197. Fehrenbacher L, Cecchini RS, Geyer Jr CE, Rastogi P, Costantino JP, Atkins JN, et al. NSABP B-47/NRG oncology phase III randomized trial comparing adjuvant chemotherapy with or without trastuzumab in high-risk invasive breast cancer negative for HER2 by FISH and with IHC 1+ or 2+. J Clin Oncol. 2020;38(5):444.
198. Zagami P, Carey LA. Triple negative breast cancer: Pitfalls and progress. NPJ Breast Cancer. 2022;8(1):95.
199. Foulkes WD, Smith IE, Reis-Filho JS. Triple-Negative Breast Cancer. N Engl J Med. 2010;363(20):1938-48.
200. Bianchini G, De Angelis C, Licata L, Gianni L. Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. Nat Rev Clin Oncol. 2022;19(2):91-113.
201. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol. 2016;13(11):674-90.
202. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol. 2010;11(2):174-83.
203. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: Recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst. 2011;103(22):1656-64.
204. Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, et al. Assessment of Ki67 in breast cancer: Updated recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst. 2021;113(7):808-19.
205. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006;7:96.
206. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003;100(14):8418-23.
207. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.
208. Fredlund E, Staaf J, Rantala JK, Kallioniemi O, Borg A, Ringnér M. The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. Breast Cancer Res. 2012;14(4):R113.
209. 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(1):83.
210. Dias K, Dvorkin-Gheva A, Hallett RM, Wu Y, Hassell J, Pond GR, et al. Claudin-low breast cancer; clinical & pathological characteristics. PLoS One. 2017;12(1):e0168669.
211. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):1-18.

212. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346-52.
213. Fougner C, Bergholtz H, Kuiper R, Norum JH, Sørlie T. Claudin-low-like mouse mammary tumors show distinct transcriptomic patterns uncoupled from genomic drivers. Breast Cancer Res. 2019;21(1):1-13.
214. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27(8):1160-7.
215. Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. Ann Oncol. 2014;25(2):339-45.
216. Filipits M, Nielsen TO, Rudas M, Greil R, Stöger H, Jakesz R, et al. The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. Clin Cancer Res. 2014;20(5):1298-305.
217. 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.
218. Ohnstad HO, Borgen E, Falk RS, Lien TG, Aaserud M, Sveli MAT, et al. Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up. Breast Cancer Res. 2017;19(1):120.
219. Lænkholm AV, Jensen MB, Eriksen JO, Rasmussen BB, Knoop AS, Buckingham W, et al. PAM50 Risk of Recurrence Score Predicts 10-Year Distant Recurrence in a Comprehensive Danish Cohort of Postmenopausal Women Allocated to 5 Years of Endocrine Therapy for Hormone Receptor-Positive Early Breast Cancer. J Clin Oncol. 2018;36(8):735-40.
220. Sestak I, Cuzick J, Dowsett M, Lopez-Knowles E, Filipits M, Dubsky P, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. J Clin Oncol. 2015.
221. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, et al. Comparison of PAM50 risk of recurrence score with onco type DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. J Clin Oncol. 2013;31(22):2783-90.
222. Gnant M, Sestak I, Filipits M, Dowsett M, Balic M, Lopez-Knowles E, et al. Identifying clinically relevant prognostic subgroups of postmenopausal women with node-positive hormone receptor-positive early-stage breast cancer treated with endocrine therapy: A combined analysis of ABCSG-8 and ATAC using the PAM50 risk of recurrence score and intrinsic subtype. Ann Oncol. 2015;26(8):1685-91.

223. Lænkholm AV, Jensen MB, Eriksen JO, Roslind A, Buckingham W, Ferree S, et al. Population-based study of Prosigna-PAM50 and outcome among postmenopausal women with estrogen receptor-positive and HER2-negative operable invasive lobular or ductal breast cancer. Clin Breast Cancer. 2020;20(4):e423-e32.
224. Bueno-de-Mesquita J, Linn S, Keijzer R, Wesseling J, Nuyten D, Van Krimpen C, et al. Validation of 70-gene prognosis signature in node-negative breast cancer. Breast Cancer Res Treat. 2009;117:483-95.
225. Buyse M, Loi S, Van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst. 2006;98(17):1183-92.
226. Van De Vijver MJ, He YD, Van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002;347(25):1999-2009.
227. Mook S, Schmidt MK, Viale G, Pruneri G, Eekhout I, Floore A, et al. The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1–3 positive lymph nodes in an independent validation study. Breast Cancer Res Treat. 2009;116:295-302.
228. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor–positive breast cancer. J Clin Oncol. 2006;24(23):3726-34.
229. Sparano JA, Gray RJ, Ravdin PM, Makower DF, Pritchard KI, Albain KS, et al. Clinical and genomic risk to guide the use of adjuvant therapy for breast cancer. N Engl J Med. 2019;380(25):2395-405.
230. Buus R, Szijgyarto Z, Schuster EF, Xiao H, Haynes BP, Sestak I, et al. Development and validation for research assessment of Oncotype DX® Breast Recurrence Score, EndoPredict® and Prosigna®. NPJ Breast Cancer. 2021;7(1):15.
231. Guedj M, Marisa L, de Reynies A, Orsetti B, Schiappa R, Bibeau F, et al. A refined molecular taxonomy of breast cancer. Oncogene. 2012;31(9):1196-206.
232. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med. 2008;14(5):518-27.
233. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, et al. Meta-analysis of gene expression profiles in breast cancer: Toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res. 2008;10(4):R65.
234. Rueda OM, Sammut SJ, Seoane JA, Chin SF, Caswell-Jin JL, Callari M, et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. Nature. 2019;567(7748):399-404.
235. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun. 2016;7:11479.
236. Ali HR, Rueda OM, Chin SF, Curtis C, Dunning MJ, Aparicio SA, et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biol. 2014;15(8):431.

237. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530-6.
238. 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(27):2817-26.
239. Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, et al. Use of Biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol. 2016;34(10):1134-50.
240. Allott EH, Cohen SM, Geradts J, Sun X, Khoury T, Bshara W, et al. Performance of three-biomarker immunohistochemistry for intrinsic breast Cancer subtyping in the AMBER Consortium. Cancer Epidemiol Biomarkers Prev. 2016;25(3):470-8.
241. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clin Cancer Res. 2010;16(21):5222-32.
242. Staaf J, Ringnér M, Vallon-Christersson J, Jönsson G, Bendahl PO, Holm K, et al. Identification of subtypes in human epidermal growth factor receptor 2--positive breast cancer reveals a gene signature prognostic of outcome. J Clin Oncol. 2010;28(11):1813-20.
243. Jin X, Zhou YF, Ma D, Zhao S, Lin CJ, Xiao Y, et al. Molecular classification of hormone receptor-positive HER2-negative breast cancer. Nat Genet. 2023;55(10):1696-708.
244. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121(7):2750-67.
245. Lehmann BD, Jovanović B, Chen X, Estrada MV, Johnson KN, Shyr Y, et al. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLoS One. 2016;11(6):e0157368.
246. Jiang YZ, Ma D, Suo C, Shi J, Xue M, Hu X, et al. Genomic and Transcriptomic Landscape of triple-negative breast cancers: Subtypes and treatment strategies. Cancer Cell. 2019;35(3):428-40.e5.
247. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res. 2015;21(7):1688-98.
248. Ring BZ, Hout DR, Morris SW, Lawrence K, Schweitzer BL, Bailey DB, et al. Generation of an algorithm based on minimal gene sets to clinically subtype triple negative breast cancer patients. BMC Cancer. 2016;16:143.
249. Alkner S, Tang MH, Brueffer C, Dahlgren M, Chen Y, Olsson E, et al. Contralateral breast cancer can represent a metastatic spread of the first primary tumor: determination of clonal relationship between contralateral breast cancers using next-generation whole genome sequencing. Breast Cancer Res. 2015;17(1):102.

250. Begg CB, Ostrovnaya I, Geyer FC, Papanastasiou AD, Ng CKY, Sakr RA, et al. Contralateral breast cancers: Independent cancers or metastases? Int J Cancer. 2018;142(2):347-56.
251. Akdeniz D, Schmidt MK, Seynaeve CM, McCool D, Giardiello D, van den Broek AJ, et al. Risk factors for metachronous contralateral breast cancer: A systematic review and meta-analysis. Breast. 2019;44:1-14.
252. Hartman M, Czene K, Reilly M, Adolfsson J, Bergh J, Adami HO, et al. Incidence and prognosis of synchronous and metachronous bilateral breast cancer. J Clin Oncol. 2007;25(27):4210-6.
253. Heron DE, Komarnicky LT, Hyslop T, Schwartz GF, Mansfield CM. Bilateral breast carcinoma: risk factors and outcomes for patients with synchronous and metachronous disease. Cancer. 2000;88(12):2739-50.
254. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, Zhang Y, et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. J Clin Oncol. 2008;26(20):3324-30.
255. Brandt J, Garne JP, Tengrup I, Manjer J. Age at diagnosis in relation to survival following breast cancer: A cohort study. World J Surg Oncol. 2015;13(1):33.
256. Liao S, Hartmaier RJ, McGuire KP, Puhalla SL, Luthra S, Chandran UR, et al. The molecular landscape of premenopausal breast cancer. Breast Cancer Res. 2015;17(1):1-13.
257. Yau C, Fedele V, Roydasgupta R, Fridlyand J, Hubbard A, Gray JW, et al. Aging impacts transcriptomes but not genomes of hormone-dependent breast cancers. Breast Cancer Res. 2007;9(5):1-16.
258. Ma D, Jiang YZ, Xiao Y, Xie MD, Zhao S, Jin X, et al. Integrated molecular profiling of young and elderly patients with triple-negative breast cancer indicates different biological bases and clinical management strategies. Cancer. 2020;126(14):3209-18.
259. Schonberg MA, Marcantonio ER, Li D, Silliman RA, Ngo L, McCarthy EP. Breast cancer among the oldest old: Tumor characteristics, treatment choices, and survival. J Clin Oncol. 2010;28(12):2038.
260. Pirmohamed M. Pharmacogenomics: current status and future perspectives. Nat Rev Genet. 2023;24(6):350-62.
261. Henricks LM, Lunenburg C, de Man FM, Meulendijks D, Frederix GWJ, Kienhuis E, et al. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: A prospective safety analysis. Lancet Oncol. 2018;19(11):1459-67.
262. Henricks LM, Lunenburg C, de Man FM, Meulendijks D, Frederix GWJ, Kienhuis E, et al. A cost analysis of upfront DPYD genotype-guided dose individualisation in fluoropyrimidine-based anticancer therapy. Eur J Cancer. 2019;107:60-7.
263. Fontanella C, Lederer B, Gade S, Vanoppen M, Blohmer JU, Costa SD, et al. Impact of body mass index on neoadjuvant treatment outcome: A pooled analysis of eight prospective neoadjuvant breast cancer trials. Breast Cancer Res Treat. 2015;150:127-39.

264. Wang H, Yee D, Potter D, Jewett P, Yau C, Beckwith H, et al. Impact of body mass index on pathological response after neoadjuvant chemotherapy: results from the I-SPY 2 trial. Breast Cancer Res Treat. 2024.
265. Candido dos Reis FJ, Wishart GC, Dicks EM, Greenberg D, Rashbass J, Schmidt MK, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. Breast Cancer Res. 2017;19:1-13.
266. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res. 2010;12(1):1-10.
267. Grootes I, Wishart GC, Pharoah PDP. An updated PREDICT breast cancer prognostic model including the benefits and harms of radiotherapy. NPJ Breast Cancer. 2024;10(1):6.
268. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol. 2013;24(9):2206-23.
269. Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet. 2011;378(9793):771-84.
270. Purdie C, Quinlan P, Jordan L, Ashfield A, Ogston S, Dewar J, et al. Progesterone receptor expression is an independent prognostic variable in early breast cancer: A population-based study. Br J Cancer. 2014;110(3):565-72.
271. Grootes I, Keeman R, Blows FM, Milne RL, Giles GG, Swerdlow AJ, et al. Incorporating progesterone receptor expression into the PREDICT breast prognostic model. Eur J Cancer. 2022;173:178-93.
272. 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(14):1273-83.
273. 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(9938):164-72.
274. Smith I, Robertson J, Kilburn L, Wilcox M, Evans A, Holcombe C, et al. Long-term outcome and prognostic value of Ki67 after perioperative endocrine therapy in postmenopausal women with hormone-sensitive early breast cancer (POETIC): an open-label, multicentre, parallel-group, randomised, phase 3 trial. Lancet Oncol. 2020;21(11):1443-54.
275. Goncalves R, Ma C, Luo J, Suman V, Ellis MJ. Use of neoadjuvant data to design adjuvant endocrine therapy trials for breast cancer. Nat Rev Clin Oncol. 2012;9(4):223-9.
276. Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, et al. 21-gene assay to inform chemotherapy benefit in node-positive breast cancer. N Engl J Med. 2021;385(25):2336-47.

277. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med. 2018;379(2):111-21.
278. Piccart M, van 't Veer LJ, Poncet C, Lopes Cardozo JMN, Delaloge S, Pierga JY, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: Updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. Lancet Oncol. 2021;22(4):476-88.
279. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. N Engl J Med. 2016;375(8):717-29.
280. Halsted WS. I. The results of operations for the cure of cancer of the breast performed at the Johns Hopkins Hospital from June, 1889, to January, 1894. Ann Surg. 1894;20(5):497.
281. Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med. 2002;347(16):1227-32.
282. Litière S, Werutsky G, Fentiman IS, Rutgers E, Christiaens MR, Van Limbergen E, et al. Breast conserving therapy versus mastectomy for stage I-II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. Lancet Oncol. 2012;13(4):412-9.
283. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med. 2002;347(16):1233-41.
284. de Boniface J, Szulkin R, Johansson ALV. Survival After breast conservation vs mastectomy adjusted for comorbidity and socioeconomic status: A swedish national 6-Year follow-up of 48 986 women. JAMA Surg. 2021;156(7):628-37.
285. van Maaren MC, de Munck L, de Bock GH, Jobsen JJ, van Dalen T, Linn SC, et al. 10 year survival after breast-conserving surgery plus radiotherapy compared with mastectomy in early breast cancer in the Netherlands: a population-based study. Lancet Oncol. 2016;17(8):1158-70.
286. Agarwal S, Pappas L, Neumayer L, Kokeny K, Agarwal J. Effect of breast conservation therapy vs mastectomy on disease-specific survival for early-stage breast cancer. JAMA Surg. 2014;149(3):267-74.
287. Houssami N, Macaskill P, Marinovich ML, Morrow M. The association of surgical margins and local recurrence in women with early-stage invasive breast cancer treated with breast-conserving therapy: A meta-analysis. Ann Surg Oncol. 2014;21(3):717-30.
288. Macmillan RD, McCulley SJ. Oncoplastic Breast Surgery: What, when and for whom? Curr Breast Cancer Rep. 2016;8:112-7.
289. Giuliano AE, Ballman KV, McCall L, Beitsch PD, Brennan MB, Kelemen PR, et al. Effect of axillary dissection vs no axillary dissection on 10-year overall survival among women with invasive breast cancer and sentinel node metastasis: The ACOSOG Z0011 (Alliance) randomized clinical trial. JAMA. 2017;318(10):918-26.

290. Giuliano AE, Hunt KK, Ballman KV, Beitsch PD, Whitworth PW, Blumencranz PW, et al. Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: A randomized clinical trial. JAMA. 2011;305(6):569-75.
291. Giuliano AE, McCall L, Beitsch P, Whitworth PW, Blumencranz P, Leitch AM, et al. Locoregional recurrence after sentinel lymph node dissection with or without axillary dissection in patients with sentinel lymph node metastases: The American College of Surgeons Oncology Group Z0011 randomized trial. Ann Surg. 2010;252(3):426.
292. Bartels SAL, Donker M, Poncet C, Sauvé N, Straver ME, van de Velde CJH, et al. Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer: 10-Year results of the randomized controlled EORTC 10981-22023 AMAROS Trial. J Clin Oncol. 2023;41(12):2159-65.
293. Boniface Jd, Tvedskov TF, Rydén L, Szulkin R, Reimer T, Kühn T, et al. Omitting Axillary dissection in breast cancer with sentinel-node metastases. N Engl J Med. 2024;390(13):1163-75.
294. Gentilini OD, Botteri E, Sangalli C, Galimberti V, Porpiglia M, Agresti R, et al. Sentinel lymph node biopsy vs no axillary surgery in patients with small breast cancer and negative results on ultrasonography of axillary lymph nodes: The SOUND randomized clinical trial. JAMA Oncol. 2023;9(11):1557-64.
295. Meattini I, Becherini C, Caini S, Coles CE, Cortes J, Curigliano G, et al. International multidisciplinary consensus on the integration of radiotherapy with new systemic treatments for breast cancer: European Society for Radiotherapy and Oncology (ESTRO)-endorsed recommendations. Lancet Oncol. 2024;25(2):e73-e83.
296. Darby S, McGale P, Correa C, Taylor C, Arriagada R, Clarke M, et al. 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. Lancet. 2011;378(9804):1707-16.
297. Alkner S, de Boniface J, Lundstedt D, Mjaaland I, Ryden L, Vikstrom J, et al. Protocol for the T-REX-trial: Tailored regional external beam radiotherapy in clinically node-negative breast cancer patients with 1-2 sentinel node macrometastases - an open, multicentre, randomised non-inferiority phase 3 trial. BMJ Open. 2023;13(9):e075543.
298. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2019;30(8):1194-220.
299. Brunt AM, Haviland JS, Kirby AM, Somaiah N, Wheatley DA, Bliss JM, et al. Five-fraction radiotherapy for breast cancer: FAST-Forward to implementation. Clin Oncol (R Coll Radiol). 2021;33(7):430-9.
300. Murray Brunt A, Haviland JS, Wheatley DA, Sydenham MA, Alhasso A, Bloomfield DJ, et al. Hypofractionated breast radiotherapy for 1 week versus 3 weeks (FAST-Forward): 5-year efficacy and late normal tissue effects results from a multicentre, non-inferiority, randomised, phase 3 trial. Lancet. 2020;395(10237):1613-26.

301. Beatson GT. On the treatment of inoperable cases of carcinoma of the mamma: Suggestions for a new method of treatment, with illustrative cases. Trans Med Chir Soc Edinb. 1896;15:153-79.
302. Harper M, Walpole A. A new derivative of triphenylethylene: effect on implantation and mode of action in rats. Reproduction. 1967;13(1):101-19.
303. Jordan VC. Fourteenth Gaddum Memorial Lecture. A current view of tamoxifen for the treatment and prevention of breast cancer. Br J Pharmacol. 1993;110(2):507.
304. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Early Breast Cancer Trialists' Collaborative Group. Lancet. 1992;339(8784):1-15.
305. Aromatase inhibitors versus tamoxifen in early breast cancer: Patient-level meta-analysis of the randomised trials. Lancet. 2015;386(10001):1341-52.
306. Howell A, Cuzick J, Baum M, Buzdar A, Dowsett M, Forbes JF, et al. Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. Lancet. 2005;365(9453):60-2.
307. Group BIG-C. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. N Engl J Med. 2005;353(26):2747-57.
308. van de Velde CJ, Rea D, Seynaeve C, Putter H, Hasenburg A, Vannetzel JM, et al. Adjuvant tamoxifen and exemestane in early breast cancer (TEAM): A randomised phase 3 trial. Lancet. 2011;377(9762):321-31.
309. Ruhstaller T, Giobbie-Hurder A, Colleoni M, Jensen MB, Ejlertsen B, de Azambuja E, et al. Adjuvant letrozole and tamoxifen alone or sequentially for postmenopausal women with hormone receptor-positive breast cancer: Long-term follow-up of the BIG 1-98 trial. J Clin Oncol. 2019;37(2):105-14.
310. Cuzick J, Sestak I, Baum M, Buzdar A, Howell A, Dowsett M, et al. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. Lancet Oncol. 2010;11(12):1135-41.
311. Adjuvant bisphosphonate treatment in early breast cancer: meta-analyses of individual patient data from randomised trials. Lancet. 2015;386(10001):1353-61.
312. 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(2):122-37.
313. Pagani O, Walley BA, Fleming GF, Colleoni M, Láng I, Gomez HL, et al. Adjuvant Exemestane with ovarian suppression in premenopausal breast cancer: Long-term follow-up of the combined TEXT and SOFT trials. J Clin Oncol. 2023;41(7):1376-82.
314. Harbeck N, Rastogi P, Martin M, Tolaney SM, Shao ZM, Fasching PA, et al. Adjuvant abemaciclib combined with endocrine therapy for high-risk early breast cancer: Updated efficacy and Ki-67 analysis from the monarchE study. Ann Oncol. 2021;32(12):1571-81.
315. Johnston SR, Harbeck N, Hegg R, Toi M, Martin M, Shao ZM, et al. Abemaciclib combined with endocrine therapy for the adjuvant treatment of HR+, HER2−, node-positive, high-risk, early breast cancer (monarchE). J Clin Oncol. 2020;38(34):3987.

316. Rastogi P, O'Shaughnessy J, Martin M, Boyle F, Cortes J, Rugo HS, et al. Adjuvant Abemaciclib plus endocrine therapy for hormone receptor-positive, human epidermal growth factor receptor 2-negative, high-risk early breast cancer: Results from a preplanned monarchE overall survival interim analysis, including 5-Year efficacy outcomes. J Clin Oncol. 2024:Jco2301994.
317. Slamon D, Lipatov O, Nowecki Z, McAndrew N, Kukielka-Budny B, Stroyakovskiy D, et al. Ribociclib plus endocrine therapy in early breast cancer. N Engl J Med. 2024;390(12):1080-91.
318. Peto R, Davies C, Godwin J, Gray R, Pan HC, Clarke M, et al. Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet. 2012;379(9814):432-44.
319. Polychemotherapy for early breast cancer: An overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet. 1998;352(9132):930-42.
320. Anthracycline-containing and taxane-containing chemotherapy for early-stage operable breast cancer: A patient-level meta-analysis of 100 000 women from 86 randomised trials. Lancet. 2023;401(10384):1277-92.
321. Increasing the dose intensity of chemotherapy by more frequent administration or sequential scheduling: a patient-level meta-analysis of 37 298 women with early breast cancer in 26 randomised trials. Lancet. 2019;393(10179):1440-52.
322. Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: Meta-analysis of individual patient data from ten randomised trials. Lancet Oncol. 2018;19(1):27-39.
323. Geyer CE, Sikov WM, Huober J, Rugo HS, Wolmark N, O'Shaughnessy J, et al. Long-term efficacy and safety of addition of carboplatin with or without veliparib to standard neoadjuvant chemotherapy in triple-negative breast cancer: 4-year follow-up data from BrighTNess, a randomized phase III trial. Ann Oncol. 2022;33(4):384-94.
324. Rugo HS, Olopade OI, DeMichele A, Yau C, van 't Veer LJ, Buxton MB, et al. Adaptive randomization of veliparib-carboplatin treatment in breast cancer. N Engl J Med. 2016;375(1):23-34.
325. Yu K-D, Ye F-G, He M, Fan L, Ma D, Mo M, et al. Effect of adjuvant paclitaxel and carboplatin on survival in women with triple-negative breast cancer: a phase 3 randomized clinical trial. JAMA oncology. 2020;6(9):1390-6.
326. Loibl S, O'Shaughnessy J, Untch M, Sikov WM, Rugo HS, McKee MD, et al. Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to standard neoadjuvant chemotherapy in triple-negative breast cancer (BrighTNess): a randomised, phase 3 trial. Lancet Oncol. 2018;19(4):497-509.
327. Masuda N, Lee SJ, Ohtani S, Im YH, Lee ES, Yokota I, et al. Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. N Engl J Med. 2017;376(22):2147-59.
328. Harbeck N. Neoadjuvant and adjuvant treatment of patients with HER2-positive early breast cancer. Breast. 2022;62:S12-S6.

329. 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(4):320-68.
330. 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. Lancet Oncol. 2021;22(8):1139-50.
331. 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(16):1659-72.
332. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-åositive breast cancer. N Engl J Med. 2005;353(16):1673-84.
333. von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, et al. Adjuvant pertuzumab and trastuzumab in early HER2-åositive breast cancer. N Engl J Med. 2017;377(2):122-31.
334. Gianni L, Pienkowski T, Im Y-H, Tseng L-M, Liu M-C, Lluch A, et al. 5-year analysis of neoadjuvant pertuzumab and trastuzumab in patients with locally advanced, inflammatory, or early-stage HER2-positive breast cancer (NeoSphere): A multicentre, open-label, phase 2 randomised trial. Lancet Oncol. 2016;17(6):791-800.
335. Gianni L, Pienkowski T, Im Y-H, Roman L, Tseng L-M, Liu M-C, et al. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. Lancet Oncol. 2012;13(1):25-32.
336. 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(7):617-28.
337. 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(12):1688-700.
338. 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(1):80-91. e7.
339. Park JW, Liu MC, Yee D, Yau C, van 't Veer LJ, Symmans WF, et al. Adaptive Randomization of neratinib in early breast cancer. N Engl J Med. 2016;375(1):11-22.
340. Holmes FA, Moy B, Delaloge S, Chia SK, Ejlertsen B, Mansi J, et al. Overall survival with neratinib after trastuzumab-based adjuvant therapy in HER2-positive breast cancer (ExteNET): A randomised, double-blind, placebo-controlled, phase 3 trial. Eur J Cancer. 2023;184:48-59.
341. Hurvitz SA, Hegg R, Chung WP, Im SA, Jacot W, Ganju V, et al. Trastuzumab deruxtecan versus trastuzumab emtansine in patients with HER2-positive metastatic breast cancer: Updated results from DESTINY-Breast03, a randomised, open-label, phase 3 trial. Lancet. 2023;401(10371):105-17.

342. Modi S, Saura C, Yamashita T, Park YH, Kim SB, Tamura K, et al. Trastuzumab deruxtecan in previously treated HER2-positive breast cancer. N Engl J Med. 2020;382(7):610-21.
343. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. N Engl J Med. 2022;387(1):9-20.
344. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005;434(7035):917-21.
345. 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(10):1526-33.
346. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23(4):517-25.
347. Tutt ANJ, Garber JE, Kaufman B, Viale G, Fumagalli D, Rastogi P, et al. Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer. N Engl J Med. 2021;384(25):2394-405.
348. Geyer CE, Jr., Garber JE, Gelber RD, Yothers G, Taboada M, Ross L, et al. Overall survival in the OlympiA phase III trial of adjuvant olaparib in patients with germline pathogenic variants in BRCA1/2 and high-risk, early breast cancer. Ann Oncol. 2022;33(12):1250-68.
349. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer. 2012;12(4):252-64.
350. Nanda R, Liu MC, Yau C, Shatsky R, Pusztai L, Wallace A, et al. Effect of pembrolizumab plus neoadjuvant chemotherapy on pathologic complete response in women with early-stage breast cancer: An analysis of the ongoing phase 2 adaptively randomized I-SPY2 trial. JAMA Oncol. 2020;6(5):676-84.
351. Schmid P, Cortes J, Dent R, Pusztai L, McArthur H, Kümmel S, et al. Event-free survival with pembrolizumab in early triple-negative breast cancer. N Engl J Med. 2022;386(6):556-67.
352. Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, et al. Pembrolizumab for early triple-negative breast cancer. N Engl J Med. 2020;382(9):810-21.
353. Mittendorf EA, Zhang H, Barrios CH, Saji S, Jung KH, Hegg R, et al. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. Lancet. 2020;396(10257):1090-100.
354. Simón L, Campos A, Leyton L, Quest AFG. Caveolin-1 function at the plasma membrane and in intracellular compartments in cancer. Cancer Metastasis Rev. 2020;39(2):435-53.
355. Williams TM, Lisanti MP. Caveolin-1 in oncogenic transformation, cancer, and metastasis. Am J Physiol Cell Physiol. 2005;288(3):C494-506.

356. Patani N, Martin LA, Reis-Filho JS, Dowsett M. The role of caveolin-1 in human breast cancer. Breast Cancer Res Treat. 2012;131(1):1-15.
357. Low JY, Nicholson HD. Epigenetic modifications of caveolae associated proteins in health and disease. BMC Genet. 2015;16:71.
358. Huertas-Martínez J, Rello-Varona S, Herrero-Martín D, Barrau I, García-Monclús S, Sáinz-Jaspeado M, et al. Caveolin-1 is down-regulated in alveolar rhabdomyosarcomas and negatively regulates tumor growth. Oncotarget. 2014;5(20):9744-55.
359. Arpaia E, Blaser H, Quintela-Fandino M, Duncan G, Leong HS, Ablack A, et al. The interaction between caveolin-1 and Rho-GTPases promotes metastasis by controlling the expression of alpha5-integrin and the activation of Src, Ras and Erk. Oncogene. 2012;31(7):884-96.
360. Díaz-Valdivia NI, Calderón CC, Díaz JE, Lobos-González L, Sepulveda H, Ortíz RJ, et al. Anti-neoplastic drugs increase caveolin-1-dependent migration, invasion and metastasis of cancer cells. Oncotarget. 2017;8(67):111943-65.
361. Moreno-Càceres J, Caballero-Díaz D, Nwosu ZC, Meyer C, López-Luque J, Malfettone A, et al. The level of caveolin-1 expression determines response to TGF-β as a tumour suppressor in hepatocellular carcinoma cells. Cell Death Dis. 2017;8(10):e3098.
362. Pellinen T, Blom S, Sánchez S, Välimäki K, Mpindi JP, Azegrouz H, et al. ITGB1-dependent upregulation of Caveolin-1 switches TGFβ signalling from tumour-suppressive to oncogenic in prostate cancer. Sci Rep. 2018;8(1):2338.
363. Nwosu ZC, Ebert MP, Dooley S, Meyer C. Caveolin-1 in the regulation of cell metabolism: a cancer perspective. Mol Cancer. 2016;15(1):71.
364. Asterholm IW, Mundy DI, Weng J, Anderson RG, Scherer PE. Altered mitochondrial function and metabolic inflexibility associated with loss of caveolin-1. Cell Metab. 2012;15(2):171-85.
365. Yu DM, Jung SH, An HT, Lee S, Hong J, Park JS, et al. Caveolin-1 deficiency induces premature senescence with mitochondrial dysfunction. Aging Cell. 2017;16(4):773-84.
366. Fridolfsson HN, Kawaraguchi Y, Ali SS, Panneerselvam M, Niesman IR, Finley JC, et al. Mitochondria-localized caveolin in adaptation to cellular stress and injury. FASEB J. 2012;26(11):4637-49.
367. Shyu P, Jr., Wong XFA, Crasta K, Thibault G. Dropping in on lipid droplets: insights into cellular stress and cancer. Biosci Rep. 2018;38(5).
368. Antalis CJ, Arnold T, Rasool T, Lee B, Buhman KK, Siddiqui RA. High ACAT1 expression in estrogen receptor negative basal-like breast cancer cells is associated with LDL-induced proliferation. Breast Cancer Res Treat. 2010;122(3):661-70.
369. Antalis CJ, Uchida A, Buhman KK, Siddiqui RA. Migration of MDA-MB-231 breast cancer cells depends on the availability of exogenous lipids and cholesterol esterification. Clin Exp Metastasis. 2011;28(8):733-41.
370. Li J, Gu D, Lee SS, Song B, Bandyopadhyay S, Chen S, et al. Abrogating cholesterol esterification suppresses growth and metastasis of pancreatic cancer. Oncogene. 2016;35(50):6378-88.

371. Gao Y, Li L, Li T, Ma L, Yuan M, Sun W, et al. Simvastatin delays castration-resistant prostate cancer metastasis and androgen receptor antagonist resistance by regulating the expression of caveolin-1. Int J Oncol. 2019;54(6):2054-68.
372. Guruswamy S, Rao CV. Synergistic effects of lovastatin and celecoxib on caveolin-1 and its down-stream signaling molecules: Implications for colon cancer prevention. Int J Oncol. 2009;35(5):1037-43.
373. Chatterjee M, Ben-Josef E, Robb R, Vedaie M, Seum S, Thirumoorthy K, et al. Caveolae-Mediated Endocytosis Is Critical for Albumin Cellular Uptake and Response to Albumin-Bound Chemotherapy. Cancer Res. 2017;77(21):5925-37.
374. Wang Z, Wang N, Li W, Liu P, Chen Q, Situ H, et al. Caveolin-1 mediates chemoresistance in breast cancer stem cells via β-catenin/ABCG2 signaling pathway. Carcinogenesis. 2014;35(10):2346-56.
375. Zhao Y, Lv F, Chen S, Wang Z, Zhang J, Zhang S, et al. Caveolin-1 expression predicts efficacy of weekly nab-paclitaxel plus gemcitabine for metastatic breast cancer in the phase II clinical trial. BMC Cancer. 2018;18(1):1019.
376. Williams TM, Schneeweiss A, Jackisch C, Shen C, Weber KE, Fasching PA, et al. Caveolin gene expression predicts clinical outcomes for early-stage HER2-negative breast cancer treated with paclitaxel-based chemotherapy in the GeparSepto trial. Clin Cancer Res. 2023;29(17):3384-94.
377. Scatena C, Fanelli G, Fanelli GN, Menicagli M, Aretini P, Ortenzi V, et al. New insights in the expression of stromal caveolin 1 in breast cancer spread to axillary lymph nodes. Sci Rep. 2021;11(1):2755.
378. Witkiewicz AK, Dasgupta A, Sotgia F, Mercier I, Pestell RG, Sabel M, et al. An absence of stromal caveolin-1 expression predicts early tumor recurrence and poor clinical outcome in human breast cancers. Am J Pathol. 2009;174(6):2023-34.
379. Witkiewicz AK, Dasgupta A, Nguyen KH, Liu C, Kovatich AJ, Schwartz GF, et al. Stromal caveolin-1 levels predict early DCIS progression to invasive breast cancer. Cancer Biol Ther. 2009;8(11):1071-9.
380. Sloan EK, Ciocca DR, Pouliot N, Natoli A, Restall C, Henderson MA, et al. Stromal cell expression of caveolin-1 predicts outcome in breast cancer. Am J Pathol. 2009;174(6):2035-43.
381. Chung YC, Kuo JF, Wei WC, Chang KJ, Chao WT. Caveolin-1 Dependent Endocytosis Enhances the Chemosensitivity of HER-2 Positive Breast Cancer Cells to Trastuzumab Emtansine (T-DM1). PLoS One. 2015;10(7):e0133072.
382. Pereira PMR, Sharma SK, Carter LM, Edwards KJ, Pourat J, Ragupathi A, et al. Caveolin-1 mediates cellular distribution of HER2 and affects trastuzumab binding and therapeutic efficacy. Nat Commun. 2018;9(1):5137.
383. Chung YC, Chang CM, Wei WC, Chang TW, Chang KJ, Chao WT. Metformin-induced caveolin-1 expression promotes T-DM1 drug efficacy in breast cancer cells. Sci Rep. 2018;8(1):3930.

384. Indira Chandran V, Månsson AS, Barbachowska M, Cerezo-Magaña M, Nodin B, Joshi B, et al. Hypoxia attenuates trastuzumab uptake and trastuzumab-emtansine (T-DM1) cytotoxicity through redistribution of phosphorylated caveolin-1. Mol Cancer Res. 2020;18(4):644-56.
385. Bourseau-Guilmain E, Menard JA, Lindqvist E, Indira Chandran V, Christianson HC, Cerezo Magaña M, et al. Hypoxia regulates global membrane protein endocytosis through caveolin-1 in cancer cells. Nat Commun. 2016;7(1):11371.
386. Ha TK, Her NG, Lee MG, Ryu BK, Lee JH, Han J, et al. Caveolin-1 increases aerobic glycolysis in colorectal cancers by stimulating HMGA1-mediated GLUT3 transcription. Cancer Res. 2012;72(16):4097-109.
387. Martinez-Outschoorn UE, Sotgia F, Lisanti MP. Caveolae and signalling in cancer. Nat Rev Cancer. 2015;15(4):225-37.
388. Jin L, Shen F, Weinfeld M, Sergi C. Insulin growth factor binding protein 7 (IGFBP7)-related cancer and IGFBP3 and IGFBP7 Crosstalk. Front Oncol. 2020;10:727.
389. Murphy M, Pykett MJ, Harnish P, Zang KD, George DL. Identification and characterization of genes differentially expressed in meningiomas. Cell Growth Differ. 1993;4(9):715-22.
390. Evdokimova V, Tognon CE, Benatar T, Yang W, Krutikov K, Pollak M, et al. IGFBP7 binds to the IGF-1 receptor and blocks its activation by insulin-like growth factors. Sci Signal. 2012;5(255):ra92.
391. Artico LL, Laranjeira ABA, Campos LW, Corrêa JR, Zenatti PP, Carvalheira JBC, et al. Physiologic IGFBP7 levels prolong IGF1R activation in acute lymphoblastic leukemia. Blood Adv. 2021;5(18):3633-46.
392. Burger AM, Leyland-Jones B, Banerjee K, Spyropoulos DD, Seth AK. Essential roles of IGFBP-3 and IGFBP-rP1 in breast cancer. Eur J Cancer. 2005;41(11):1515-27.
393. Sato J, Hasegawa S, Akaogi K, Yasumitsu H, Yamada S, Sugahara K, et al. Identification of cell-binding site of angiomodulin (AGM/TAF/Mac25) that interacts with heparan sulfates on cell surface. J Cell Biochem. 1999;75(2):187-95.
394. Akaogi K, Okabe Y, Sato J, Nagashima Y, Yasumitsu H, Sugahara K, et al. Specific accumulation of tumor-derived adhesion factor in tumor blood vessels and in capillary tube-like structures of cultured vascular endothelial cells. Proc Natl Acad Sci U S A. 1996;93(16):8384-9.
395. St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, et al. Genes expressed in human tumor endothelium. Science. 2000;289(5482):1197-202.
396. Rosendahl AH, Björner S, Ygland Rödström M, Jirström K, Borgquist S, Ingvar C, et al. Pre- and postoperative circulating IGF-I, IGFBP-3, and IGFBP-7 levels in relation to endocrine treatment and breast cancer recurrence: A nested case-control study. Front Oncol. 2021;11(292).
397. Seth A, Kitching R, Landberg G, Xu J, Zubovits J, Burger AM. Gene expression profiling of ductal carcinomas in situ and invasive breast tumors. Anticancer Res. 2003;23(3A):2043-51.

398. Landberg G, Östlund H, Nielsen NH, Roos G, Emdin S, Burger AM, et al. Downregulation of the potential suppressor gene IGFBP-rP1 in human breast cancer is associated with inactivation of the retinoblastoma protein, cyclin E overexpression and increased proliferation in estrogen receptor negative tumors. Oncogene. 2001;20(27):3497-505.
399. Esseghir S, Kennedy A, Seedhar P, Nerurkar A, Poulsom R, Reis-Filho JS, et al. Identification of NTN4, TRA1, and STC2 as prognostic markers in breast cancer in a screen for signal sequence encoding proteins. Clin Cancer Res. 2007;13(11):3164-73.
400. Burger AM, Zhang X, Li H, Ostrowski JL, Beatty B, Venanzoni M, et al. Down-regulation of T1A12/mac25, a novel insulin-like growth factor binding protein related gene, is associated with disease progression in breast carcinomas. Oncogene. 1998;16(19):2459-67.
401. Keren P, Joshua RA, Tammy YNT, Kezia G, Trishna D, Chibuzor FO, et al. Identifying proteomic risk factors for cancer using prospective and exome analyses: 1,463 circulating proteins and risk of 19 cancers in the UK Biobank. medRxiv. 2023:2023.07.28.23293330.
402. Zhang L, Smyth D, Al-Khalaf M, Blet A, Du Q, Bernick J, et al. Insulin-like growth factor-binding protein-7 (IGFBP7) links senescence to heart failure. Nature Cardiovasc Res. 2022;1(12):1195-214.
403. Bracun V, van Essen B, Voors AA, van Veldhuisen DJ, Dickstein K, Zannad F, et al. Insulin-like growth factor binding protein 7 (IGFBP7), a link between heart failure and senescence. ESC Heart Fail. 2022;9(6):4167-76.
404. Shah AM, Myhre PL, Arthur V, Dorbala P, Rasheed H, Buckley LF, et al. Large scale plasma proteomics identifies novel proteins and protein networks associated with heart failure development. Nat Commun. 2024;15(1):528.
405. Gandhi PU, Gaggin HK, Sheftel AD, Belcher AM, Weiner RB, Baggish AL, et al. Prognostic usefulness of insulin-like growth factor-binding protein 7 in heart failure with reduced ejection fraction: A novel biomarker of myocardial diastolic function? Am J Card. 2014;114(10):1543-9.
406. Gandhi Parul U, Gaggin Hanna K, Redfield Margaret M, Chen Horng H, Stevens Susanna R, Anstrom Kevin J, et al. Insulin-like growth factor–binding protein-7 as a biomarker of diastolic dysfunction and functional capacity in heart failure with preserved ejection fraction. JACC: Heart Fail. 2016;4(11):860-9.
407. Chugh S, Ouzounian M, Lu Z, Mohamed S, Li W, Bousette N, et al. Pilot study identifying myosin heavy chain 7, desmin, insulin-like growth factor 7, and annexin A2 as circulating biomarkers of human heart failure. Proteomics. 2013;13(15):2324-34.
408. Rodier F, Campisi J, Bhaumik D. Two faces of p53: Aging and tumor suppression. Nucleic Acids Res. 2007;35(22):7475-84.
409. Boonstra J, Post JA. Molecular events associated with reactive oxygen species and cell cycle progression in mammalian cells. Gene. 2004;337:1-13.
410. Wajapeyee N, Serra RW, Zhu X, Mahalingam M, Green MR. Oncogenic BRAF induces senescence and apoptosis through pathways mediated by the secreted protein IGFBP7. Cell. 2008;132(3):363-74.

411. Zuo S, Liu C, Wang J, Wang F, Xu W, Cui S, et al. IGFBP-rP1 induces p21 expression through a p53-independent pathway, leading to cellular senescence of MCF-7 breast cancer cells. J Cancer Res Clin Oncol. 2012;138(6):1045-55.
412. Lopez-Bermejo A, Khosravi J, Fernández-Real JM, Hwa V, Pratt KL, Casamitjana R, et al. Insulin resistance is associated with increased serum concentration of IGF-binding protein–related protein 1 (IGFBP-rP1/MAC25). Diabetes. 2006;55(8):2333-9.
413. Januzzi JL, Mohebi R, Liu Y, Sattar N, Heerspink HJL, Tefera E, et al. Cardiorenal biomarkers, canagliflozin, and outcomes in diabetic kidney disease: The CREDENCE trial. Circulation. 2023;148(8):651-60.
414. Januzzi JL, Jr., Butler J, Sattar N, Xu J, Shaw W, Rosenthal N, et al. Insulin-like growth factor binding protein 7 predicts renal and cardiovascular outcomes in the canagliflozin cardiovascular assessment study. Diabetes Care. 2020;44(1):210-6.
415. Kashani K, Al-Khafaji A, Ardiles T, Artigas A, Bagshaw SM, Bell M, et al. Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. Crit Care. 2013;17(1):R25.
416. Schulz KF, Grimes DA, Horton RC, Schulz KF. Essential concepts in clinical research : randomised controlled trials and observational epidemiology / Kenneth Schulz, David A. Grimes ; foreword by Richard Horton. Second edition. ed. Edinburgh: Elsevier; 2019.
417. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58(5):295-300.
418. Sandén E, Khazaei S, Tryggvadottir H, Borgquist S, Isaksson K, Jirström K, et al. Re-evaluation of HER2 status in 606 breast cancers-gene protein assay on tissue microarrays versus routine pathological assessment. Virchows Arch. 2020.
419. 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(2):e158-e68.
420. 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(1):20.
421. 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):94.
422. Wolf DM, Yau C, Wulfkuhle J, Brown-Swigart L, Gallagher RI, Lee PRE, et al. Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies. Cancer Cell. 2022;40(6):609-23.e6.

423. Pusztai L, Yau C, Wolf DM, Han HS, Du L, Wallace AM, et al. Durvalumab with olaparib and paclitaxel for high-risk HER2-negative stage II/III breast cancer: Results from the adaptively randomized I-SPY2 trial. Cancer Cell. 2021;39(7):989-98.e5.
424. Yee D, DeMichele AM, Yau C, Isaacs C, Symmans WF, Albain KS, et al. Association of Event-Free and Distant Recurrence-Free Survival With Individual-Level Pathologic Complete Response in Neoadjuvant Treatment of Stages 2 and 3 Breast Cancer: Three-Year Follow-up Analysis for the I-SPY2 Adaptively Randomized Clinical Trial. JAMA Oncol. 2020;6(9):1355-62.
425. Wulfkuhle JD, Yau C, Wolf DM, Vis DJ, Gallagher RI, Brown-Swigart L, et al. Evaluation of the HER/PI3K/AKT family signaling network as a predictive biomarker of pathologic complete response for patients with breast cancer treated with neratinib in the I-SPY 2 trial. JCO Precis Oncol. 2018;2.
426. Ringnér M, Fredlund E, Häkkinen J, Borg Å, Staaf J. GOBO: Gene expression-based outcome for breast cancer online. PLoS One. 2011;6(3):e17911.
427. Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res. 2011;13(5):R97.
428. 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(7):e47.
429. Hu C, Dignam JJ. Biomarker-Driven Oncology Clinical Trials: Key Design Elements, Types, Features, and Practical Considerations. JCO Precis Oncol. 2019;3.
430. Sukswai N, Khoury JD. Immunohistochemistry innovations for diagnosis and tissue-based biomarker detection. Curr Hematol Malig Rep. 2019;14(5):368-75.
431. Yagi H, Nishigori M, Murakami Y, Osaki T, Muto S, Iba Y, et al. Discovery of novel biomarkers for atherosclerotic aortic aneurysm through proteomics-based assessment of disease progression. Sci Rep. 2020;10(1):6429.
432. Kononen J, Bubendorf L, Kallioniemi A, Bärlund M, Schraml P, Leighton S, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med. 1998;4(7):844-7.
433. Meyerholz DK, Beck AP. Principles and approaches for reproducible scoring of tissue stains in research. Lab Invest. 2018;98(7):844-55.
434. Sauter G, Simon R, Hillan K. Tissue microarrays in drug discovery. Nat Rev Drug Discov. 2003;2(12):962-72.
435. 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(1):19737.
436. Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. PLoS One. 2017;12(12):e0190152.
437. Chaitankar V, Karakülah G, Ratnapriya R, Giuste FO, Brooks MJ, Swaroop A. Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. Prog Retin Eye Res. 2016;55:1-31.

438. Lima LS, Galiciolli MEA, Pereira ME, Felisbino K, Machado-Souza C, de Oliveira CS, et al. Modification by genetic polymorphism of lead-induced IQ alteration: a systematic review. Environ Sci Pollut Res. 2022;29(29):43435-47.
439. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A network for understanding the genetic architecture of common cancers. Cancer Epidemiol Biomarkers Prev. 2017;26(1):126-35.
440. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2(8):e124.
441. Ioannidis JP. Why most discovered true associations are inflated. Epidemiology. 2008;19(5):640-8.
442. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. PLoS Med. 2007;4(10):e297.
443. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc. 1999;94(446):496-509.
444. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011;45(3):1 - 67.
445. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
446. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9.
447. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353-d61.
448. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.
449. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45(10):1113-20.
450. Godina C, Belting M, Vallon-Christersson J, Isaksson K, Bosch A, Jernström H. Caveolin-1 gene expression provides additional prognostic information combined with PAM50 risk of recurrence (ROR) score in breast cancer. Sci Rep. 2024;14(1):6675.
451. Pilch PF, Liu L. Fat caves: caveolae, lipid trafficking and lipid metabolism in adipocytes. Trends Endocrinol Metab. 2011;22(8):318-24.
452. Razani B, Combs TP, Wang XB, Frank PG, Park DS, Russell RG, et al. Caveolin-1-deficient mice are lean, resistant to diet-induced obesity, and show hypertriglyceridemia with adipocyte abnormalities. J Biol Chem. 2002;277(10):8635-47.

453. Mukherjee R, Kim SW, Choi MS, Yun JW. Sex-dependent expression of caveolin 1 in response to sex steroid hormones is closely associated with development of obesity in rats. PLoS One. 2014;9(3):e90918.
454. Cohen AW, Razani B, Wang XB, Combs TP, Williams TM, Scherer PE, et al. Caveolin-1-deficient mice show insulin resistance and defective insulin receptor protein expression in adipose tissue. Am J Physiol Cell Physiol. 2003;285(1):C222-35.
455. Shiroto T, Romero N, Sugiyama T, Sartoretto JL, Kalwa H, Yan Z, et al. Caveolin-1 is a critical determinant of autophagy, metabolic switching, and oxidative stress in vascular endothelium. PLoS One. 2014;9(2):e87871.
456. Sonveaux P, Martinive P, DeWever J, Batova Z, Daneau G, Pelat M, et al. Caveolin-1 expression is critical for vascular endothelial growth factor-induced ischemic hindlimb collateralization and nitric oxide-mediated angiogenesis. Circ Res. 2004;95(2):154-61.
457. Wang S, Wang N, Zheng Y, Yang B, Liu P, Zhang F, et al. Caveolin-1 inhibits breast cancer stem cells via c-Myc-mediated metabolic reprogramming. Cell Death Dis. 2020;11(6):450.
458. Jögi A, Ehinger A, Hartman L, Alkner S. Expression of HIF-1α is related to a poor prognosis and tamoxifen resistance in contralateral breast cancer. PLoS One. 2019;14(12):e0226150.
459. Tutzauer J, Sjöström M, Holmberg E, Karlsson P, Killander F, Leeb-Lundberg LMF, et al. Breast cancer hypoxia in relation to prognosis and benefit from radiotherapy after breast-conserving surgery in a large, randomised trial with long-term follow-up. Br J Cancer. 2022;126(8):1145-56.
460. Goetz JG, Minguet S, Navarro-Lérida I, Lazcano JJ, Samaniego R, Calvo E, et al. Biomechanical remodeling of the microenvironment by stromal caveolin-1 favors tumor invasion and metastasis. Cell. 2011;146(1):148-63.
461. Luca BA, Steen CB, Matusiak M, Azizi A, Varma S, Zhu C, et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell. 2021;184(21):5482-96.e28.
462. Schmitt CA, Wang B, Demaria M. Senescence and cancer — role and therapeutic opportunities. Nature Rev Clin Oncol. 2022;19(10):619-36.
463. Koelwyn GJ, Newman AAC, Afonso MS, van Solingen C, Corr EM, Brown EJ, et al. Myocardial infarction accelerates breast cancer via innate immune reprogramming. Nat Med. 2020;26(9):1452-8.
464. Meijers WC, Maglione M, Bakker SJL, Oberhuber R, Kieneker LM, de Jong S, et al. Heart Failure Stimulates Tumor Growth by Circulating Factors. Circulation. 2018;138(7):678-91.
465. Ekyalongo RC, Yee D. Revisiting the IGF-1R as a breast cancer target. NPJ Precis Oncol. 2017;1.
466. Akshintala S, Sundby RT, Bernstein D, Glod JW, Kaplan RN, Yohe ME, et al. Phase I trial of ganitumab plus dasatinib to cotarget the insulin-like growth factor 1 receptor and Src family kinase YES in rhabdomyosarcoma. Clin Cancer Res. 2023;29(17):3329-39.