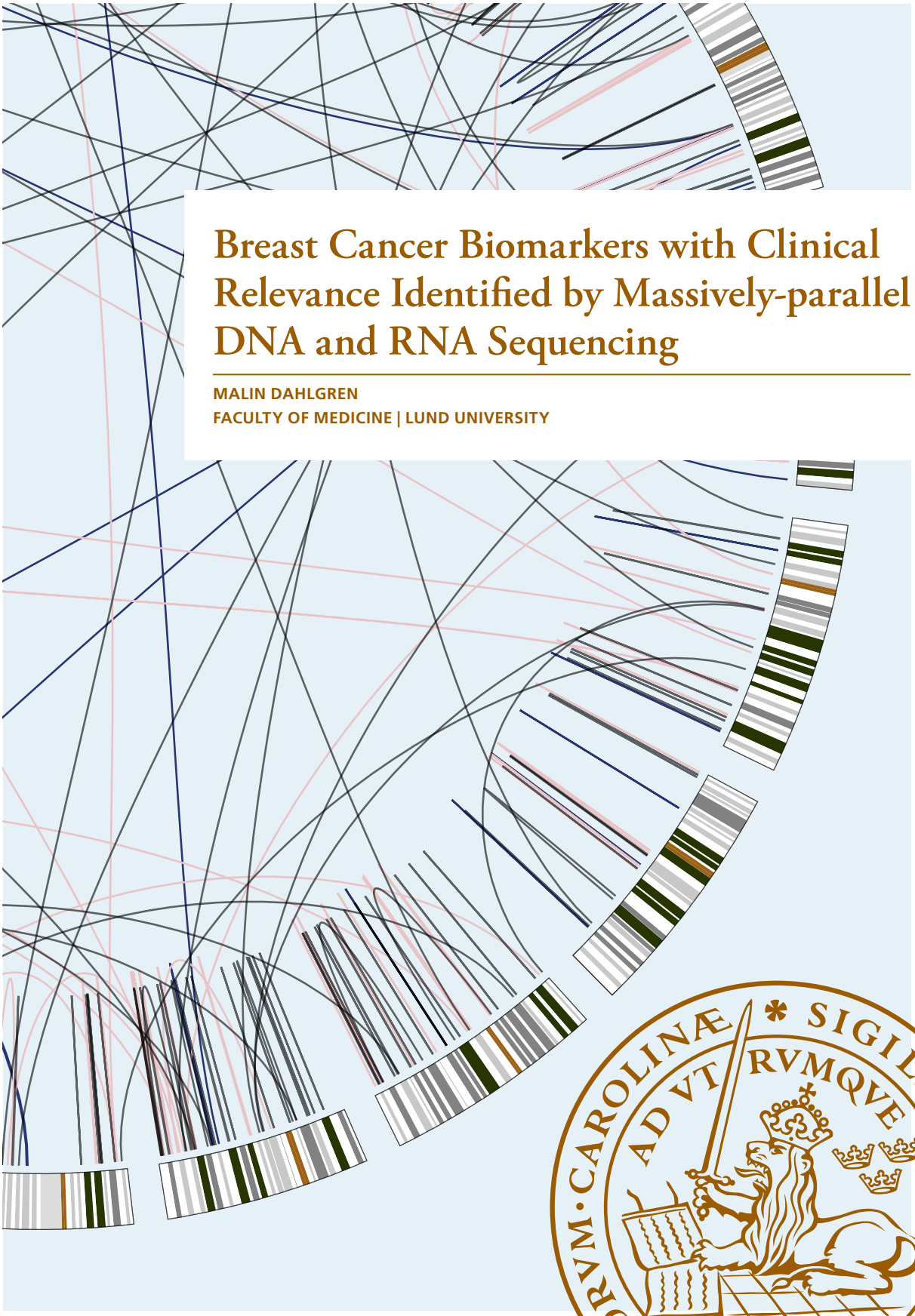
Breast Cancer Biomarkers with Clinical Relevance Identified by Massively-parallel DNA and RNA Sequencing

Dahlgren, Malin

2021

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



Breast Cancer Biomarkers with Clinical Relevance Identified by Massively-parallel DNA and RNA Sequencing

MALIN DAHLGREN
FACULTY OF MEDICINE | LUND UNIVERSITY

Breast Cancer Biomarkers with Clinical Relevance Identified by Massively-parallel DNA and RNA Sequencing

Breast cancer was the most commonly diagnosed form of cancer in the world in 2020 and 10% of women will develop breast cancer during their lifetime. Although the prognosis is relatively good, breast cancer is still a significant contributor to morbidity and death, both in Sweden and worldwide. Cancer is thought of as a disease with genomic aberrations at its root, but many aspects of tumor genomics and its effects on patient prognosis, treatment response, and risk for relapse are still not fully understood. In this thesis, we have employed massively-parallel sequencing methods to identify genomic aberrations in breast cancer with clinical implications.

We have characterized chromosomal rearrangements and investigated the similarity between primary tumors and distant metastases as well as between contralateral tumors. Moreover, we have interrogated RNA sequencing data from a collection of more than 3000 primary breast cancers from the SCAN-B study and detected known hormonal treatment resistance mutations in estrogen receptor alpha as well as characterized the transcriptional levels and impact on prognosis of estrogen receptor beta. This research contributes to further increasing our understanding of tumor dynamics and evolution and moreover highlights breast tumor biomarkers that may carry relevance for clinical diagnostics and treatment decisions.

Printed by Media-Tryck, Lund 2021 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2021:115
ISBN 978-91-8021-122-2
ISSN 1652-8220

Breast Cancer Biomarkers with Clinical Relevance Identified by Massively-parallel DNA and RNA Sequencing

Malin Dahlgren

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Segerfalk Lecture Hall, Thursday November 4th, 2021.

Faculty opponent
Antonis Valachis, MD PhD
Örebro University Hospital

Organization	Document name
LUND UNIVERSITY	Doctoral dissertation
Faculty of Medicine Department of Clinical Sciences Lund Division of Oncology	**Date of issue** 2021-11-04
Author Malin Dahlgren	Sponsoring organization

Title and subtitle Breast Cancer Biomarkers with Clinical Relevance Identified by Massively-parallel DNA and RNA Sequencing

Abstract

Women have a 10% lifetime risk of developing breast cancer, and the disease has surpassed lung cancer as the most frequently diagnosed type of cancer in the world. Breast cancer originates in the epithelial cells of the mammary gland and tumor cells have undergone a series of genetic and phenotypic changes that confer tumor promoting properties.

Genomic rearrangement is a common phenomenon in cancer, involving breakage and dysfunctional repair of chromosomes. With the aim to characterize such variants and their progression from primary to metastatic disease, we performed whole-genome sequencing of paired primary tumors and metastases (study I) and paired contralateral breast cancers (CBC) (study II). Metastasis rearrangement profiles bore a remarkable resemblance to the respective primary tumors (median 89% shared), indicating that the rearrangements were early events in tumor development, remaining stable throughout progression. Our study on CBC (study II) subsequently allowed us to identify 1 in 10 tumor pairs that likely represented metastatic spread rather than a new primary tumor (76% of rearrangements shared).

One of the risk factors for breast cancer is high exposure to estrogens; signaling via estrogen receptor (ER) α is considered the most important driver for the 75% of tumors expressing this marker. Mutations in the gene for ERα are known to be common in endocrine therapy-refractory breast cancer and confer resistance to standard anti-hormonal treatment. In study III, we interrogated RNA-seq data from 3217 primary breast tumors from the SCAN-B initiative and found that 1% of tumors were positive for one of the mutations at surgery. For those patients that received adjuvant endocrine therapy, the mutations were associated to worse overall and relapse-free survival.

In study IV, we further explored the SCAN-B dataset to investigate the phenotypic properties and prognosis associated to high expression of the much less well studied ERβ. We discovered that this receptor was not abundantly expressed, with 1/3 of tumors entirely negative. Further, we saw that patients with high levels of ERβ mRNA had slightly improved overall survival and that the expression of ERβ was associated to expression of genes involved in immune cell activation.

In summary, we have employed sequencing technology to study breast cancer patient material to identify and assess the validity of genomic and transcriptomic changes that may both be of value as potential biomarkers, and in elucidating biological mechanisms that drive or suppress breast cancer progression.

Key words Breast cancer, DNA sequencing, RNA sequencing, Metastasis, Biomarker, Estrogen receptor alpha, Estrogen receptor beta, ESR1, ESR2

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN and key title 1652-8220		**ISBN** 978-91-8021-122-2
Recipient's notes	**Number of pages** 97	Price
	Security classification	

Signature  Date 2021-09-22

Breast Cancer Biomarkers with Clinical Relevance Identified by Massively-parallel DNA and RNA Sequencing

Malin Dahlgren

LUND
UNIVERSITY

Cover illustration: Circos plot of structural variants in the primary tumor and metastasis of a breast cancer patient. (Adapted from Study I, Supplementary Figure S1 P18, CC-BY 3.0)

Faculty of Medicine
Department of Clinical Sciences Lund

Lund University, Faculty of Medicine Doctoral Dissertation Series 2021:115
ISBN 978-91-8021-122-2
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2021

"Coming back to where you started
is not the same as never leaving."

Terry Pratchett in *A Hat Full of Sky* (2004)

Table of Contents

List of studies

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I **Remarkable similarities of chromosomal rearrangements between primary human breast cancers and matched distant metastases as revealed by whole-genome sequencing.**

Tang MH*, **Dahlgren M***, Brueffer C, Tjitrowirjo T, Winter C, Chen Y, Olsson E, Wang K, Törngren T, Sjöström M, Grabau D, Bendahl PO, Rydén L, Nimeus E, Saal LH, Borg Å, Gruvberger-Saal SK.

Oncotarget, 2015. **6**(35): p. 37169-84

II **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.**

Alkner S*, Tang MH*, Brueffer C, **Dahlgren M**, Chen Y, Olsson E, Winter C, Baker S, Ehinger A, Rydén L, Saal LH, Fernö M, Gruvberger-Saal SK.

Breast Cancer Res, 2015. **17**(1): 102

III **Pre-existing somatic mutations of estrogen receptor alpha (ESR1) in early-stage primary breast cancer predict failure of adjuvant endocrine therapy and poor survival.**

Dahlgren M, George AM, Brueffer C, Gladchuk S, Chen Y, Vallon-Christersson J, Hegardt C, Häkkinen J, Rydén L, Malmberg M, Larsson C, Gruvberger-Saal SK, Ehinger A, Loman N, Borg Å, Saal LH.

JNCI Cancer Spectrum, 2021. **5**(2)

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

Dalal H, **Dahlgren M**, Gladchuk S, Brueffer C, Gruvberger-Saal SK, Saal LH.

Manuscript.

* = equal contribution

Study not included in the thesis

Serial monitoring of circulating tumor DNA in patients with primary breast cancer for detection of occult metastatic disease.
Olsson E*, Winter C*, George A, Chen Y, Howlin J, Tang MH, **Dahlgren M**, Schulz R, Grabau D, van Westen D, Fernö M, Ingvar C, Rose C, Bendahl PO, Rydén L, Borg Å, Gruvberger-Saal SK, Jernström H, Saal, LH.
EMBO Mol Med, 2015. **7**(8): p. 1034-47

* = equal contribution

Abstract

Women have a 10% lifetime risk of developing breast cancer, and the disease has surpassed lung cancer as the most frequently diagnosed type of cancer in the world. Breast cancer originates in the epithelial cells of the mammary gland and tumor cells have undergone a series of genetic and phenotypic changes that confer tumor promoting properties.

Genomic rearrangement is a common phenomenon in cancer, involving breakage and dysfunctional repair of chromosomes. With the aim to characterize such variants and their progression from primary to metastatic disease, we performed whole-genome sequencing of paired primary tumors and metastases (study I) and paired contralateral breast cancers (CBC) (study II). Metastasis rearrangement profiles bore a remarkable resemblance to the respective primary tumors (median 89% shared), indicating that the rearrangements were early events in tumor development, remaining stable throughout progression. Our study on CBC (study II) subsequently allowed us to identify 1 in 10 tumor pairs that likely represented metastatic spread rather than a new primary tumor (76% of rearrangements shared).

One of the risk factors for breast cancer is high exposure to estrogens; signaling via estrogen receptor (ER) α is considered the most important driver for the 75% of tumors expressing this marker. Mutations in the gene for ERα are known to be common in endocrine therapy-refractory breast cancer and confer resistance to standard anti-hormonal treatment. In study III, we interrogated RNA-seq data from 3217 primary breast tumors from the SCAN-B initiative and found that 1% of tumors were positive for one of the mutations at surgery. For those patients that received adjuvant endocrine therapy, the mutations were associated to worse overall and relapse-free survival.

In study IV, we further explored the SCAN-B dataset to investigate the phenotypic properties and prognosis associated to high expression of the much less well studied ERβ. We discovered that this receptor was not abundantly expressed, with 1/3 of tumors entirely negative. Further, we saw that patients with high levels of ERβ mRNA had slightly improved overall survival and that the expression of ERβ was associated to expression of genes involved in immune cell activation.

In summary, we have employed sequencing technology to study breast cancer patient material to identify and assess the validity of genomic and transcriptomic changes that may both be of value as potential biomarkers, and in elucidating biological mechanisms that drive or suppress breast cancer progression.

Populärvetenskaplig sammanfattning

Bröstcancer var den vanligast diagnosticerade cancertypen i världen år 2020, med mer än 2,2 miljoner nya fall. Årligen i Sverige diagnosticeras ungefär 8000 nya fall av bröstcancer och cirka 1400 personer dör i sjukdomen. Cancer är en sjukdom där celler i kroppen börjar dela sig snabbare än normalt något som i bröstcancer kan yttra sig som en knöl i bröstet, d v s en tumör. Elakartade cancerceller kan också erhålla förmågan att sprida sig till andra delar av kroppen och bilda nya tumörer – metastaser – och i det stadiet anses sjukdomen i princip vara obotlig. En viktig orsak till utveckling av såväl bröstcancer som andra cancerformer är att celler utvecklar förändringar i arvsmassan (DNA), som sedan nedärvs under generationer av celldelningar. Dessa förändringar är ett resultat av genetiska skador och kan exempelvis bestå av trasiga och omkastade kromosomer, så kallade kromosomala rearrangemang, eller förändringar av enstaka molekyler i DNA, så kallade punktmutationer. Vissa av förändringarna är aktiva och driver på utvecklingen av tumören, medan andra är passiva och påverkar inte cancersjukdomens förlopp. Genom forskning på mutationer och andra förändringar i cancer kan man få ledtrådar om hur tumören har utvecklats. Om tumören sprids till en ny plats i kroppen kan man också analysera DNA i metastasen för att se hur den skiljer sig från den ursprungliga cancern och försöka hitta faktorer som ökar risken för metastassjukdom.

Avhandlingen omfattar fyra studier. I den första studien har vi karakteriserat kromosomala rearrangemang, hos elva bröstcancerpatienter och studerat hur de skiljer sig mellan den primära – ursprungliga – brösttumören och metastasen. Vi fann att förändringarna ser ganska likartade ut i primärtumör jämfört med en metastas från samma patient, vilket kan vara ett tecken på att de har uppstått tidigt under cancerns utveckling och är gemensamma för de flesta cellerna i en brösttumör.

I nästa del av avhandlingen jämförde vi kromosomala rearrangemang från tumörer hos tio patienter som fått cancer i båda brösten, så kallad kontralateral bröstcancer. När en patient drabbas av detta är det viktigt att ta reda på ifall det rör sig om en spridning av den första tumören eller om det är en helt ny tumör, eftersom prognos och behandling ser annorlunda ut för de två diagnoserna. Genom att jämföra med data från vår tidigare studie på metastaser, kunde vi visa att minst en av de tio patienterna i studien som diagnosticerats med kontralateral bröstcancer i själva verket hade en metastas. Om även framtida forskning styrker att kontralateral cancer kan representera spridning av den första tumören, skulle det vara relevant att överväga tumörgenetisk diagnostik för dessa patienter för att bättre veta vilken behandling som är mest lämpad i varje enskilt fall.

De två sista studierna är baserade på data från SCAN-B-initiativet. SCAN-B initierades 2010 och erbjuder deltagande till alla patienter med nydiagnostiserad bröstcancer vid nio svenska cancerkliniker med målet är att utföra genetisk sekvensanalys på tumörmaterialet. Idag har mer än 17 000 patienter gått med och 11 000 tumörer har analyserats.

I den tredje studien studerade vi punktmutationer i östrogenreceptor alfa (ERα). Tre fjärdedelar av alla brösttumörer är beroende av signalering med könshormonet östrogen för att kunna växa. En förutsättning för detta är att de har en mottagare – receptor – för östrogen (ERα) som östrogen binder till och aktiverar, vilket sätter i gång cellens svar på östrogensignalen. Därför är antihormonella läkemedel som blockerar ERα en viktig del i behandlingen för patienter vars tumörer har denna receptor. Men ibland utvecklas mutationer i ERα som gör att medicinerna inte fungerar. Dessa mutationer har tidigare nästan uteslutande hittats i ett senare skede av sjukdomen och har visat sig mycket ovanlig innan tumören har utsatts för hormonbehandling.

För att undersöka förekomsten och relevansen av resistensmutationer som finns redan vid diagnos av bröstcancer använde vi sekvensdata från mer än 3200 tumörer från SCAN-B-studien. Vi fann att resistensmutationer i ERα förekom i ca 1% av brösttumörerna och analyser visade att de inte var medfödda mutationer utan att de hade utvecklats i tumören. Vi såg dessutom att de patienter som hade resistensmutationerna oftare fick återfall och oftare avled under uppföljningstiden, vilket kan vara ett tecken på sämre svar på den antihormonella behandling som gavs. Även om mutationerna är relativt ovanliga skulle det alltså kunna vara relevant att som del av diagnostiken leta efter de här mutationerna för att hitta de patienter som kan tänkas behöva alternativa behandlingar.

Slutligen, i den sista studien använde vi data från de 3200 SCAN-B-patienterna för att studera ERαs ”kusin” – östrogenreceptor β (ERβ). Denna receptor har mer otydlig roll i bröstcancer och det finns motstridiga åsikter om huruvida dess roll är att undertrycka eller öka aggressiviteten i cancer. Vi mätte förekomsten av ERβ-RNA som ett mått på hur mycket av receptorn som tumörcellerna försöker producera och fann att en tredjedel av tumörerna i princip saknade ERβ-RNA. Den tredjedel av patienterna som hade högst mängd ERβ-RNA hade en något bättre överlevnad jämfört med övriga patienter. De här resultaten stödjer hypotesen att ERβ har anti-canceregenskaper.

Sammanfattningsvis har vi karakteriserat olika typer av genetiska förändringar och i brösttumörer och undersökt huruvida de skulle kunna användas för att studera tumörers utveckling och för att förbättra diagnosmetoderna för patienter med bröstcancer. Vi har också studerat förekomsten och betydelsen av ERβ i bröstcancer. I framtiden kan denna kunskap bli viktig för att åstadkomma mer detaljerad profilering av brösttumörer, identifiera nya mål för anti-cancerbehandling samt möjliggöra personligt skräddarsydda behandlingsplaner.

Abbreviations

AI	Aromatase inhibitor
AJCC	American Joint Committee on Cancer
AF-1	Activation Function 1
AF-2	Activation Function 2
AP-1	Activator protein 1
BAM	Binary alignment/map file format
BER	Base excision repair
bp	Basepairs
CBC	Contralateral breast cancer
CNV	Copy number variant
DNA	Deoxyribonucleic acid
ER	Estrogen receptor
E2	Estradiol
ERα	Estrogen receptor α
ERβ	Estrogen receptor β
ERE	Estrogen response element
ESMO	European Society for Medical Oncology
ESR1	The gene symbol for estrogen receptor α
ESR2	The gene symbol for estrogen receptor β
FPKM	Fragments per kilobase of exon per million mapped reads
HRR	Homologous recombination repair
HER2	Human epidermal growth factor receptor 2 (gene symbol: *ERBB2*)
HR	Hazard ratio
HRR	Homologous recombination repair
IHC	Immunohistochemistry
Indel	Short insertion or deletion
kb	Kilobases/kilobasepairs
KM	Kaplan-Meier
LBD	Ligand-binding domain
MMR	Mismatch repair
mRNA	Messenger ribonucleic acid
NER	Nucleotide excision repair
NHG	Nottingham histological grade

NHEJ	Non-homologous end-joining repair
OS	Overall survival
PAM50	Prediction Analysis of Microarrays 50 gene signature
PARP	Poly-(ADP-ribose) polymerase
PCR	Polymerase chain reaction; ddPCR: droplet digital PCR
PI3K	Phosphoinositide 3-kinase
PR	Progesterone receptor (gene symbol: *PGR*)
RFI	Relapse-free interval
RFS	Relapse-free survival
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RPKM	Reads per kilobase of exon per million mapped reads
SCAN-B	Sweden Cancerome Analysis Network – Breast
SERM	Selective estrogen receptor modulator
SERD	Selective estrogen receptor degrader
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SV	Structural variant
TCGA	The Cancer Genome Atlas
TNBC	Triple-negative breast cancer
TNM	AJCC Tumor, Node, Metastasis staging system
TPM	Transcripts per million reads

Unless otherwise defined, genes and gene products are referred to by their gene symbols according to the HUGO Gene Nomenclature Committee (https://www.genenames.org/).

Aims of the thesis

General aims

- Characterize genomic and transcriptomic changes in breast cancer and their possible roles for
    - tracking the evolution of tumors
    - use as biomarkers
    - predictors of response and outcome

Specific aims

- Identify large-scale genomic changes in the form of chromosomal rearrangements and evaluate the clonal relationship between primary tumors and paired distant metastases (study I).
- Using chromosomal rearrangements for determining clonal relationship between contralateral breast tumors (study II)
- Detecting *ESR1* mutations that confer resistance to endocrine therapy in primary breast cancer and determine whether they have an effect on outcome (study III).
- Characterizing mRNA expression profile of estrogen receptor β (ERβ) in breast cancer and its clinical associations (study IV).

Introduction

Cancer

> "The body is a cell state in which every cell is a citizen. Disease is merely the conflict of the citizens of the state brought about by the action of external forces."
>
> Rudolf Virchow in *Die Cellularpathologie* (1858),
> via *A Dictionary of Scientific Quotations* [1]

The word cancer has roots in Latin (cancer) and Greek (karkinos), in both cases signifying "crab". The term, dating to Hippocrates, is descriptive of the hardness, the symptoms of pain, and later of the pattern of swollen blood vessels, reminiscent of a crustacean, observed in some tumors. The word tumor (from Latin "swelling" or "lump") refers to a mass of abnormally growing cells. Solid tumors may be benign, without capacity for damaging or invading surrounding tissues, but in the case of malignant cancer, the tumor cells are more aggressive, often with lethal results should the condition remain untreated.

The development of cancer is thought of as a step-wise accumulation of cellular and genomic damage, throughout generations of cell division, gradually disabling the intrinsic control mechanisms for growth. Tissues in the body that have a high rate of self-renewal or are more exposed to damaging foreign agents, like the skin, lungs and intestinal mucosa, are particularly vulnerable. Although some pre-cancerous or cancerous cells will perish as a result of cellular and genomic injury, those that survive can give rise to yet other cancer cells, making the tumor a micro-ecosystem where aggressive and proliferative properties can undergo positive selection.

All multicellular organisms are vulnerable to the development of tumors. For humans, who have a relatively long lifespan, cancerous disease is common. In Sweden, it is estimated that one in three persons will at some point during their lifetime be diagnosed with cancer, and out of those around 25% will die within five years [2]. Worldwide, cancer is the second leading cause of death [3] with an estimated 10 million deaths in 2020 [4].

Cancer causes morbidity and mortality through several mechanisms. Locally growing tumors can compromise the integrity of the organ of origin through the process of inflammation, tissue destruction, and suppression of healthy and functioning tissue. Additionally, tumor cells may consume nutrients to such an extent that the host is starved. Some tumors produce signaling molecules that disrupt the homeostasis of the entire body, including inflammatory factors and hormones. Once cancer starts to spread, not only does the tumor burden increase, but additional tissues may suffer destruction, resulting in organ failure and eventually death.

Hallmarks of cancer

Human cells are part of a cooperative system and, for the preservation of the organism as a whole, they are subject to rigid control mechanisms that prevent them from reverting to selfish, survival-of-the-fittest behavior. In the reviews "Hallmarks of cancer" and "Hallmarks of cancer – the next generation" [5, 6], Hanahan and Weinberg describe the principal abilities that a cancer cell acquires in order to survive the antitumor environment of the human body (Figure 1). This includes both changes in individual cellular behavior and interaction with surrounding tissues and the immune system.

Cancer cells disable or circumvent intrinsic and extrinsic apoptotic pathways as well as develop independence from growth signals, either by producing growth factors themselves or by hijacking the downstream intracellular pathways. They also achieve replicative immortality, through maintenance and regeneration of chromosomal telomeres, to enable indefinite mitotic cycling. The cellular metabolism is reprogrammed to extract energy and metabolites through the process of glycolysis and the increased need for nutrients and oxygen is addressed through the induction of new blood vessel growth. Tumor cells can also create a tumor-permissive microenvironment, ignore contact inhibition signals from neighboring cells and may ultimately, through replicative number as well as upregulation of proteolytic enzymes, invade surrounding tissues.

Of particular importance in this thesis is the hallmark "Genomic instability and mutation". Not only is this one of the ways in which tumor cells can adapt to and survive new challenges, like tumoricidal pharmacological therapies, but cancer is also largely thought of as a disease with genetic and genomic aberrations at its root. As will be described later in this thesis, our work on estrogen receptors is also linked to several hallmarks, particularly "Evading growth suppressors", "Sustaining proliferative signaling", and "Resisting cell death".

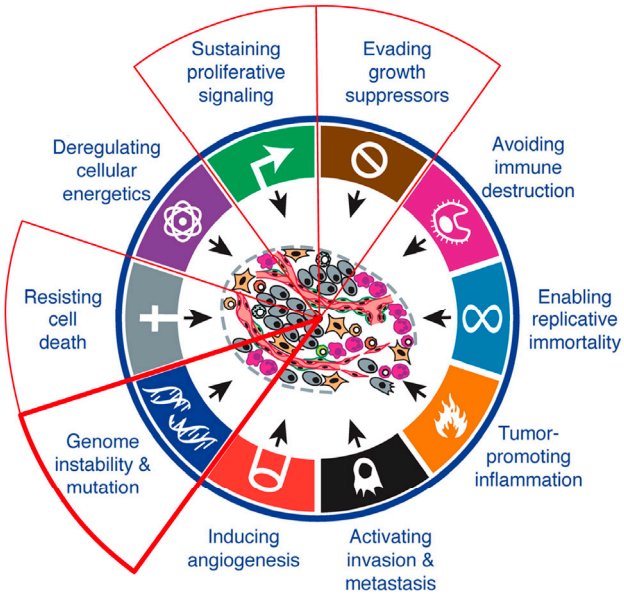


Figure 1: The hallmarks of cancer. The hallmarks highlighted in red are of particular importance in this work. Adapted from Hanahan and Weinberg (2011) [6] with permission from the copyright holder.

Mutational processes

> "The defects in genome maintenance and repair are selectively advantageous and therefore instrumental for tumor progression, if only because they accelerate the rate at which evolving premalignant cells can accumulate favorable genotypes."
>
> Douglas Hanahan and Robert A. Weinberg
> in "Hallmarks of Cancer" (2011) [6]

Germline and somatic mutations

Mutations and genomic alterations that are studied in cancer are either present from conception (germline mutation) or have been acquired during the life of the organism and are only found in a subset of cells (somatic mutation). Germline mutations are studied as the mechanism for hereditary cancer, for example *BRCA1* and *BRCA2* mutations, which confer an estimated 20-40% lifetime risk of developing ovarian cancer and 40-85% risk of breast cancer [7]. Another example is Lynch syndrome, which is caused by mutation in one of the mismatch repair

(MMR) genes or in the *EPCAM* gene and greatly increases the risk of several cancer types, first and foremost colorectal and endometrial cancer [7]. However, overall it is estimated that only 5-10% of all cancer is caused by strong single-gene germline mutations.

Somatic mutations accumulate and propagate during the lifetime of multicellular organisms, and a small subset of these will confer tumorigenic properties to the cell. At diagnosis, a tumor typically contains 2-8 "driver" alterations [8]; the remaining alterations providing no apparent advantage to the tumor cells are often referred to as "passenger" mutations. Solid tumors have varying amounts of somatic mutations: lung and melanoma tumors have on average around 150 non-synonymous mutations whereas pediatric cancers only harbor around a tenth of that number [8]. Two principally different processes are at play in the generation of mutations: 1) replicative errors and 2) environmental mutagens [9, 10].

DNA replication error

Considering the vastness of the human genome – over 3 billion base pairs (bp) in the haploid genome – it is almost surprising that it remains so stable throughout generations and throughout the life of the individual person. The genome is safeguarded by several important enzymatic and supportive protein systems, ensuring correct DNA replication and repair of any damage before it can give rise to persisting mutations.

Replication error accounts for part of the "bad luck" factor and is a major contributor to cancer development. The lifetime risk for many cancer types has been reported to be directly proportional to the number of stem cell divisions in the tissue of origin [9]. Some mutations are introduced during each cycle of replication due to infidelity of the cell's replication machinery for example through mismatch base-pairing [11]. DNA polymerase proof-reading is therefore crucial in maintaining the integrity of the genome, and deficiencies in this process are known to give rise to tumors [12, 13].

Environmental and intrinsic mutagens

Environmental mutagens include toxins and ionizing radiation such as ultraviolet (UV) light. Often these agents result in distinct mutational signatures due to their individual chemical reactivity profiles; for example UV exposure creates pyrimidine dimers that are prone to give rise to C>T and CC>TT mutations [14].

Importantly, mutagens can also be intrinsic to the human body. For example, in the metabolic processes, reactive oxygen species are generated, which have the potential to damage DNA [15]. Additionally, DNA molecules *in vivo* exist in aqueous state, which renders them vulnerable to hydrolysis of the glycosylic bond and release of a free base, creating an apurinic/apyramidinic site [16].

DNA repair systems

In a healthy cell, acquired genetic damage can often be corrected by DNA repair processes and enzymes. Disruption or dysregulation of these enzymes, as is often the case in cancer, will increase the proportion of genomic damage that is propagated. Germline or acquired mutations in genes critical to maintaining DNA stability and/or repairing errors can drastically increase the chance for DNA replication errors in cells and tissues, and thus are generally associated with increased risk for cancer.

When only one DNA strand is damaged, the other strand still holds the information needed to restore the missing or ambiguous bases. This can be done with base excision repair (BER), nucleotide excision repair (NER) or mismatch repair (MMR). All three rely on endonucleases and involve breaking the bonds of the DNA backbone and removing either just one nucleotide (BER) or a larger section of the damaged strand (NER, MMR) followed by resynthesis with the other strand as a template and re-ligation with ligases. Some base modifications can be resolved by direct enzymatic modification, without breaking the phosphodiester bonds [17]. The MMR system monitors newly synthesized DNA for incorrectly copied DNA sequences and is particularly important in highly repetitive regions. Together, the DNA proof-reading and MMR systems account for the astoundingly low replication error rate of 1 in a billion nucleotides [17], equivalent to about 6 errors per cell replication.

Double-stranded breaks of the DNA backbone are especially harmful to the cell as they may lead to genomic rearrangements wherein stretches of DNA fuse to the wrong site or even the wrong chromosome. Optimally, such a break may be resolved by the homologous recombination repair system (HRR), wherein an identical or nearly identical DNA molecule is used as a template to restore the DNA integrity [18]. Non-homologous end-joining (NHEJ) repair constitutes an alternative and more error-prone pathway [19].

Genomic changes in cancer

Tumor cells have dozens to hundreds or more coding mutations at the time of diagnosis, but the total number of alterations, coding and silent, can be much higher than that. The initiation of a tumor in many tissues may be attributed to mutations that arise stochastically during normal stem cell divisions [9, 10], but once a tumor has developed, it soon harbors too many genetic alterations to be readily accounted for by normal tissue dynamics. As an explanation, the concept of a mutator phenotype was introduced more than forty years ago based on the observation that once a neoplasm has developed, the rate of genetic change increases [20]. Although mutations and chromosomal damage are potentially lethal to a cell, they are also mechanisms through which cells can acquire a

proliferative advantage; or resistance, both to the human body's own anti-tumor mechanisms, and to cancer therapies [6].

Single nucleotide variants and indels

Genomic changes are commonly classified according to size and type as well as other characteristics. Single nucleotide variants (SNV) are the most commonly occurring alterations and are subdivided into nucleotide transitions and transversions. Transversion refers to the mutational process typically driven by ionizing radiation and alkylating agents, when a purine (A and G) is substituted by a pyrimidine (C and T) or *vice versa* (C>A, C>G, T>A, and T>G), whereas transitions are the conversion of purine to purine (G>A, A>G) or pyrimidine to pyrimidine (C>T, T>C), often driven by oxidative deamination and tautomerization. A single nucleotide change can be part of normal human variation and is, if present in at least 1% of the population, referred to as a single nucleotide polymorphism (SNP). A "non-synonymous" SNV by definition leads to an altered amino acid, a lost stop codon, or a premature stop codon, each resulting in an abnormal mRNA and protein. SNVs in coding sequence may also result in an unaltered amino acid sequence and are then referred to as "synonymous". Synonymous variants are often disregarded at the analytical stage, but they may have unanticipated effects, such as exon-skipping [21] and it has recently been estimated that 6-8% of cancer driver mutations are synonymous variants [22].

Indels (insertions and deletions) refer to changes, that either add or remove bases, typically less than 50 nucleotides (small indels) but sometimes much larger. Indels of a base pair length divisible by three are "in-frame", and do not lead to a truncated protein, since the down-stream codons are still intact. Frame-shift alterations disturb the reading frame and typically lead to a premature stop codon and a truncated protein and/or translation-coupled degradation of the mRNA through the nonsense-mediated decay pathway [23].

The mechanisms for acquiring genomic changes in cancer are similar to those at work in all somatic cells, with the crucial difference that one or several protective systems may be dysfunctional. Some pathways are well-described in tumor mutagenesis and can be tracked through their distinct mutational profiles.

Enzymatic deamination of cytidine to uracil is a key part of the innate and adaptive immune system, providing a mechanism for hypermutation and chromosomal rearrangements to allow greater antibody diversity. Dysregulation of the activation-induced deaminase (AID) and APOBEC family deaminases are also a source for hypermutation in human cancers [24, 25]. APOBEC-deregulation is manifested as a substantial over-representation of C to T and C to G substitutions, primarily within TCA and TCT trinucleotides [24]. The APOBEC mutational pattern has also been identified in other cancers, including bladder, cervical, head and neck, and lung cancers [26]. Another example of a mutational signature is the

pattern of G:C to T:A transversions present in MUTYH-associated polyposis, a hereditary syndrome with high risk for developing colorectal cancers. This signature arises due to a defective BER DNA glycosylase (MUTYH) [27].

Structural variants

Larger genomic changes are referred to as structural variants (SV), which includes rearrangements of the chromosomes and copy number variations (CNV). Interchromosomal rearrangements involve movement of genetic material between chromosomes, whereas intrachromosomal rearrangements are limited to a single chromosome (Figure 2). When a segment of DNA is missing, resulting in a lower copy number state, this is referred to as a deletion. Similarly, a large genomic segment may be duplicated one or several times, resulting in increased copy number, referred to as a duplication or amplification. Such copy number changes can lead to either decreased or increased amounts of gene product, which may have profound effects in a cell, and may also physically disrupt the coding sequence of genes, thereby inactivating normal function. Inversion refers to when a segment of DNA reverses orientation so that the sequence is oriented "backwards", which alters the coding sequence and can likewise lead to loss or disruption of genes [28]. Another possible consequence of SV is the generation of fusion genes. The archetypal example of this is the hybrid Philadelphia chromosome in chronic myeloid leukemia resulting in the fusion of *BCR* and *ABL* genes into *BCR-ABL* which gives rise to a constitutively active tyrosine kinase with the ability to drive cell division [29-31].

Cancer cells are also prone to develop aneuploidy and highly complex genome copy states. The human karyotype, normally consisting of 23 chromosome pairs, may acquire extra chromosomes (e.g., trisomy) or lost chromosomes. Whole-genome duplication is not uncommon in cancer and results in tetraploid cells [32], a phenomenon some report occur in 37% of human cancers [33].

Clustered rearrangements in cancer

> "Whatever the mechanism of damage, the consequences are profound. Faced with hundreds of DNA breaks, the cell's DNA repair machinery attempts to rescue the genome. The resultant hodgepodge bears little resemblance to its original structure, and the genomic disruption has wholesale and potentially oncogenic effects."
>
> Philip J Stephens and colleagues (2011) [34]

From cytogenetic studies of cancers, it has long been known that chromosomal rearrangements occur and may constitute drivers of cancer growth. With the

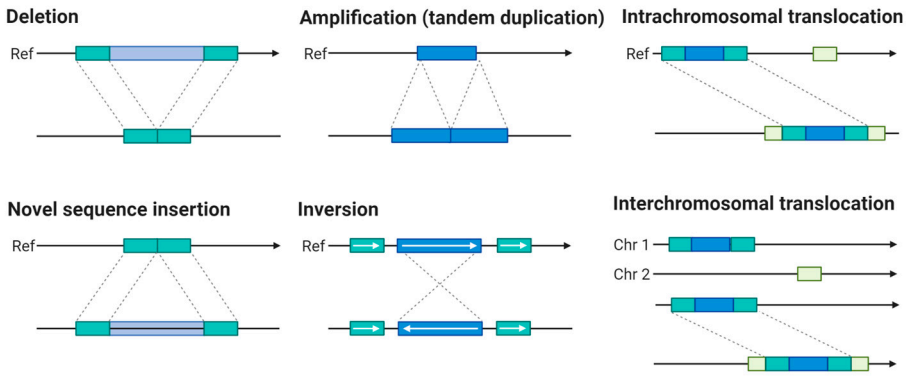


Figure 2: Types of chromosomal rearrangements. Chr: chromosome; Ref: reference chromosome. Created with BioRender.com.

advances of genetic methodologies and bioinformatics, analyses of structural in cancer have revealed diverse patterns in the type and location of rearrangements and copy number variations that may give clues to how they arose. Even before Watson and Crick's discovery of the DNA double-helix, Barbara McClintock's work on maize genetics in the 1930's and -40's described the phenomenon of repeated chromosome breakage and fusion through cycles of mitosis that would result in complex, clustered rearrangements of chromosomes [35]. Fusion of chromosomes may not immediately lead to a problem, but in the anaphase of mitosis, when chromosomes must segregate into separate daughter cells, there may be strands of chromatin still connecting them, which renders the genome vulnerable to fracturing. In cancer, cytogenetic studies have revealed significant telomere attrition and instability, which can lead to such breakage-fusion-bridge cycles with resulting chromosomal inconsistencies [36, 37].

In 2011, Stephens and colleagues coined the expression "chromothripsis" to describe the occurrence of tens to hundreds of rearrangements clustered within a limited space, from a few megabases up to the length of a chromosome arm [34]. Multiple plausible hypotheses exist to explain chromothripsis. A cell that has initiated apoptosis will start to digest its genome, resulting in chromosome fragmentation and if this process is interrupted, the restoration of the damaged regions is likely to be highly error-prone. Another proposed cause is that a burst of ionizing radiation during mitosis likewise could explain the juxtapositioning of the breakpoints [34, 38].

Analyses of breakpoint sequences in chromothripsis regions indicate that fusion of the chromosomes is mainly mediated by NHEJ repair, and, to a lesser extent, alternative end-joining repair and microhomology-mediated break-induced replication [39, 40]. Initially, using SNP chip microarray data, chromothripsis was

estimated to be present in 2-3% of cancers [34], but a recent analysis of whole genome sequencing data from a pan-cancer collection of tumors showed it was found in 22.3% of tumors, and some tumor types, including breast cancer, had a prevalence of more than 50% [40, 41].

Other patterns of clustered SVs have also been described. Chromoanasynthesis refers to the process of rearrangement through replication error involving template switching and microhomology-mediated break-induced replication [42]. A third term, chromoplexy, was invented for yet another type of rearrangement pattern found in prostate cancer, characterized by large chains of rearrangements that affect multiple chromosomes in a coordinated fashion [43]. Chromothripsis, chromoanasynthesis and chromoplexy are grouped under the umbrella concept of chromoanagenesis [44].

Tumor evolution and heterogeneity

> "As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it vary however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of surviving, and thus be naturally selected. From the strong principle of inheritance, any selected variety will tend to propagate its new and modified form."
>
> Charles Darwin in *On the Origin of Species* (1859)

As our understanding of tumor biology has grown, many tumor diseases have been divided into subtypes according to their molecular and pathological characteristics. Even though subclassification of tumors is highly useful for determining prognosis and treatment, it is only an approximation of reality since each cancer patient will have a genetically and phenotypically unique tumor disease. We now know that even within a single tumor, there can be considerable intratumoral clonal diversity (Figure 3).

It is commonly stated that the approximate limit for clinical detection is a tumor size of 1 cm^3, roughly equivalent to one gram, and contains 10^9 tumor cells. Although cell size and density will vary, it is estimated that a tumor at diagnosis contains millions to billions of cells [45]. It is likely that all these cells originated from one or a few common precursor cells, accumulating genetic and phenotypic changes throughout generations of cell divisions.

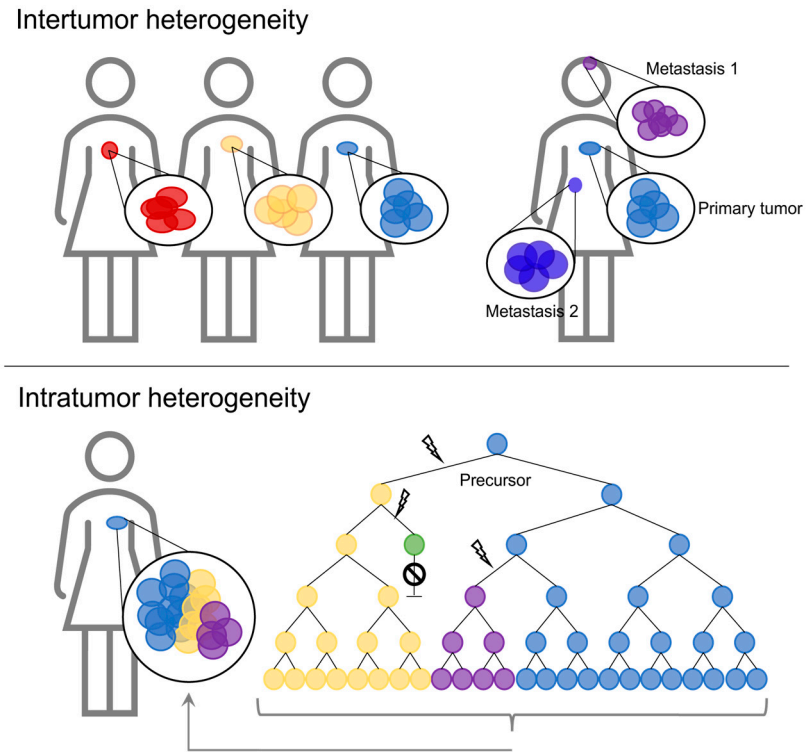


Figure 3: Schematic illustration of the concepts of inter- and intratumor heterogeneity. Colored circles represent tumor cells. Different colors are used to illustrate genotype and phenotype differences between (top) and within (bottom) tumors. Lightning symbol: mutation event; stop sign: unable to propagate.

Genomic instability of tumor cells increases the likelihood of acquiring new genetic changes which propagate throughout generations [46]. It is possible that a small subpopulation of tumor cells with stem cell-like properties are responsible for upholding the population of cancer cells [47]. The concept of cancer stem cells was first introduced in leukemia [48]. Since then, solid tumors have also been found to hold cells with stem-like properties, including gastric cancer [49], melanoma [50], colorectal cancer [51], prostate cancer [52], and others. In breast cancer, for example, a $CD44^{+}CD24^{-}$ subset of tumor cells have increased ability to initiate new tumors [53]. Additionally, the claudin-low subtype in breast cancer exhibits properties reminiscent of the stem-like phenotype [54, 55]. The implication for tumor heterogeneity is that the cancer stem cells make up only a small portion of the bulk of the tumor but act as precursors and may regenerate or repopulate the mass if they remain and survive surgery and cytotoxic treatments. The stem cell model is sometimes pitted against the classical model of tumor clonality, but elements of both mechanisms may be at work depending on the tumor type.

Metastatic processes

> "When a plant goes to seed, its seeds are carried in all directions; but they can only live and grow if they fall on congenial soil."
>
> "Then as regards "metastasis." Here, too, we shall find evidences of predisposition; we shall see that one remote organ is more prone to be the seat of secondary growth than another."
>
> Stephen Paget (1889) [56]

Inflammation, destruction of surrounding tissues and mere physical bulk associated with a primary tumor will potentially cause disruption of the function of the organ in which it grows. Tumors may also deprive the host organism of nutrients and produce signal substances that disturb homeostasis. Ultimately however, only 10% of cancer deaths are attributed to a primary tumor – the remaining 90% are caused by metastatic cancer [57].

The dynamics of tumor evolution and metastatic spread have been modelled to account for the clonal properties of tumors. The linear model hypothesizes that metastatic spread is a late event and that metastatic cells are most closely related to the most advanced of the primary tumor clones. The parallel model holds the view that metastasis may begin very early during tumor evolution and that the primary and metastasis tumors then continue to evolve in parallel, possibly with continuous seeding or re-seeding of daughter tumors [58] (Figure 4).

Four hallmarks of metastasis have been proposed [59]. These include 1) Motility and invasion, 2) Modulation of microenvironment, 3) Plasticity and 4) Colonization. The first barrier a cancer cell must overcome to achieve migration to a distant locale is usually the basement membrane. The next step is to gain access to anatomical spaces that allow for traveling throughout the body. In some compartments, there is room for extensive growth and spread, such as the abdominal cavity, but tumor cells can also be transported using blood vessels or lymphatic vessels. Intravasation is the process through which cancer cells gain access to the circulation, resulting in tumor cells circulating through the blood stream. For colonization from blood vessels, tumor cells must first adhere to the endothelial lining and extravasate to the tissues. Then, if the microenvironment of the new tissue is permissive, the tumor cells may form a colony [59].

Tumors that preferentially spread through the circulatory system often seed metastases at the first capillary bed that they encounter – for example the lungs (via pulmonary circulation) or the liver (via portal circulation). Lymphatogenic spread seeds metastases to nearby draining lymph nodes and therefore, analysis of these tissues is often integral to cancer diagnostics. The lymphatic system drains

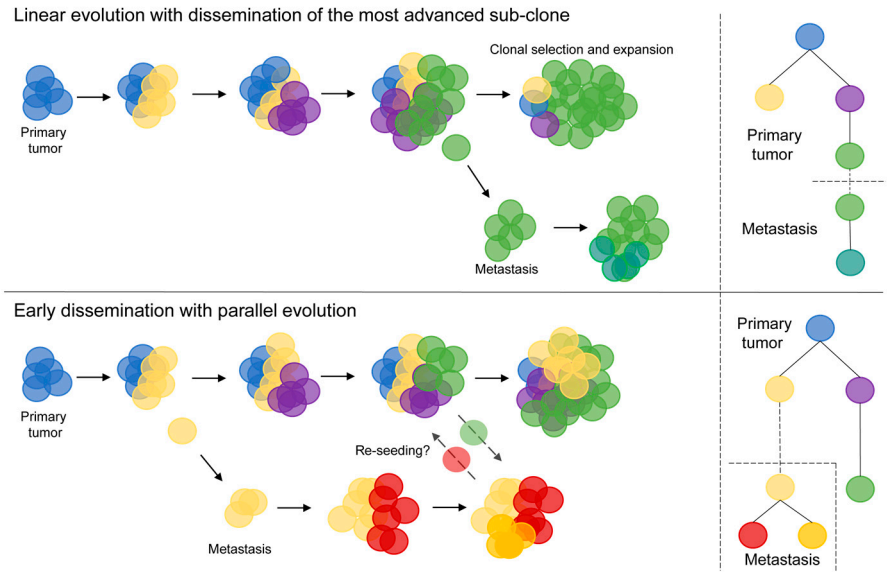


Figure 4: Schematic illustration of proposed models for tumor evolution. Right hand panels show proposed evolutionary path from precursor phenotype (blue). Different colors are used to illustrate the diverse tumor cell clones and offspring.

into the venous system, and from there the tumor cells can spread hematogenically, with potential access to the entire body [60].

In 1889, Stephen Paget coined the "seed and soil" theory, based on his post-mortem examination of cancer patients. He observed that both the properties of the tumor cells (seed) and the metastasis locale (soil) dictate the conditions for the metastatic seeding process. In other words, tumor cells will only spread to tissues wherein the microenvironment is permissive to their growth, and this varies between tumor types [56]. Carcinomas, especially of the breast and prostate, preferentially spread to bone – one study reports that around 70% of patients that have died from prostate or breast cancer have skeletal metastases at autopsy [61]—whereas melanoma metastases are more common in the lungs, liver and brain [62].

Studying inter-tumor and intra-tumor heterogeneity

Most biochemical methods require many cells for analysis and for clinical biopsy, efforts are made to retrieve as representative a sample as possible. But as we have seen, a growing, developing tumor potentially harbors many diverse subpopulations and therefore it is of interest to examine the general structure and clonality of tumor cells within a cancer and determine the likelihood of capturing most subclones in one biopsy.

Tumor evolution and heterogeneity can be studied through a variety of ways. One method is multiregion sampling wherein multiple samples are extracted from the

same tumor, hoping to capture subpopulations of tumor cells. This has been successfully achieved in several solid malignancies including breast cancer [63, 64], kidney cancer [65], pancreatic cancer [66], lymphoma [67], lung cancer [68], melanoma [69] and prostate cancer [70], revealing varying degrees of intratumor heterogeneity. The result of such studies depends greatly on the genetic markers used and the interpretation of the data, but in general, it is acknowledged that many solid tumors display at least some degree of genetic heterogeneity, with indications that larger tumors are more heterogeneous [63]. With advances in technology, tumor heterogeneity can be studied at higher and higher resolution, allowing for characterization of single tumor cells [71, 72].

Clonal relationship between primary tumor and metastasis

With the background of tumor evolution and metastatic properties, a distant metastasis may be very different from the primary tumor that seeded it. This population of cells has undergone the tests of invasion, spread and colonization and the disease at this point is often regarded as incurable. Changes that are uniquely present in metastases potentially have a role in tumor aggressiveness and spread so there have been numerous efforts to characterize metastases and compare them to the matched primary tumors.

One approach is to track coding mutations in known cancer driver genes, which has been done for lung cancer, colorectal cancer, breast cancer and others. Limiting the analysis to known driver genes in cancer, the mutational profiles in primary tumors and metastases is more than two thirds concordant and generally there is a net gain of driver mutations throughout the evolution of the tumor [73-80]. Whole-genome and exome sequencing studies include both driver and non-driver mutations and detect a larger heterogeneity between primary tumors and metastases, but it seems that approximately 50-60% of mutations can still be expected to be shared, although the concordance percentage varies widely between patients (range 6% to 95%) [81-85].

Even when the number of novel mutations in metastases is low, they can potentially represent a mechanism for relapse after the selection pressure of adjuvant treatment. In breast cancer, mutations in the *ESR1* gene, encoding estrogen receptor α, have recently been found to occur almost exclusively in advanced breast cancer, after treatment with anti-estrogen compounds [86, 87]. It is debated whether treatment resistance variants may actually be present already in the primary tumor, but at sub-detection levels. Mathematical modeling of tumor mutations in colorectal cancers speaks in favor of this hypothesis [88, 89]. If resistance mutations precede pharmaceutical therapy, the treatment will cause a temporary disease regression while selecting for and allowing expansion of the clones that are resistant. For tumor diseases where relapse follows a very distinct pattern of progression within just a few months, such as BRAF inhibitor-treated melanoma [90, 91] and anti-EGFR-treated colorectal cancer [92] this scenario is particularly plausible.

Breast Cancer

> "The conclusion I draw from [these cases] is this, that we must look in the female to the ovaries as the seat of the exciting cause of carcinoma, certainly of the mamma, in all probability of the female generative organs generally, and possibly of the rest of the body."
>
> George Beatson (1896) [93]

The human breast

Mammary gland development begins in the developing fetus and only completes after pregnancy, when hormonal influences initiate the process of lactation. The breast is composed of three types of tissue: glandular, adipose and connective tissue. The mammary glands have 15-20 lobes which are further subdivided into lobules. The functional unit of the breast is the terminal duct lobular unit which is composed of a cluster of alveoli with adjoining ductule (Figure 5). Epithelial cells line the ducts and alveoli and produce milk through a combination of merocrine and apocrine secretion and myoepithelial cells contribute to the ejection of milk through the converging ducts that terminate in the nipple [94]. Breast tissue undergoes repeated remodeling during the fertile years, precipitated by puberty, menstrual cycles, pregnancy and lactation. In conjunction with menopause, the mammary gland normally undergoes involution, where the terminal duct lobular units senesce and reduce in size. Incomplete involution is one of the biological processes associated with breast cancer development [95].

Female sex hormones

The steroid hormones, estrogens and progesterone, orchestrate the processes of female secondary sex characteristics and reproduction. Four estrogen hormones are found in the human body, estrone, estradiol, estriol and estetrol, with estradiol (E2) regarded as the dominant one in fertile years. Biosynthesis of estrogens in women of reproductive age mainly takes place in the ovaries, but there is significant contribution to estrogen production and metabolism by other tissues, such as adipose tissue, osteoblasts, chondrocytes and vascular endothelium [96].

In the human breast, estrogen is responsible for breast duct development and instrumental in inducing prolactin secretion by the pituitary. Progesterone is required for alveolar growth and development, but estrogen is also indirectly involved, by inducing transcription of the progesterone receptor (PR; gene symbol *PGR*), enabling progesterone signaling [97].

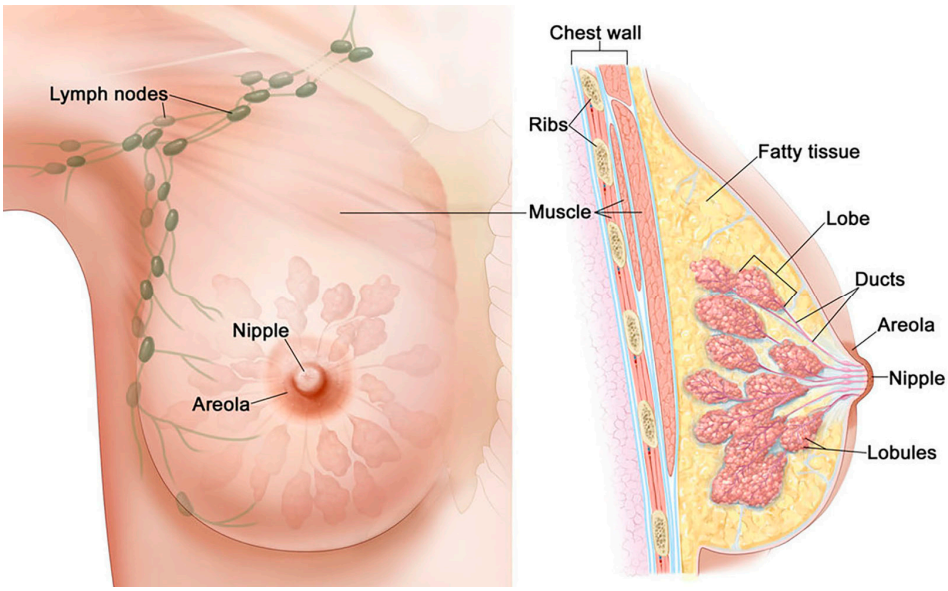


Figure 5: The anatomy of the female breast. © 2011 Terese Winslow LLC. US Govt has certain rights. Reprinted and adapted with permission from the copyright holder.

Breast cancer epidemiology and etiology

Cancer is the second leading cause of death worldwide [4]. More than 19 million people were diagnosed with a tumor disease in 2019, and almost 10 million died because of cancer [4, 98]. Breast cancer represents a sizeable portion of these numbers, comprising 11.7% of new diagnoses and 6.9% of deaths [98]. In Sweden during 2019, 8288 women received a diagnosis and 1353 died from the disease [99]. The incidence for breast cancer has increased steadily over the last few decades [2], and simultaneously, the prognosis for breast cancer has improved drastically, with a 40% reduction in mortality rate since 1989 [100]. The 5-year survival rate is now estimated at 91% and the 10-year survival rate at 84%, making breast cancer one of the invasive tumor diseases with the best prognosis.

A wide variety of tumorigenic and metastatic processes can initiate tumor formation in the breast. The World Health Organization classification of “breast tumors” includes fibroepithelial tumors, hamartomas, mesenchymal tumors, lymphomas and metastases to the breast; but breast carcinoma (breast cancer), as the name implies, develops from epithelial cells of the breast. Most forms of breast cancer arise from epithelial cells in the terminal duct lobular unit [101].

It is estimated that familial genetic factors account for 5-10% of breast cancer cases [102], but even in patients with a relevant family history, a predisposing gene is found in less than 30% of cases. Pathogenic mutations in *BRCA1* or

BRCA2, are autosomal-dominant and confer a greatly increased risk of developing breast cancer before the age of 70 (around 60% for *BRCA1* and 50% for *BRCA2*) [103-105]. Inactivating mutations of the tumor suppressor gene *PTEN* are rare but associated with an 85% lifetime risk of breast cancer. Carcinogenic variants of *TP53*, *CDH1,* and *STK11* are about half as penetrant, but also associated with development of other cancers [106].

Even though a large number of predisposing genes have been identified, most breast cancer cases are so-called "sporadic" and arise stochastically without any obvious familial or hereditary component. As with many other cancer types, the risk of contracting breast cancer increases with age. High body mass index [107, 108] and diabetes [109] are known predisposing factors, along with modifiable lifestyle aspects such as tobacco use [110] and alcohol consumption [111]. Breast cancer arises in hormone responsive tissue, and exposure to female sex hormones is also an important epidemiological risk factor. This includes both exogenous exposure such as postmenopausal estrogen substitution, and endogenous exposure such as early menarche and late menopause, nulliparity and higher age at first birth [102, 112]. High levels of endogenous circulating steroid sex hormones in postmenopausal women have also been linked to the development of breast cancer [113, 114].

Estrogen receptors in breast cancer

> "If cancer cells have receptors for these hormones, it's not necessarily a bad thing. It means that they are at least TRYING to perform the tasks of a normal breast cell. They're behaving — somewhat."
>
> Marisa Weiss, Chief Medical Officer, Breastcancer.org (2020)

Estrogen receptors

ERα was first discovered by Elwood Jensen in the late 1950's. While studying the estrogen-retaining properties of uterine tissues, he isolated the receptor as a protein [115] that mainly resided in the cell cytoplasm of target tissues, and, upon binding of estrogens, translocated to the nucleus [116]. It is now known that ERα is encoded by the *ESR1* gene on chromosome 6 and spans 595 amino acids [117].

Decades after the discovery of ERα, a different team of scientists isolated a protein from rat prostate tissue that was highly homologous to the estrogen receptor and was dubbed ERβ [118]. It is encoded by the *ESR2* gene on chromosome 14, which is transcribed and translated to a 530 amino acid long protein, containing similar functional domains of ERα, with varying degrees of homology (Figure 6) [119, 120].

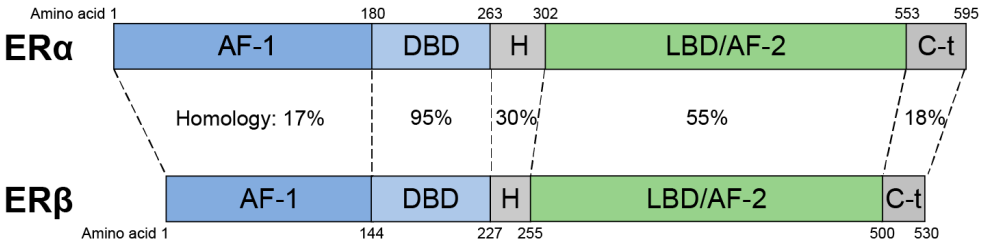


Figure 6: Protein domains and homology of full-length estrogen receptors α and β. AF-1: Activation function 1; DBD: DNA-binding domain; H: hinge domain; LBD: Ligand-binding domain; AF-2: Activation function 2; C-t: C-terminal domain.

Through alternative promoter usage and alternative splicing, dozens of transcript variants can be produced from the ER genes. Many of the reported transcripts have not been well characterized, but several are known to vary in their occurrence across tissue types and result in increased or decreased receptor activity and co-factor affinity [121].

Although somewhat outside the scope of this thesis, it should be mentioned that there is a third estrogen receptor, the G protein coupled estrogen receptor (GPER, formerly GPR30) which is structurally and functionally distinct from ERα and β [122-124]. Upon ligand-binding, it enacts a rapid cellular response through the production of second messengers. It is reported to be expressed in around 60% of invasive breast tumors [125, 126], and to correlate to poor prognosis [127], and tamoxifen resistance [128, 129].

Estrogen receptor signaling mechanisms

ERα and β are members of the nuclear receptor superfamily, subgroup 3 – the steroid receptors. [130]. Like most other nuclear steroid receptors, they are composed of four functionally distinct domains (Figure 6). The N-terminal activation function 1 (AF-1) domain permits ligand-independent activation through phosphorylation and interaction with transcriptional co-factors. Downstream of this region is the DNA-binding region, which contains a zinc finger motif and allows binding of the receptor to estrogen response elements (ERE) on DNA [131]. The hinge region contains a nuclear localization signal [118] and also composes a flexible connection to the C-terminal part of the protein, the ligand-binding domain (LBD). The LBD also includes a second activation function section (AF-2), responsible for conformational change in response to ligands as well as association to transcriptional co-factors [131]. A comparison between ERα and β reveals that the DNA-binding domain is 95% homologous, the ligand-binding domain 53% and the co-factor association domain AF-1, only 17% homologous (Figure 6) [118], indicating that the two receptors may have similar affinity for binding sites on DNA but diverge in their ligand- and co-factor-binding profiles.

When ERs are activated, they translocate to the nucleus and undergo a conformational change, enabling them to form dimeric complexes. ER dimers exert their effects through interaction with the genome and associated co-factors, resulting in transcription of target genes. ERs form both homodimers (α/α, β/β) and heterodimers (α/β) [132, 133]. ERα and ERβ compete for genomic binding sites, and while they share a substantial number of transcriptional targets, ERα has a higher affinity for EREs, and co-expression of both receptors results in displacement of ERβ to more ERE-poor regions of the genome [134].

ERs can interact directly with DNA containing ERE palindromic sequences (the classical/direct pathway) or indirectly, through other transcriptional regulators such as SP1 [135] and the transcription factor complex Activator Protein 1 (AP-1) [136, 137], called the tethered/indirect pathway. A third mechanism of ER activation is through phosphorylation, independent of ligand binding. ER activity is regulated through phosphorylation via a diverse set of pathways, including the epidermal growth factor receptor family and the insulin-like growth factor receptor with associated downstream effectors (such as phosphoinositide 3-kinase (PI3K)/AKT and RAS) [138-144], as well as chemokine receptor pathways [145, 146], estrogen signaling pathways (ERα) [147] and cyclin-dependent kinase pathways (ERα) [148, 149] (Figure 7).

Biological impact of estrogen receptors

The estrogen-responsive transcriptome has been extensively studied and reviewed [150-155]. Methods for identifying target genes include computational sequence analysis for estrogen-responsive DNA motifs, gene expression analysis, chromatin immunoprecipitation with DNA analysis and protein detection assays. Estrogen receptors are involved in both upregulation and downregulation of hundreds to thousands of genes. Important targets include transcription factors and co-factors such as *PGR*, *FOS*, *GREB1*, *RARA* and *MYC*, as well as genes involved in growth hormone signaling such as *VEGFA*, *IGF1*, and *EGFR*.

As we have seen, ERα and β have a high degree of homology in their DNA-binding domain (95%), but while they share some transcriptional targets, they have opposite effect on others. This can be exemplified by ER regulation of some of the cell cycle proteins. Whereas ERα is known to stimulate cell cycle progression through transcriptional upregulation of cyclin D1, cyclin A and MYC [156-158], introduction of ERβ reduces the transcription of these genes and causes a G2 arrest [159].

One of the mechanisms through which ERβ opposes the action of ERα is through regulation of transcription sites dependent on the AP-1 transcriptional complex. ERβ affects this process on multiple levels, altering recruitment of AP-1 factors FOS and JUN to estrogen-responsive promoters, and reducing production of FOS

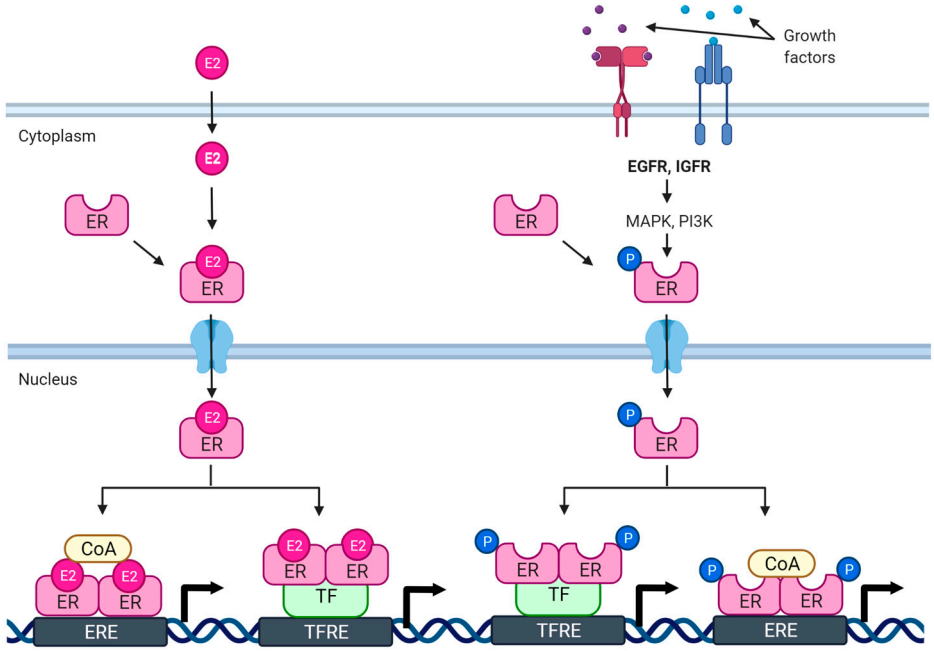


Figure 7: Estrogen receptor (ER) mechanisms for activation of transcription. Upon activation by ligand or through phosphorylation, ERs translocate to the nucleus, dimerize and bind response elements either directly or via transcriptional co-factor complexes. ER: Estrogen receptor; EGFR: epidermal growth factor receptor; IGFR: insulin-like growth factor receptor; E2: estradiol; CoA: Co-activator; TF: Transcription factor; P: phosphoryl group; ERE: estrogen response element; TFRE: transcription factor response element. Created with BioRender.com.

protein, resulting in decreased transcription of some ERα targets, including *PGR* and *TFF1* [160].

The ERα/β concerted effect on cell transcription has been described by Williams *et al*. In a transcriptome-wide study of ERα and β-responsive genes, performed using a cell line with inducible ERβ expression they showed that, while estrogenic activation of ERα alone caused differential expression of 1432 genes, introduction of ERβ abrogated this effect for 998 genes (70%). ERβ also affected the expression of an additional 152 genes not regulated by ERα. The biological processes induced by ERα included mitotic cell cycle, cell proliferation and cell cycle check point, whereas ERβ induced genes involved in negative regulation of cell cycle, energy pathways and apoptosis [161].

Expression patterns of estrogen receptors

The estrogen steroid receptors have a profound impact on cell fate, regulate a wide variety of downstream genes and affect cell survival and proliferation. ERα is required for breast development and drives proliferation in the normal mammary

gland, but ERα positive cells themselves do not seem to actively proliferate [162, 163]; instead, they may act through paracrine signaling mechanisms to induce cell division in neighboring cells [164]. ERβ is not necessary for mammary duct development, although ERβ knock-out mice have reduced ovarian function and require supplementation of progesterone to induce normal pubertal mammary growth [165]. ERα is mainly expressed in epithelial breast cells; around 10-25% of acinar and ductal epithelial cells located near the lumen are positive for this receptor in immunohistochemistry (IHC) assays [166-168], whereas ERβ is expressed both in luminal epithelial, myoepithelial and stromal cells [166, 168-170].

In breast cancer, ERα is expressed in approximately 75% of breast tumors [171, 172] and in contrast to normal breast tissue, it is present in a larger number of cells and co-expressed with markers for proliferation [162]. IHC detection of ERβ in breast cancer specimens has yielded conflicting results, with some studies reporting a high percentage of tumors expressing ERβ (56-76%) [173-177] and others a moderate to absent number (0-35%) [170, 178, 179]. Some authors have remarked that the lack of well validated antibodies and assays may be at the root of these conflicting results and that comparison with *ESR2* mRNA levels indicate that estimates in the lower range may be more accurate [170, 179].

Clinical relevance of estrogen receptors

ERα has the capability to initiate and drive tumor growth through its activation of a plethora of tumor promoting genes. The tactic of reducing estrogen signaling to combat breast cancer has been used since the 19th century, when ovariectomy was found to have a beneficial effect for patients with inoperable breast cancer [93]. Pharmacological endocrine therapies in clinical use can inhibit ER signaling through estrogen depletion, partial antagonism or targeted ERα degradation; treatment strategies targeting ERα is further discussed in a later section. As we shall see, ERα is a well-established biomarker in breast cancer diagnostics, and is, despite its tumor-promoting properties, associated with better prognosis. The terms “ER-positive” and “ER-negative” are frequently used to describe breast tumors and it should be noted that this generally refers to the expression of ERα only.

The role of ERβ in breast cancer is controversial. In vitro studies indicate that it has antitumorigenic properties, and it is proposed as a tumor suppressor in breast- and other cancers [180], and ERβ expression levels are reported to decline during breast tumor development [181-183]. Expression of ERβ in breast cancer has been shown in numerous studies to be linked to a favorable outcome, with some authors reporting the effect across all subtypes and treatments [184, 185], and some for ERα-positive disease only [186]. Association for ERβ with better outcome in tamoxifen-treated breast cancer [187-190] has also been reported, with some studies indicating that this is particularly pronounced in ERα-negative disease [187, 188]. However, in contradiction with these results, ERβ has also widely been

reported as a potential marker for increased risk for relapse and/or death – both in hormone-receptor negative [191, 192] and hormone receptor-positive disease [193, 194], as well as in a cohort unselected for receptor status and treatment [195].

Given the conflicting reports and our incomplete understanding of ERβ function in breast cancer, the exact role and potential clinical use of this receptor is yet to be determined. However, a meta-analysis, incorporating data from most of the aforementioned studies as well as several others, showed that high ERβ expression was associated with improved outcome for both disease-free and overall survival [196]. Should the hypothesis of ERβ as a tumor suppressor be confirmed, there are ERβ-selective agonists [197, 198] that could conceivably provide new treatment options in the future.

Genomic changes in ESR1

For some malignancies, genomic rearrangements resulting in gene fusions are of great importance and constitute druggable targets. Breast cancer does not to the same extent exhibit fusion genes, but some recurrent fusion genes have been described. For example, *ESR1* gene fusions can result in a protein lacking the ligand-binding domain and may lead to ligand-independent, constitutive activation of the receptor [199, 200]. RNA sequencing (RNA-seq) has revealed *ESR1* fusion transcripts that enable estrogen-independent growth as well as complete resistance to selective estrogen receptor modulators (SERMs) [201].

A large portion of breast cancer patients receive some form of endocrine therapy to suppress estrogen-driven tumor proliferation. Similar to other tumor drivers, ERα is subject to mutations that increase the activity of the receptor, even in the absence of agonists. Variants like this, including E380Q [202] and Y537N [203] were described already in the 1990's, but the clinical relevance only became apparent in 2013, when a series of studies on *ESR1* mutations in advanced breast cancer reported that *ESR1* mutations were common in metastatic breast cancers, especially among patients who had previously received endocrine therapy [86, 87, 200, 204]. In these studies, both previously known and novel mutations were described, most of them located in the ligand-binding domain of the estrogen receptor [86], leading to constitutive activation and downstream transcription. LBD mutations include at least 60 variants (reviewed in [205]), several of which have been experimentally verified to confer increased activity in the absence of estrogenic ligands [86, 87, 200, 204]. The more potent mutations, like Y537S, even confer some resistance to ER suppression by ER modulators and degraders, requiring higher doses to inhibit *in vitro* proliferation [206]. In patients, the major reported effect is resistance to aromatase inhibitors both as monotherapy [207] and in combination with the new PI3K-blocking compound alpelisib [208], although there are novel endocrine therapies in development that are effective even against *ESR1* mutant tumors [209, 210].

Biomarkers and molecular profiling in breast cancer

In addition to the estrogen receptors, there are several biologically and clinically relevant markers that affect tumor cell phenotype, patient prognosis and treatment options.

Progesterone receptors

PR A and B are nuclear steroid receptors encoded by the same gene (*PGR*) and are produced by start of transcription at one of two alternative promoters [211]. PR B contains an additional 164 amino acids at the N-terminal end [212] and has a higher capability of inducing expression of downstream target genes [213]. *PGR* is a classical transcriptional target of ERα [211] and presence of the receptor is considered a marker for ER activation. It is involved in extensive cross-talk with ER and although its exact functions in breast cancer are debated, it is routinely assessed in the clinic together with ERα and human epidermal growth factor receptor 2 (HER2), and is associated to a more favorable outcome [214-216].

HER2/ERBB2/neu

HER2, encoded by the gene *ERBB2*, was identified in the 1980's as an oncogene in neuroblastoma [217] and was later revealed to be a tyrosine kinase receptor and a member of the EGFR family [218, 219]. HER2 is located in the cellular membrane, where it responds to growth signals through dimerization and autophosphorylation, initiating a signaling cascade through several downstream effector systems [220]. In breast cancer, HER2 is genomically amplified and/or overexpressed in 15-20% of tumors, leading to excessive intracellular growth signaling and tumor proliferation, and its overexpression is associated with poor prognosis [221, 222], although the outcome for HER2-positive breast cancer has improved since the advent of targeted therapies against this receptor [223].

Intrinsic subtypes

In the early 2000's, with the use of microarray technology, pioneering work by Perou and Sørlie resulted in a novel classification of breast tumors. Using transcriptional profiles, they stratified a collection of breast tumors into groups that transcriptome-wise resembled either luminal or basal epithelial mammary cells, with a third group characteristic of the HER2-enriched subtype [172]. These gene expression profiles have since crystallized into five subtypes that are associated to clinical characteristics and outcomes: Luminal A, Luminal B, HER2-amplified, Basal-like, and Normal-like [221]. The intrinsic subtypes are closely linked to the receptor status of the tumors: Luminal A and B are mainly composed of ER-positive tumors; the HER2-enriched group is enriched for HER2-amplified tumors (ER-positive or -negative); and the Basal-like group is generally characterized as being ER-, PR- and HER2-negative (triple negative breast cancers; TNBC). The role and even the biological existence of the Normal-like

subtype, so named because its gene expression profile resembles that of normal breast tissue, is debated. It is not used in clinical diagnostics, and many argue that it may be an artifact from contamination by normal breast tissue [224, 225].

Breast tumors can now be classified into the intrinsic subtypes based on expression levels of a 50-gene panel, Prediction Analysis of Microarrays 50 (PAM50), that approximates the gene expression groups [225]. Although expression profiling is not usually part of the clinical routine, surrogate immunohistochemistry and morphological markers can approximate the expression subtypes. The concordance between the IHC- and PAM50-defined subtypes ranges from 45% for the HER2-enriched group to 86% for the Basal-like/TNBC group [226].

Mutations and genomic dysregulation

Breast cancer, like other malignancies, exhibits recurrent alterations in certain genes, particularly genes that are related to regulation of the cell cycle, cell survival and apoptosis.

TP53, dubbed "the guardian of the genome" [227], is the most frequently mutated gene across all cancer types [228] and *TP53* mutations occur in approximately 35% of breast tumors [228, 229]. The protein encoded by this gene responds to diverse stress factors, including DNA damage, and restricts inappropriate clonal expansion by controlling cell cycle progression, DNA repair and apoptotic mechanisms [230]; oncogenic mutations disable these tumor suppressive functions. Multiple strategies are being investigated to develop treatments targeting TP53 dysregulation in tumors, including achieving synthetic lethality by disrupting DNA repair pathways, inducing an immunogenic response against mutant TP53 and restoring proper protein conformational structure through small molecule compounds. For some anti-mutant TP53 therapies, clinical trials are underway [231].

PIK3CA encodes the catalytic subunit of PI3K, p110α, which catalyzes the production of second messengers with diverse downstream effects, including the activation of the AKT/mTOR (mammalian target of rapamycin) signaling pathway [232]. *PIK3CA* mutations are present in approximately 35% of breast tumors [228, 229, 233-235], and can lead to constitutive activation of the PI3K pathway, resulting in dysregulation of processes such as cell growth, proliferation, migration and differentiation. *PIK3CA* mutations are generally associated to better outcomes [236], but predict poor prognosis in the metastatic hormone receptor-positive subtypes, with poorer response to chemotherapy [237].

In normal tissues, the PI3K pathway is negatively regulated by the phosphatase PTEN, which catalyzes the opposite reaction to PI3K, reducing the levels of second messengers and counteracting the downstream effects. Although *PTEN* mutations are not particularly common in breast cancer, the PTEN phosphatase is

frequently lost through other processes such as genomic disruption and promoter hypermethylation [235, 238-240].

The most important markers for hereditary breast cancer, *BRCA1* and *BRCA2,* both encode proteins involved in double-stranded DNA break repair via the HRR repair system. Inactivation of *BRCA1/2* function causes genomic instability and chromosomal rearrangements. Characterization of rearrangement signatures has shown that mutant *BRCA1* is associated predominantly with smaller (<10 kb) tandem duplications and smaller (<100 kb) deletions, whereas mutant *BRCA2* is associated more exclusively with smaller deletions [233]. Via categorization of rearrangement and mutational signatures it is possible to identify tumors with a phenotype matching that of germline *BRCA* mutations – so-called "*BRCA*-ness". Although not subject to *BRCA1/2* inactivating mutations these may have acquired *BRCA* methylation or other HRR defects that render them susceptible to treatments that target this deficiency [241].

Metastatic breast cancer

When breast cancer spreads to one or more secondary locations it is usually considered incurable, but novel targeted and chemotherapeutic agents have contributed to extending the survival of patients with systemic disease. Breast cancer preferentially spreads to the bone, but lung and liver metastases are also common [242]. Depending on the molecular subtype, breast tumors have different affinities for different metastatic locales. The luminal subtypes show preference for bone, whereas the basal-like subtype more often metastasizes to lung and brain [243].

Although the prognosis of breast cancer has improved dramatically, tumor recurrence is still common, and late recurrence is more common than in most other cancer types, especially in ER-positive breast cancer. Risk for relapse in ER-negative breast cancer subtypes is greatest in the first five years and for ER-positive breast cancer the risk is greatest at 6-10 years [244]. One meta-analysis showed that distant relapse occurred in the 5-20 year period after endocrine therapy for 13-41% of patients with ER-positive disease, with the cumulative risk increasing within this timeframe depending on tumor stage and node status [245]. Thus, overall the 5-year survival from breast cancer is very high, however over a quarter of women will, in the longer term, suffer a relapse.

As late relapse is more common in breast malignancies with a relatively good prognosis, risk-stratification is of great interest since it could spare patients from unnecessary treatment. Multiple attempts have been made to define risk-factors for late relapse. Suggested markers include lymph node metastasis [246] and expression of proteins related to tumor dormancy [247].

Discordance of biomarkers in primary and metastatic tissues

One challenge in the treatment of metastatic breast cancer is that the primary and metastatic tumors, although descended from a common progenitor, are not always genetically and phenotypically similar. Studies indicate that the receptor status (ER, PR or HER2) changes from the primary tumor to the metastasis in more than 40% of patients with PR being the most frequently affected marker [248-252]. A meta-analysis concluded that ER was discordant in 23%, PR in 41%, and HER2 in 10% of cases with matched samples [253].

Considering the differences seen between primary tumor and metastasis biology, it is now recommended by the European Society for Medical Oncology (ESMO) to secure biopsies from relapse sites for evaluation of the receptor status whenever possible. But in cases with discordant results, it has still not been ascertained whether primary or metastasis receptor status should guide treatment, although clinical studies to answer this question are underway [254].

Contralateral breast cancer

Contralateral breast cancer (CBC) is defined as breast cancer occurring in both breasts, either synchronously or in sequence (metachronously). Principally, CBC is regarded as a new primary tumor, and diagnosed and treated as such. Risk factors for developing CBC include young age at diagnosis, family history of breast cancer, large tumor size and a high number of tumor-positive lymph nodes [255].

Breast cancer patients have a higher risk of developing a new tumor in the other breast, compared to previously healthy individuals. A meta-analysis study showed a CBC median annual incidence rate of 0.5% after diagnosis of a first primary tumor [256]. Some studies also report worse prognosis of CBC compared to other breast cancer patients [257, 258], particularly if the second tumor occurs within 4-5 years or is synchronous with the first [257, 259].

Clinical diagnostics of breast cancer

The most common clinical presentation of breast cancer is a lump in the breast [260]. In the last 10 years in Sweden, 62% of breast cancers have been diagnosed via mammographic screening [261]. Both Swedish guidelines and ESMO recommend a triple approach to diagnosing breast cancer. It consists of clinical examination of the breasts and regional lymph nodes, radiological imaging and pathological evaluation of a biopsy or cytological sample [260, 262].

Histology and pathology

Invasive breast tumors exhibit a variety of growth patterns from almost completely differentiated, glandular-like to completely solid or diffusely growing. The most

common morphology, previously referred to as “invasive ductal carcinoma” and lately renamed “invasive breast carcinoma of no special type”, comprises most breast cancers. Carcinomas with more than 90% of special type morphology are classified as “special type”, for example invasive lobular, mucinous or tubular carcinomas [101]. There is limited knowledge of the biological processes resulting in one special type or another. Correlations have been found between certain histology types and molecular characteristics, for example invasive lobular carcinoma is often classified to the luminal or HER2 positive subtypes [263]. However, there is still much work to be done to reconcile the molecular and morphologic phenotypes.

Tumor grading is performed according to the Nottingham histological grade (NHG) system. Points are scored according to tubule and gland formation, nuclear pleomorphism and mitoses, and the tumors are classified into one of three grades, where grade 1 is the most indolent and grade 3 is the most malignant. More extensive glandular formation is a sign of higher differentiation and lower grade, but nuclear pleomorphism – diversity in the shapes and sizes of the tumor cell nuclei – and a high mitotic count are a sign of malignancy and high grade [264].

Intrinsic subtyping through gene expression analysis is increasingly recommended, but still not routinely performed in the clinic. Presently in Sweden, gene expression profiling is recommended for a subset of luminal breast cancers with uncertain risk categorization to better determine which patients would benefit from chemotherapeutic agents [260]. However, in most cases, IHC markers and tumor grade are used for approximation of tumor subtype to help guiding the choice of treatment. The key biomarkers are ER, PR, and HER2 receptor statuses, tumor grade and the expression of the proliferation marker Ki67 (Figure 8) [260, 262]. The ER-positive subtypes, Luminal A and B, have a favorable prognosis and can be treated with hormonal therapy, but are more at risk for late or very late relapse. HER2-positive disease has a more aggressive biology but is sensitive to treatment with anti-HER2 targeted therapies (usually combined with chemotherapy and/or endocrine therapy), and generally has a favorable outcome. The most aggressive and difficult-to-treat tumors fall in the Basal-like and triple negative subgroups, which are mostly, but not completely overlapping. They lack expression of either treatment targets (ER/PR or HER2) and are often poorly differentiated with an aggressive tumor biology and a high risk for relapse in the short term.

Tumor staging for breast cancer takes into account the anatomical properties of the tumor and is classified in categories of T (tumor size and local invasion), N (lymph node involvement) and M (distant metastasis) [265]. Stage is denoted from 0-IV and subdivided further using A-C, for example Stage IA signifies a tumor smaller than 20 mm without lymph node involvement or metastasis, whereas Stage IV refers to any tumor with distant spread. Further, the pathological stage, likewise designated 0-IV, combines TNM stage with grade and biomarkers for improved prognostic prediction.

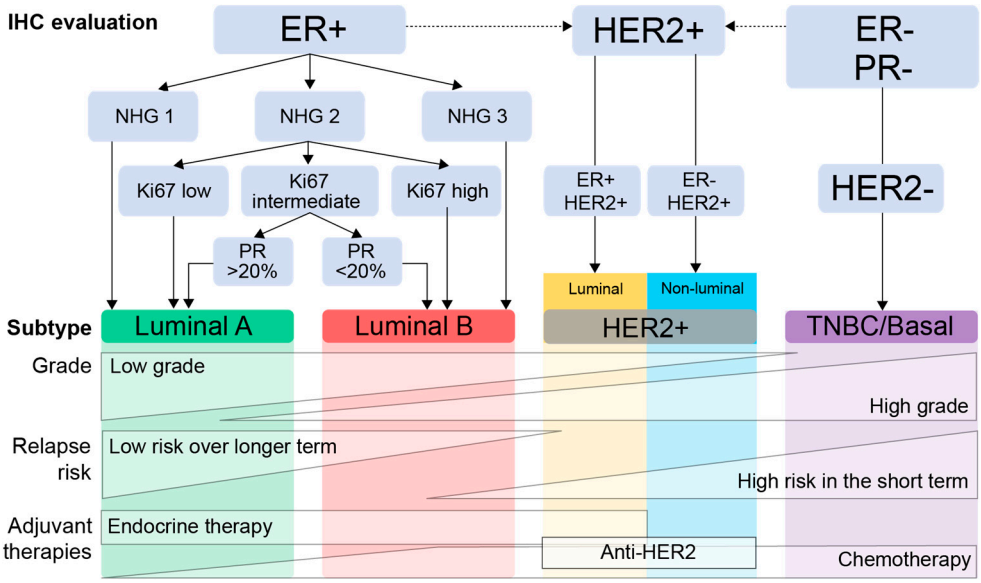


Figure 8: Surrogate instrinsic subtyping using immunohistochemistry. Luminal subtypes have a better prognosis with a tendency towards lower histological grade and lower risk for relapse, but are at risk for recurrence during a longer period. ER-negative subtypes have a higher risk for relapse in the short term and are often high grade at diagnosis. ER: estrogen receptor α; PR: progesterone receptor; HER2: human epidermal growth factor 2; NHG: Nottingham histological grade; + sign denotes positive status; - sign denotes negative status.

Treatment for breast cancer

Treatment regimens for breast cancer depend on the properties of the tumor disease, and include surgery, radiotherapy, endocrine treatment, immunotherapy, chemotherapy, and targeted therapies.

Surgery and radiotherapy

For most patients, surgery is the first step in their treatment regimen. Breast-conserving surgery combined with radiotherapy results in equal recurrence-rates compared to mastectomy [266] and in Sweden, 65% of non-metastatic tumors in the years 2015-2020 were excised using breast-conserving surgery [261]. Depending on growth characteristics, size and lymph node involvement, some tumors require pre-operative systemic treatment – neo-adjuvant therapy – to shrink the tumor and prevent relapse. Full mastectomy is indicated when partial mastectomy would not be sufficient for radical removal, or when adjuvant radiotherapy is contraindicated. In conjunction with surgery, sentinel node biopsy is normally performed [260]. Radiotherapy is recommended for all patients that have undergone breast-conserving tumor removal [260, 262].

Endocrine treatment

Lifetime exposure to estrogens is a risk factor for developing breast cancer and estrogens can likewise act as growth stimulators on breast tumor tissue. For hormone receptor-positive breast cancer, standard of care includes hormonal therapy with the aim of reducing ER activity. This can be achieved through estrogen deprivation, inhibition of ER or induction of ER degradation.

Selective estrogen receptor modulators (SERM) like tamoxifen have both agonistic and antagonistic effects depending on the tissue. Upon oral administration it is converted into potent metabolites, e.g. 4'-OH-tamoxifen, which have high affinity for the estrogen receptor. In certain tissues, such as breast, it has anti-estrogenic effects, whereas in other tissues, it acts as an agonist, thus reducing some of the adverse effects. This tissue-specific effect is not fully understood, but is at least in part due to differential expression of ER isoforms and co-factors [267].

Aromatase inhibitors (AI) are the first line hormonal adjuvant treatment for post-menopausal breast cancer patients. Anastrozole, letrozole and other AIs work by inhibiting the conversion of androgens into estrogens in the breast and other non-ovarian tissues [268]. This effectively results in estrogen deprivation and reduced ER activity in post-menopausal women. Pre-menopausal women have active ovaries as well as compensatory mechanisms through the hypothalamo-pituitary axis and do not get a consistent reduction in plasma estrogens on AI therapy alone, which is why tamoxifen is generally the first line treatment for these women [269].

Another approach to suppressing estrogen signaling is using selective estrogen receptor degraders (SERDs) like fulvestrant, which is used mainly in metastatic/advanced breast cancer [254, 260]. It acts both as an antagonist of the receptor and induces protein degradation thus reducing the total ER content of the cells. Meta-analyses have showed equal efficacy to other endocrine therapies in advanced breast cancer [270, 271] and it can be used in conjunction with targeted therapy with cyclin-dependent kinase 4 and 6 (CDK4/6) inhibitors [272, 273].

In pre-menopausal women, estrogen levels can also be drastically reduced by suppression of ovarian function. This can be achieved through surgical removal of the ovaries, through radiotherapy or treatment with luteinizing hormone-releasing hormone (LHRH) agonists/gonadotropin-releasing hormone analogs. LHRH agonists have been shown to reduce the risk of relapse in ER+ breast cancer in premenopausal women [274].

Continuation of endocrine therapy for ten years versus five years results in an additional reduction in recurrence [275]. Swedish and European guidelines recommend endocrine therapy for at least five years for patients with ER-positive disease (unless very-low risk), and also recommend extension of hormonal therapy for up to ten years, depending on individual risk and adverse effect profiles [260].

Chemotherapy

Cytotoxic agents are an important part of the treatment repertoire for most cancer types, but they are not used indiscriminately, due to the risk of adverse effects. In breast cancer, patients are stratified based on several risk factors, including patient age, tumor size, lymph node positivity, receptor status and tumor stage to determine whether to administer chemotherapy. Most patients with Luminal B, HER2-positive or triple negative breast cancer receive adjuvant chemotherapy which usually consists of anthracyclines and/or taxanes [262].

HER2 immunotherapy

Passive immunotherapy with the monoclonal antibody trastuzumab (Herceptin) was approved for use in advanced breast cancer in 1998 and has, in combination with chemotherapy, become standard-of-care for patients with HER2-positive tumors [260]. Trastuzumab and its biosimilars directly inhibit the extracellular portion of the HER2 receptor, reducing downstream intracellular signaling. For patients with HER2-positive disease that is advanced or has poor prognosis, additional HER2-targeted agents can be combined with chemotherapy and trastuzumab to enhance treatment response in pre- or postoperative therapy regimens, including the monoclonal antibody pertuzumab [276], and small molecule inhibitors such as lapatinib [277] and neratinib [278].

There are also available antibody-cytotoxic drug conjugates that take advantage of the cancer-targeting effect of the anti-HER2 antibodies to deliver a chemotherapy dose directly to the cancer cells. For example trastuzumab-emtansine improves invasive disease-free survival in HER2-positive treatment-refractory breast cancer [279] and is now part of the treatment repertoire for these patients.

New and future targeted therapies

Hormone therapy resistance can occur through increased activation of a number of pathways, including the CDK4/6-RB1 pathway. CDK4 and CDK6 are activated by cyclin D1 which leads to phosphorylation of RB1, allowing the cell to move from G1 to S phase [280]. One of the available CDK4/6 inhibitors is palbociclib, a small molecule that prevents the activation of CDK4/6 and when added to fulvestrant or AI treatment improves the prognosis in advanced breast cancer [281, 282]. CDK4/6 inhibition is now recommended as an addition to endocrine therapy in advanced ER-positive disease [254, 260].

As discussed previously, dysregulation of the PI3K pathway through *PIK3CA* mutations or *PTEN* loss are common occurrences in breast cancer, and there are therapeutic agents targeting the PI3K pathway. Some compounds have been implemented in the clinic for some time, such as the mTOR inhibitor everolimus in metastatic breast cancer, whereas others are still in development.

For the ~35% of breast cancer patients with a mutation in *PIK3CA* there will be new avenues of treatment, such as small molecule inhibitors of PI3K. Alpelisib, a specific inhibitor of PI3Kα, administered together with endocrine therapy, was recently found to improve progression-free survival, although with a statistically non-significant effect on overall survival (OS), in hormone receptor-positive, HER2-negative, *PIK3CA*-mutated breast cancer. [283, 284]. Alpelisib has been approved by the American Food and Drug Administration as well as the European Medicines Agency and will be available in Sweden soon. There are also pan-PI3K inhibitors in development, such as buparlisib and pictilisib which have completed phase III clinical trials [285, 286], as well as compounds that target the PI3K downstream effector AKT that have shown promise in phase II studies [287-289].

Cancers with a mutation in one of the *BRCA* genes are more vulnerable to treatments that block other DNA repair proteins. PARP inhibitors target poly-ADP-ribose polymerase (PARP), which is part of the single-strand DNA break repair system. PARP inhibitors trap the PARP repair complex, leading to stalled replication forks. In normal, HRR-sufficient cells, these replication forks are resolved by the HRR machinery, but in HRR-deficient cells, inhibition of PARP in combination with cytotoxic chemotherapy give rise to synthetic lethality as the cancerous cells can no longer repair DNA damage [290].

Tumors have developed strategies to evade destruction by the immune system, including upregulation of immunosuppressive molecules, like the PD-1/PD-L1 proteins. Antibodies and small molecule inhibitors of PD-L1 have emerged as a promising therapy in several cancer types, including melanoma and non-small cell lung cancer [291]. PD-L1 antibody atezolizumab together with chemotherapy is now also a treatment option in PD-L1-positive metastatic TNBC, with clinical trial results showing extended progression-free survival and overall survival [292, 293].

Methods

Patients and samples

All patient studies were performed in compliance with the World Medical Association Declaration of Helsinki, with approvals from the regional ethical committee. Trained health professionals provided written and oral information and all patients gave signed written informed consent. All confidential data was handled in accordance with Sweden's Personal Data Act and the General Data Protection Regulation, where applicable.

Study I

Eleven patients diagnosed in the south Swedish healthcare region during 1986-1997 with available frozen specimens from both primary tumor and at least one distant metastasis were included in the study. Blood samples were obtained from three of the patients and from 7 unrelated individuals which were used as normal controls in whole-genome sequencing analysis. DNA was extracted from fresh-frozen tissue using AllPrep (Qiagen). Clinical data was retrieved from patient charts, except for HER2 status, which was determined through DNA copy number analysis.

Study II:

Ten patients with metachronous CBC, previously part of a larger cohort of CBC [294] with available fresh frozen tumor tissue and detailed patient information were included. The patients were diagnosed with a first invasive breast cancer. Nine developed a subsequent contralateral invasive breast tumor and one a contralateral *in situ* lesion. Thirteen out of the 20 tumors had paraffin-embedded material available, allowing for IHC re-evaluation of standard clinical biomarkers by a pathologist, and for the remainder, clinicopathological data was obtained from patient charts. DNA was extracted using AllPrep (Qiagen).

Study III and IV:

The Sweden Cancerome Analysis Network – Breast (SCAN-B) study is a multi-center, prospective observational study initiated in 2010 that aims to include all patients diagnosed with primary breast cancer at each of the participating hospitals [295, 296]. Today, nine cancer care units in the south of Sweden as well as

Uppsala participate and more than 17,000 patients have been enrolled, an inclusion rate of 99%. Patients receive standard-of-care and tumor samples, where available after clinical pathological analysis, are collected for research and, in the first phase, are all RNA sequenced (>11,000 to date). The patients in study III and IV are a subset of consecutively diagnosed patients (Figure 9) with invasive primary tumors diagnosed between September 1 2010 and March 31 2015, previously described by Brueffer *et al*. [297], which was reduced to 3207 patients in study IV due to additional quality checks and implementation of advanced analysis pipelines. The cohort is population-based and patients were included based on no prior contra- or ipsilateral disease. Tumor samples are preserved in RNAlater (Ambion/ThermoFisher) immediately after surgery, flash frozen, and processed according to standardized SCAN-B protocols [295-297].

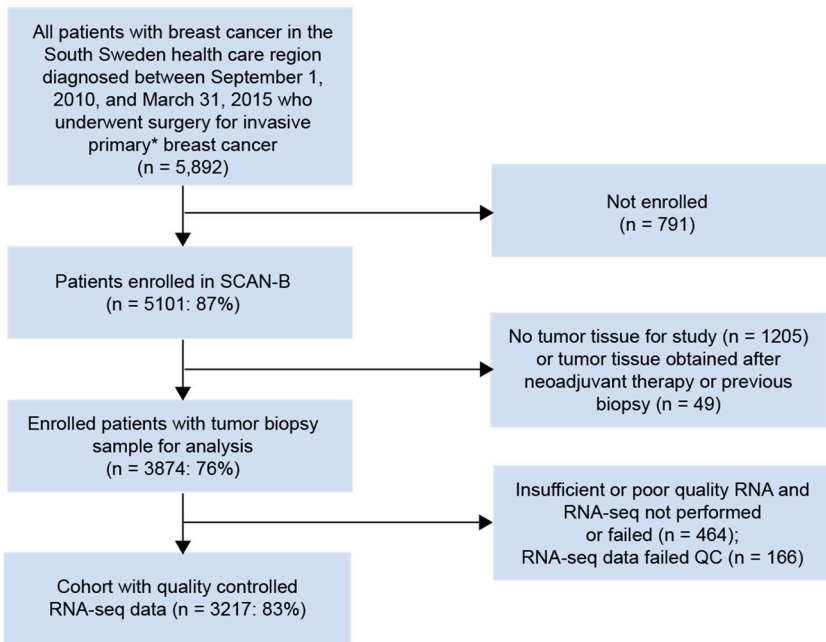


Figure 9: The SCAN-B cohort selection diagram for study III. QC: quality control. * non-metastatic unilateral primary breast cancer, excluding patients with synchronous contralateral tumor (within 3 months of diagnosis). Adapted from Brueffer *et al*. [297] (CC-BY 4.0).

DNA and RNA analysis

Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) was developed in the 1980's [298] and laid the groundwork for the bulk of genetic analysis methods used today. It is a method wherein specified nucleic acid sequences are copied and exponentially amplified. The reaction components include template DNA, the DNA replication enzymes from a heat resistant microorganism (Taq polymerase), primers, free nucleotides and a thermal cycler. Primers are starter sequences that dictate which specific segment will be amplified. By cycling between hotter and cooler temperatures, the DNA strands are artificially separated allowing access for the primers and polymerase, duplicating the sequence exponentially, resulting in millions to billions of copies [298].

Quantitative PCR methods

Quantitative PCR (qPCR; also called real-time PCR) is a development of the PCR technology. By using fluorophores or sequence-specific fluorescent probes and measuring fluorescent intensity, it is thus possible to performing relative quantification by comparing the fluorescent intensity to reference sequences.

Digital PCR (dPCR) is a further development of qPCR, and the most common type is droplet digital PCR (ddPCR). Rather than running a bulk amplification, the aqueous template and reagent mix is randomly partitioned into reaction compartments – droplets separated by oil in ddPCR – with a volume in the nanoliter range. Ideally, each droplet contains either zero, one or only a few of the nucleic acid sequence(s) of interest. The amplification reaction is then performed and a fluorescent probe (or some other indicator) is used to generate signal if the target sequence(s) is present within the droplet [299]. The "digital" designation refers to that each droplet is scored either "positive" or "negative" based on the signal reaching above a threshold indicating presence of the target. Since the partitioning is a random process following Poisson statistics, from the number of positive and negative droplets, a highly accurate quantification can be made.

Importantly, dPCR-based methods yield an absolute measurement of the number of sequence copies in the sample, and thus do not rely on relative quantifications using reference standards as do other methods such as qPCR. This method enables highly sensitive and specific detection and quantification of minute amounts of genetic material [300], such as tumor DNA released into the blood stream of cancer patients [301], and an improved variant of dPCR called IBSAFE was used in study III to validate presence of mutations in DNA of tumor samples and detect mutations in relapse material.

DNA and RNA sequencing

Principles of sequencing

The Human Genome Project – coordinated by the Human Genome Organization (HUGO) – was, at the time of its launch in 1990, a massive undertaking which aimed to map out the complete human DNA sequence. It took 13 years to complete and has since been a foundation for human molecular genomic research [302]. In the past two decades, advances in sequencing technology and data processing have greatly reduced the required time and cost of whole genome sequencing and as a result, thousands of genomes have been sequenced from breast tumors alone [228, 303, 304].

The first method of sequencing developed by Sanger *et al*. in the seventies exploits the properties of dideoxynucleotides triphosphates (ddNTPs), which lack the two hydroxyl chemical groups on the ribose which are necessary for further elongation of the DNA strand. When incorporated into an elongating DNA strand, they terminate replication and result in a shortened fragment. A PCR reaction with for example ddTTP nucleotides mixed in will result in amplicons of varying length, always terminated on a thymine base. Thus, four separate reactions, each containing either ddATP, ddTTP, ddGTP or ddCTP, and subsequent fragment separation with gel electrophoresis could reveal the DNA sequence [305]. Sanger sequencing subsequently evolved to include fluorophore-labeled ddNTPs, so that the reactions could be pooled, and the sequence decoded through detection of fluorescent signals [306].

Massively-parallel sequencing

Next generation sequencing (NGS), also known as massively-parallel sequencing, is fundamental to all four studies (study I-IV) presented here. There are two principally different methods. One is sequencing-by-ligation [307], which utilizes fluorescently labeled oligonucleotides that, upon complementary hybridization to the template material, is ligated to the polynucleotide backbone, creating a fluorescent signal corresponding to the ligated base. The second, sequencing-by-synthesis, refers to methods that are conceptually similar to the original Sanger approach, but modified and vastly scaled-up, allowing simultaneous sequencing of tens to hundreds of millions of template fragments in a “flow cell” where reagents are flowed over tethered clonal reactions. By hybridization of DNA strands to the surface of the flow cell, and cycles of chemicals for base additions, the sequence of millions of nucleic acid fragments can be imaged and read in parallel.

Each sequencing method has its advantages and drawbacks. Methods that rely on PCR amplification of template prior to sequencing may produce a PCR product with uneven representation of the template, leading to misrepresentation of copy number data, although this can to some extent be corrected for at the bioinformatics stage. Additionally, too many cycles of sequence amplification

increases the risk for technical errors, polymerase base misincorporation errors and also underrepresentation of GC-rich regions [308]. The length of the sequenced fragments is limited to a few hundred base pairs at most, leading to problems with highly repetitive regions. In sequencing-by-ligation methodologies, palindromic sequences are problematic [300].

"Third generation sequencing" methods are coming into use, such as single molecule real time sequencing (SMRT, Pacific Biosciences) [309] and nanopore sequencing (Oxford Nanopore Technology) [310]. These new technologies enable sequencing of single DNA molecules, circumventing the need for clonal amplification of the templates and furthermore, producing long reads, more than 10 kb, or in some cases only limited by the sample DNA molecule length [311]. A drawback has been the high error rates (~15%) but these issues are being addressed through approaches like multiple pass sequencing and improvements in data analysis bioinformatics [312].

Currently, the sequencing-by-synthesis platform commercialized by Illumina (formerly Solexa, established in 1998) is the dominating method for high-throughput sequencing [313], and was used for sequencing of samples described in studies I-IV, and the procedures described hereafter refers to the Illumina sequencing method.

Library preparation

Genomic DNA is first fragmented to a length of up to 1500 base pairs long, and often shorter with a median length between 300-800 bp. This can be achieved through sonication or cation catalyzed enzymatic digestion. The fragment ends are then repaired and the 5' end is phosphorylated through enzymatic reactions. Several adenosine residues are added to the 3' end to facilitate the next step – adapter ligation. Adapters are oligonucleotides that serve as a connection for the DNA strands to attach to the flow cells during the sequencing reaction and they also contain identifiers or "barcodes" that allow for pooling of multiple samples into the same sequencing reaction. Finally, the DNA library is amplified in a PCR reaction and undergoes quality controls [314]. PCR cycling, although commonly used for library amplification, is known to introduce some biases. In general, a shorter program is used and some library preparation protocols omit the PCR step altogether [315]. In studies I and II, 10 cycles was used, and for RNA-seq in studies III and IV, 12 cycles of PCR.

In studies I-II, one microgram of fragmented DNA was used for library preparation with the TruSeq DNA sample preparation kit (Illumina) according to the manufacturer's instructions. Before PCR amplification, each library was size-selected for fragments between 550-950 bp using agarose gels. Libraries were analyzed with BioAnalyzer (Agilent), and the concentrations were measured using a Qubit spectrophotometer (Invitrogen).

Library preparation of mRNA, as performed in studies III and IV, requires additional preparation steps, including the production of complementary DNA (cDNA). mRNA was purified from total RNA extracts using magnetic beads targeting the poly(A) tail of the transcripts. cDNA synthesis was performed using a dUTP protocol [295-297, 316] that retains strand directionality allowing selective degradation of the second strand, prior to sequencing. In other respects, the sequencing library generated from an RNA sample is similar to a DNA library from a DNA sample.

Sequencing-by-synthesis

Sequencing takes place in a solid phase amplification reaction. The prepared library (from DNA or RNA starting material) is loaded onto a flow cell, wherein primers complementary to the library adaptors are attached to the surface at high density but with adequate spatial separation between tethered primers. The adapter-ligated DNA strands hybridize with the attached primer strands and the clustering step takes place. One DNA fragment can form a bridge between two primers and after several rounds of amplification, a "cluster" of clonal DNA forward and reverse strands is formed *in situ* (Figure 10) [317, 318]. At the end of clonal amplification, reverse strands are cleaved and washed away, leaving only forward strands. At this point, the sequencing step commences (Figure 11).

Fluorescently tagged nucleotides and enzymes are added to the flow cell. The nucleotides carry a reversible terminator that ensures that only one base is added

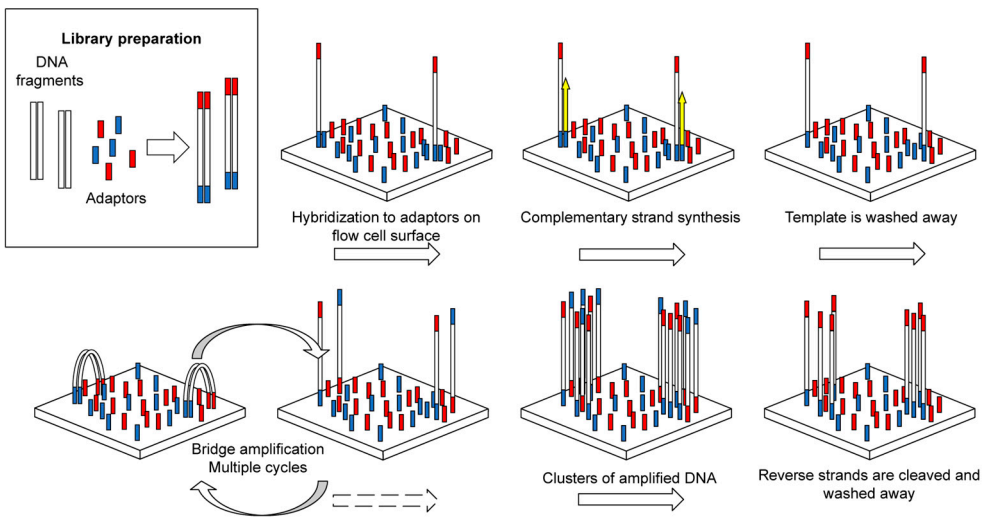


Figure 10: Cluster generation on the Illumina flow cell.

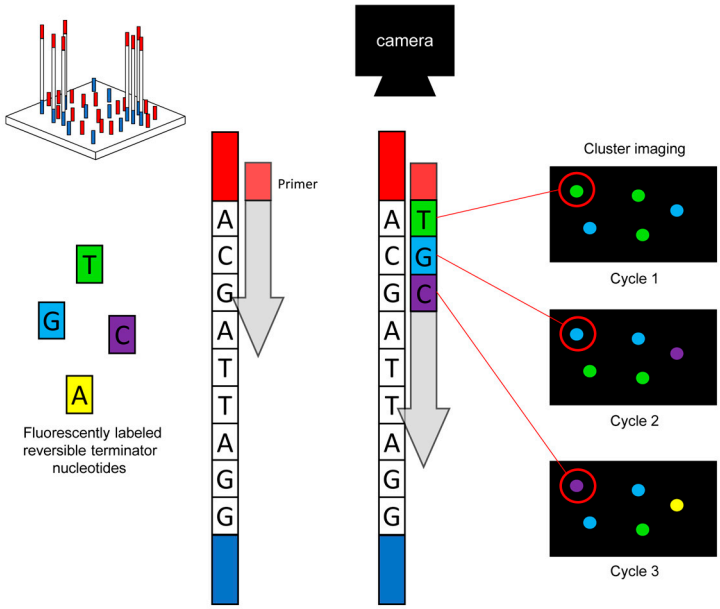


Figure 11: The Illumina sequencing reaction and imaging. Clusters of clonal DNA strands are sequenced by adding fluorescently labeled reversible terminator nucleotides in cycles. For every cycle, the clusters are visualized as a spot of color corresponding to the last added base.

for each round, but which can be cleaved to allow subsequent addition of more bases. Thus, the template DNA is copied, one base at a time, and for every added base, the clusters are photographed, recording the color of that base.

Paired-end sequencing means that the template DNA is sequenced from both ends, resulting in two short sequences of customizable size, for example 100 bp, which may overlap or may be separated by the length of the remaining DNA fragment. This allows for easier alignment and better detection of rearrangements and indels [319], and was utilized in all four studies.

Illumina has launched several machines utilizing their solid-phase, sequencing-by-synthesis method, including both benchtop and production-scale sequencers. In our studies, sequencing was performed using HiSeq 2000 (studies I and II) and for the SCAN-B patients (study III and IV), either HiSeq 2000 or NextSeq 500 were used.

Analysis of sequencing data

> "As was predicted at the beginning of the Human Genome Project, getting the sequence will be the easy part as only technical issues are involved. The hard part will be finding out what it means, because this poses intellectual problems of how to understand the participation of the genes in the functions of living cells."
>
> Sydney Brenner in "Loose End" (1995) [320]

The human genome contains more than 3 billion basepairs. Adding that massively parallel sequencing usually sequences the genome multiple times, it quickly generates massive amounts of data that requires extensive computational resources and bioinformatic processing. The greater the "sequencing depth" or "coverage" the higher the confidence with which variations from the normal sequence can be called. Greater coverage can also increase the capture subclonal variations within a population of cells.

In principle, RNA- and DNA-seq data involve similar analysis steps to yield interpretable information, but different bioinformatic tools may be required. Given equal library input, RNA-seq enables higher coverage for transcribed exonic sequences compared to DNA sequencing, but instead poses challenges due to RNA processing such as alternative splicing and RNA editing and thus requires additional data analysis steps. The RNA-seq bioinformatic analysis for studies III and IV followed the SCAN-B analysis protocol [295, 321]. The principal steps for analysis of sequencing data are described below.

Read alignment

The sequencing reaction yields millions of sequence fragments for each sample and in the first step the sequence fragments are aligned to a reference genome, containing both a consensus sequence for all human chromosomes and some alternative version sequences that are found in subpopulations, referred to as "decoys", since they prevent faulty alignment of reads [322]. The bioinformatic tool Novoalign (Novocraft Technologies) was used for alignment of whole genome sequence reads to reference genome GRCh37; SNP patched; with decoy sequences in studies I-II. For the RNA-seq studies III-IV, HISAT2 2.0.5 [323] was used for RNA-seq alignment to reference genome GRCh38.p8 (including alternative sequences and decoys), patched with dbSNP Build 147 common SNPs and the GENCODE 25 transcriptome model. Once processed, the sequencing reads from each sample is assembled in a binary alignment map/format (BAM) file, which holds the information of their position relative to the reference genome.

Tagging duplicate reads

Duplicate reads, for example from PCR amplification steps, are a common occurrence in sequencing data and may lead to over-representation of some fragments, and consequently biases downstream analyses such as copy number variation calling. With whole genome input, true identical duplicates are statistically unlikely to occur by chance, and are therefore most often attributed to technical causes, most importantly PCR amplification. Therefore, analysis software is employed to flag identical reads as duplicates so that they can be disregarded in the following analyses. In studies I-II, this was done with the tool Picard v 1.66 [324]. However, with less complex library input, such as whole mRNA, the likelihood for "natural" duplicate reads increases, and removing them may reduce analysis sensitivity and skew estimates of gene expression levels [325]. Consequently, while the SCAN-B analysis pipeline does mark duplicates, they are not removed in the subsequent gene expression analysis (study IV). Mutation detection using RNA-seq data requires a different analysis strategy and in study III, duplicate reads were marked using SAMBLASTER 0.1.24 [326] and removed in the downstream analyses.

Detection and comparison of structural variants

SVs can be identified by *in silico* analysis of the BAM file, by identifying discordant reads and split reads. Paired-end sequencing results in two short reads that are overlapping or separated by an unsequenced fragment of a length that should roughly correspond to the average library fragment size, minus the length of the reads. Discordant reads either have an unexpected 5' to 3' orientation relative one another, or they span an unexpected distance or even map to separate chromosomes. Deletions and amplifications will result in a regional increase or decrease in number of reads whereas aberrant directionality of the reads may be informative of inversion events. Paired reads that each map to a different chromosome are indicative of a chromosomal translocation. Ideally, the breakpoints of a rearrangement are covered by split reads – i.e., the exact sequence of the rearrangement breakpoint is known – but with a low coverage approach, this is not always the case, and the breakpoint is instead estimated computationally from the available sequencing data.

For calling of variants in tumors, it is optimal if there is a patient-matched normal control against which the tumor sequences can be compared, to avoid calling germline variants. This is far from always possible since normal tissue frequently is unavailable, and therefore some software allows for calling of SVs without matched controls. FREEC [327], for instance, can instead use internal sample markers for CNV characterization. In studies I-II we had available normal, non-tumor DNA from a subset of the patients as well as from some un-related, healthy individuals, and we used these samples to create a pooled normal control. FREEC was used for CNV analysis and BreakDancer [328] for SV calling.

In the comparison of SVs between paired tumors in studies I-II, overlap of rearrangements were evaluated using BedTools v2.18 and rearrangements were considered shared (present in both tumors) if one predicted breakpoint fell within 500 bases of a breakpoint in the other tumor. Through this process, rearrangements were defined as specific to either tumor 1 or tumor 2 or shared between them, enabling the calculation of percentage similarity. The shared percentage for one tumor in a pair was calculated:

$$Fraction\ of\ shared\ SVs\ in\ T1 = \frac{Number\ of\ SVs\ in\ T1\ that\ are\ also\ present\ in\ T2}{Total\ number\ of\ SVs\ in\ T1}$$

Where T1 denotes tumor 1 and T2, the paired tumor (tumor 2). In study II, the total shared percentage for both tumors was used:

$$Fraction\ of\ shared\ SVs = \frac{Number\ of\ shared\ SVs\ in\ T1\ and\ T2}{Total\ number\ of\ SVs\ in\ T1\ and\ T2}$$

Copy number states are detected as an increase or decrease of the number of sequencing reads covering that region and assigned as either gains or losses. The degree of similarity between two tumors can thus be calculated by assessing the number and length of the regions with altered copy number that are present in both samples.

RNA-seq for characterizing mutations and gene expression patterns

RNA-seq has supplanted the previously dominating technologies, microarrays, which was a popular option when exploring gene expression patterns in tissues. There are of course limitations to RNA-seq analysis, particularly for mutation detection. The analysis for sequence variants in RNA is limited to transcribed genes and is likely to miss e.g. alterations in tumor suppressor genes and variants that result in nonsense-mediated decay of the transcripts. The reverse transcription of RNA to complementary DNA can also introduce artifacts and cellular RNA editing may give rise to variants that are not present in the genome. Meanwhile, the advantages with RNA-seq are that it has superior dynamic range, does not rely on probes for detection of transcripts [329] and allows de novo characterization of RNA sequences, including fusion transcripts, RNA-editing, alternative splice variants, indels, and SNVs [330].

In study III, RNA-seq data is used to identify mutations as described by Brueffer *et al* [234]. Variants were detected using VarDict-Java [331] and were annotated using vcfanno 0.3.1 [332] with several resources, including RNA-editing databases [333-336], clinically relevant variant databases [337-342] and normal variants [343, 344], including an in-house constructed 10-sample collection of tumor-adjacent normal tissue run through the SCAN-B pipeline. Variant effects were predicted using SnpEff 4.3.1r [345].

Importantly, a series of filters were then applied to reduce noise and enhance detection of clinically relevant variants. We applied negative filters for low complexity regions, SNP status and RNA editing, and positive filters to rescue any variants present in databases for clinically relevant mutations. For the purposes of the study, we interrogated the set of mutations for 13 *ESR1* mutations experimentally verified to induce increased ER activity in the absence of ER ligands and/or presence of anti-hormonal compounds [86, 87, 200, 204, 206, 346-351].

RNA-seq gene expression

For gene expression analysis, the read count aligned to each gene must first be converted to a value relative to the total number of reads of the sample and the length of each gene. For single-end sequencing, each read represents one sequenced fragment, whereas paired-end sequencing results in two reads per fragment, which is taken into account when calculating the gene expression values. This results in measures called FPKM (paired-end sequencing) or RPKM (single-end sequencing), abbreviations for Fragments/Reads Per Kilobase of exon per Million mapped reads [352]. The values are calculated by dividing the number of reads for a gene with the total read depth of the sample and then normalizing it to the length of the gene.

$$FPKM = \frac{Number\ of\ fragments\ mapped\ to\ the\ transcript}{Transcript\ length\ (kilobases) * Total\ number\ of\ fragments * 10^{-6}}$$

TPM (transcripts per million) values are calculated by first normalizing reads to transcript length, generating the reads per kilobase of transcript (RPK) value. The RPK of the gene is then normalized using the sum of all RPK values for the sample, yielding the TPM, which will have a constant average across samples, improving comparability of gene expression, although it does not take into account global shifts in RNA quantity between populations of cells [353, 354].

$$RPK = \frac{Number\ of\ fragments\ mapped\ to\ the\ transcript}{Transcript\ length\ (kilobases)}$$

$$TPM = \frac{RPK}{Sum\ of\ all\ RPK\ values\ for\ the\ sample * 10^{-6}}$$

To generate TPM values in study IV, the software Cufflinks was used [355, 356]. The TPM values were log2 transformed (after addition of 0.1 to each value since $\log_2(0)$ is not defined) to reflect the fold-change of the transcript abundance.

Differential gene expression analysis considers differences in gene mRNA levels between samples to find genes that are co-expressed or under-expressed together with a factor of interest. There are different statistical methods to achieve this end.

In study IV, we used the voom function from the limma package [357] to normalize and log-transform the TPM values. We fitted linear regression models and mean variance was calculated using the empirical Bayes method, computing the differential gene expression statistics. Furthermore, genes were grouped according to their biological function to analyze which pathways are regulated in *ESR2* high and low contexts. This was achieved using the fgsea package [358].

Statistical analysis

> "The main purpose of a significance test is to inhibit the natural enthusiasm of the investigator."
>
> Frederick Mosteller in *Selected Quantitative Techniques* (1954) [359]

Two-group comparisons and statistical significance

In two-group hypothesis testing, the structure of the data guides the choice of the type of statistical test. Data that conforms to normal distribution can be tested using parametric tests (such as Student's t-test or Chi-squared test), whereas data that does not follow the normal distribution are tested using a non-parametric test (such as Mann-Whitney U-test or Fisher's exact test) [360]. The test evaluates the probability (p) that any observed difference between the groups should arise by chance.

A p-value of 0.05 is commonly used as an upper limit for statistical significance, however, the measure is continuous. The p-value should be interpreted from the general context; p-value of <0.05 is not necessarily statistical proof that the null hypothesis can be rejected [361]. Additionally, when a large number of tests is being performed, the likelihood of encountering a "significant" p-value increases, and it should thus be subject to correction for multiple hypothesis testing.

The confidence interval (CI) describes the range of possible values for the variable of interest with a certain precision, e.g., 95%, analogous to a p-value of 0.05. The confidence interval is one way of addressing the risk for sampling error [362]. Often both p-value and CI are presented since they complement each other [360].

Survival analysis

Survival analysis is a branch of statistics dedicated to analyzing the time from diagnosis to death or another clinically relevant endpoint, like progression or relapse in cancer. The endpoints for survival analysis in cancer research have been defined and summarized by the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) initiative [363]. In study III, overall survival (OS) and relapse-free survival (RFS) were utilized as endpoints. OS events are defined as death from any cause, whereas an RFS event signifies that a

patient has either suffered a local or distant tumor relapse, or died. In study IV, we utilized the relapse-free interval (RFI) rather than RFS for better comparison with our validation dataset. The endpoint for RFI was defined as locoregional or distant relapse.

Kaplan-Meier (KM) analysis allows for analysis of time to event for patients with different start dates by setting the start point, or in this case the date of diagnosis, to day zero. Then the time to event or until last follow-up can be readily compared between patients. A KM-plot shows the fraction of patients in each group that were exempt from suffering an event (survivors) plotted over time on the x axis. Unless an event has occurred at the time of the last follow-up a patient is excluded from the group – censored – from the plot going forward. Curves that diverge may signify that the stratification factor may have prognostic significance [364]. The statistical relevance of such a difference is most commonly investigated using the logrank test [365].

Further analysis and quantification of the effect on survival can be performed by calculating the hazard ratio (HR) using Cox regression, which also allows for multivariable analysis, where possible confounders can be incorporated into the analysis. The HR value represents the risk increase over time for one group compared to a reference group [366].

Cox regression works on the assumption of proportional hazards, i.e. that the factors separating the groups have a multiplicative impact on outcome and do not change over time. Non-proportionality can for example be suspected if the survival curves cross. Since the assumption is fundamental to the Cox regression it is important to test data for non-proportionality. In study III and IV, the proportional hazards assumption was tested using the Schoenfeld residuals method [367].

Results and Discussion

> "…We are approaching an era in which treatment decisions will be based on tumour molecular abnormality profile or "signature," rather than tumour tissue type or anatomical site of origin, improving patient prognosis and quality of life."
>
> Sarah D Jackson and John D Chester [368]

In the first part of the thesis (studies I and II), chromosomal rearrangements are identified in paired tumors from breast cancer patients and used to characterize the clonal relationship between primary and metastasis or contralateral tumors. Study III is dedicated to identifying and describing mutations in ERα in 3217 primary invasive breast tumors and their associations to patient outcome. In study IV, we have characterized mRNA expression patterns of ERβ in primary breast cancer.

Study I

In study I, our aim was to identify genomic structural variants in paired primary breast tumors and matched distant metastases and determine the level of similarity between them. We performed low-coverage (range 7.5-14.9x sequence coverage) whole genome sequencing of tumor DNA for eleven patients with matched primary tumor and metastasis tissue samples available. We found a median of 85 (range 18-404) rearrangements per tumor, with slightly fewer in the primary tumors (median: 82; range 26-310) compared to metastases (median: 87; range 18-404). In any given tumor there was a high concordance with the paired sample (median: 89%, range 61%-100%), compared to 3% between randomly matched tumor pairs. In 9 out of 11 patients, there was a net gain of rearrangements in the metastasis compared to the primary tumor. The paired tumors were less similar regarding copy number aberrations, with only 16-43% shared events, compared to 5-28% in the randomly matched tumor pairs.

Apart from a net gain in rearrangements in the metastases, we also found that in some patients, there was enrichment of subsets of rearrangements, perhaps indicating clonal expansion of a subpopulation containing those rearrangements. Rearrangement patterns varied, with some tumor pairs exhibiting variants that were spread across the entire genome and others that mainly had focal clusters

limited to a chromosome or chromosome arm, reminiscent of the pattern described for chromothripsis.

Comparison of randomly generated tumor pairs showed rearrangement match between tumors of 0-36% and although 36% is an outlier value, it shows that some degree of similarity in the background must be expected. One possible explanation for this is germline- or population-specific rearrangements – indeed, in our PCR validation we found that some of the rearrangements in tumor DNA were also detectable in matched normal DNA or pooled DNA from six unrelated control persons (8/62 rearrangements). Although this could be a source for overestimation of the percentage of shared rearrangements, our validation also showed that 10 out of 51 rearrangements (20%) that were predicted to be specific to either the primary tumor or the metastasis was detectable by PCR in both tumors, indicating that, in total, the shared percentage may still be slightly underestimated.

We found that, in general, primary and metastasis pairs are very similar with regard to genomic rearrangements, with more than half of the events in one tumor also to be found in the other. Shared genomic rearrangements can be a strong indicator of a shared clonal origin. Moreover, the absence of identical events does not completely exclude the possibility of clonality, especially if the total number of rearrangements is low. From our data, it appears that most rearrangements are early events in the process of tumor evolution and likely present in a large fraction of the tumor cell population. In contrast to point mutations, which also show extensive intratumoral heterogeneity [71], rearrangements are more readily detectable with a lower-coverage approach which is increasingly feasible as sequencing costs continue to drop. This could also make rearrangements attractive targets for creating personalized assays for tracking minimal residual disease.

Study II

Contralateral breast cancer (CBC), defined as when a patient previously diagnosed with breast cancer, suffers a second tumor in the opposite, healthy breast, is clinically treated as a new primary tumor unless there is specific suspicion of metastasis. In study II, we investigated a collection of 10 paired breast tumors from patients with CBC. We performed low coverage (sequence coverage 1.8-11.2x), whole genome sequencing and analyzed the sequencing data for structural variants and copy number aberrations. We found that rearrangements were readily detectable in both first (BC1) and second (BC2) tumors and that most pairs of contralateral tumors had a low percentage of shared rearrangements (6-15%), consistent with the tumors being independent primary disease. Three of the pairs showed a higher degree of shared events with 39%, 46% and 75%. Our previous characterization of primary tumors and paired metastases (Study I) had revealed

similarities greater than 50% between clonally related tumors and thus we could conclude that at least one contralateral tumor (with 75% shared events) certainly represented a spread of the first tumor, with two other tumors that were ambiguous.

Although the number of patients in this study was low, the finding that at least one patient in ten had a metastasis that was clinically interpreted as a new primary tumor is significant. The findings of our study has since been confirmed by other authors using exome sequencing and identifying concordant mutational profiles in 3 out of 25 contralateral tumor pairs [369]. Although more studies must be undertaken before any general conclusion can be made, the results suggest that about one in ten cases of contralateral disease may in fact represent metastatic spread. If no *in situ* component can be found in the second tumor, then it may be relevant to expand the analysis to incorporate markers that will allow the clonal origin of the new tumor to be defined with more certainty.

Study III

As previously discussed, mutations in *ESR1*, have been identified as an acquired mechanism of resistance to endocrine therapy (ET) [86, 87, 200, 204]. This group of mutations, mostly comprised of SNVs, have the potential to render the tumors unresponsive to treatment with pharmaceutical compounds that would normally repress ER function in cancer cells. ERα-activating variants have almost exclusively been found in metastatic or treatment-refractory breast cancer, although studies incorporating primary tumors report occurrence in 0-7% [86, 87, 370, 371].

We looked for the presence of known activating mutations, previously characterized to induce constitutive ERα activation in *in vitro* and *ex vivo* studies. We analyzed RNA-seq data from 3217 SCAN-B patients with invasive tumors, diagnosed in the years 2010-2015. We identified *ESR1* resistance mutations in 30 out of 3217 cases (0.9%) in the full cohort, and out of the 2720 ER-positive tumors, 29 cases (1,1%) were *ESR1* mutated. For two of the *ESR1* mutant tumors we also acquired tissue or cytological samples from relapse material and ddPCR analysis revealed that the resistance mutations were still present in these lesions.

In ET-treated disease, presence of *ESR1* mutation was significantly associated to poor relapse-free and overall survival ($p=0.0006$ and $p=0.008$, respectively), with hazard ratios of 3.00 and 2.51, respectively, which remained significant when adjusted for other prognostic factors.

This is, to our knowledge, the single largest study of *ESR1* resistance mutations in primary tumors, and the first to show that these mutations, already existing at

diagnosis and prior to treatment, are associated to a poor outcome in the cohort receiving ET. Given the small number of mutant tumors we could not stratify based on specific mutations, but other authors have demonstrated that the mutations may have diverse characteristics, both in regard to downstream transcriptional activity [372, 373] and pharmacodynamic profiles [206]. The most common mutation was E380Q (10 patients), which has a modest effect on ERα activity *in vitro*, and indeed, only one out of ten patients carrying this mutation had a death or relapse event.

If our results can be replicated, it could be considered highly relevant to screen for *ESR1* mutations already at diagnosis, since novel endocrine therapies that are more effective against the mutant tumors are in development [209, 210].

Study IV

ERβ is a receptor for estrogen, which can form both homodimers and heterodimers with ERα, resulting in transcriptional regulation of a distinct subset of target genes [161]. There are indications that ERβ is associated with better prognosis, attenuating the growth-promoting signaling by ERα, but previous studies have yielded conflicting results and protein detection has been a challenge due to the lack of well validated antibodies [170, 179].

We studied *ESR2* mRNA detected through RNA-seq in the SCAN-B sub-cohort (study III, reduced to 3207 samples due to additional quality controls) and its relation to the expression of other genes as well as impact on patient survival. *ESR2* mRNA was expressed at low levels, with one third of tumors *ESR2*-negative, and transcription levels were highest in the Normal-like, Basal-like and HER2-enriched PAM50 subtypes. We found that there was a small, but significant association with OS in the full cohort (p=0.006, logrank test) as well as the ET-treated (p=0.03, logrank test) and TNBC (p=0.01, logrank test) subgroups, with a better outcome in the tertile with highest *ESR2* expression. Our multivariable analysis, incorporating age, tumor size, lymph node status and grade showed that the effect on overall survival was independent of these co-variates for the whole group and for the TNBC group, but not the ET-treated patients. We also analyzed relapse-free interval (RFI) but could find no significant association of *ESR2* expression with this endpoint.

Furthermore, we set out to validate the findings using RNA-seq data from the TCGA primary breast tumors and found that there was a similar trend for the entire cohort (p=0.1, logrank test), and ET group (p=0.1, logrank test), but not for TNBC. Additionally, we found that high *ESR2* expression was associated with better outcome for ERα-negative and HER2-positive tumors in TCGA, both with regard to RFI and OS (p=0.02 and p=0.03 respectively, logrank test).

We performed differential gene expression analysis of *ESR2*-high versus *ESR2*-low tumors in SCAN-B ERα positive and negative subgroups, which revealed that high *ESR2* expression was associated with transcription of genes involved in immune activation, regardless of ERα status, and this could provide a rationale for the association to outcome.

In summary, we showed that *ESR2* is expressed at low levels in primary breast tumors, but that higher expression is associated to improved survival. Our results are in line with the hypothesis that ERβ possesses tumor suppressor properties and is down-regulated during tumor development. Methods relying on tissue mRNA quantification such as qPCR and RNA-seq have a different set of challenges to antibody-based assays, including the inability to discriminate between tumor cell expression and stromal expression as well as the technical challenges introduced by template amplification and RNA editing. Nevertheless, with the contradictory evidence derived from antibody-based assays, mRNA quantification may contribute valuable information to characterize the role of ERβ in breast cancer.

Conclusions

- Chromosomal rearrangements are relatively stable throughout the evolution of a tumor and can be identified through low coverage (10x) whole-genome sequencing and subsequently detected with PCR assays tailored to the tumor. Thus, they could serve as personalized biomarkers for investigating tumor clonality and tracking disease progression.

- Contralateral tumors may represent a metastatic spread and it would be relevant to expand current clinical diagnostics to more definitively ascertain the clonal origin of a second tumor and treat contralateral spread appropriately.

- Genomic endocrine therapy resistance mutations in *ESR1* occur in around 1% of ER-positive primary tumors. They are associated to poor prognosis, and screening for these variants may help clinicians to identify patients that will not benefit from standard endocrine therapies.

- Expression levels of *ESR2* are generally low in breast cancer, but higher expression is associated to improved outcomes and increased gene expression signatures of immune cell activity.

Concluding remarks

> "… Humans need to keep exploring the world around us through open-ended scientific research. The wonders of penicillin would never have been discovered had Alexander Fleming not been conducting simple experiments with Staphylococci bacteria. Recombinant DNA research—the foundation for modern molecular biology—became possible only with the isolation of DNA-cutting and DNA-copying enzymes from gut- and heat-loving bacteria. Rapid DNA sequencing required experiments on the remarkable properties of bacteria from hot springs. And my colleagues and I would never have created a powerful gene-editing tool if we hadn't tackled the much more fundamental question of how bacteria fight off viral infections."
>
> Jennifer A. Doudna in *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution* (2017)

By applying state-of-the-art technologies and bioinformatic techniques, we have described diverse genomic and transcriptomic changes that may serve to shed light on tumor biology and potentially be used for refined diagnostics and monitoring of disease progression. Since our work on genomic rearrangements in breast tumors was first published, the mechanisms behind large genomic damage events have been further elaborated upon. As discussed in the introduction, we now know more about the different rearrangement profiles that may be found in tumor cells and the putative mechanisms behind them. By characterizing patterns of DNA damage, it could become possible to identify tumors that are susceptible to different types of therapies, analogous to the synthetic lethality caused by PARP inhibitors in *BRCA*-mutated tumors. We showed in our work that the majority of rearrangements in clonal tumors are early events in the development of the tumor and rearrangements may serve as relatively stable biomarkers for sensitive detection of minimal residual disease in circulating tumor DNA from patient blood plasma [301].

Targeted therapies in cancer medicine often have a limited effect, sometimes only extending the median time to progression by a few months, and development of resistance is a frequent problem. This means that there is frequently hesitation about adopting new treatments into clinical practice; the benefit for the patient must be weighed against economical cost and the risk for adverse events. On the

other hand, each new cancer drug is another tool in the arsenal against tumor progression. In the future, sequence analysis may assist in stratifying patients to inform therapy decisions and increase our understanding of the dynamics of resistance development. That the SCAN-B project has RNA-sequenced, in real-time, more than 11,000 tumors to date demonstrates the feasibility of implementing sequencing technologies in clinical diagnostic routines. Gene expression analysis for select cases, as well as mutational analysis for certain cancers are already a clinical reality.

Genomics and transcriptomics have the potential to contribute both to general prognostics and to identifying tumor- and patient-specific drivers that could be targeted with pharmacological intervention. RNA-seq offers the dual opportunity of both performing gene expression profiling and identifying coding genomic changes, as we have shown in this and related works [234], which is beneficial in breast cancer, where both are of clinical interest. Whole genome- and/or transcriptome profiling is not yet widely utilized in the clinical setting. The work performed in this thesis and by others will hopefully contribute to elucidating which analyses could be valuable for precision medicine in cancer care. However, sequencing data requires specialized expertise for bioinformatic analysis and interpretation that is not available at all clinical centers. For the implementation of these techniques, standardized laboratory and bioinformatic protocols need to be developed that can produce easily interpretable data that is generalizable across countries and clinics.

As a final note, I would like to argue for the importance of continued advancements in our understanding of fundamental biology. Clinical research and translational research, wherein clinicians and scientist cooperate to bring biological knowledge into the clinic and *vice versa*, are essential for developing future practices in medicine. Basic research is at the other end of the spectrum, investigating the molecular and biological processes that govern the fate of cells, but is just as vital for medical research as a whole, particularly for generating sound hypotheses and developing novel techniques. Many important discoveries begin with the simple curiosity for the mechanisms of life, and although the clinical utility of a line of research may not be immediately apparent, sometimes the most clinically removed discoveries can have profound medical implications.

Acknowledgments

This work could never have been achieved without the help and assistance of many people; mentors, collaborators, colleagues, friends, and family. I am truly grateful for your work and support throughout my thesis projects.

My thanks go out especially to the following people:

All the **patients** that have participated in the studies, without whom we would never have been able to accomplish this research. Even during what may have been the most trying time in your lives, you have chosen to contribute to advancing our understanding of cancer, to the benefit of current and future patients.

My former main supervisor, **Sofia Gruvberger-Saal**, for believing in me and for introducing me to the field of pre-clinical cancer research. Thank you for all your encouragement and many great learning experiences and inspiring discussions!

Lao Saal, my current main supervisor and mentor, for working with me to reach the finish line. Thank you for all the constructive feedback, assistance, and learning opportunities, and for the great flexibility to accommodate my other endeavors.

Jillian Howlin, my assistant supervisor, for providing guidance and inspiration. Not to mention the much-needed coffee breaks and animated discussions, both about research-oriented and not-so-research-oriented things.

Åke Borg, my assistant supervisor and initiator of SCAN-B; and of course, everyone at the **SCAN-B lab**, **bioinformatics facilities**, and associated **clinical centers**, without whose work studies III and IV would not have been possible.

All my colleagues in and around the **Translational Oncogenomics Unit; Christian Brueffer**, for our project collaborations, helpful discussions and comments as well as bioinformatic scripts and tricks; **Anthony "Tony" George** and **Yilun "Alan" Chen** for contributing with your expertise in the mutations and PCR areas (among other things) and for being first-rate office and lab "roomies"; **Heena Saini**, for our great teamwork on the ERβ project; **Eleonor Olsson**, for helpful feedback and camaraderie in the research as well as medical student settings; and to **Sergii Gladchuck**, **Man-Hung Tang**, **Robert Rigo**, **Christof Winter, Tamara Tjitrowirjo,** and **all our excellent collaborators and students**. It has been a privilege to work with you!

Gabriella Honeth for your constructive comments and regular pep-talks.

Pär-Ola Bendahl, for your amazingly comprehensible explanations and sound advice about statistics-related matters.

Ralph Schultz and **Jeanette Valcich** for always being friendly and helping out with equipment and other lab-related issues, especially when I was a newcomer in the lab.

Kristina Lövgren and **Anna Ebbesson** for technical advice and assistance.

Susanne André and **Björn Frostner** for support in all things administrative.

John och Augusta Perssons stiftelse for funding my conference travel.

All my **co-authors** for your essential work, input and feedback.

The **"Friday breakfast club"** at MV, providing a welcome break with tasty sandwiches and pleasant conversations.

I also want to thank my family. My dad, **Staffan** for being my rock and teaching me, through example, to relax once in a while. My mom, **Åsa** – you inspired me to pursue science and cheered for me every step of the way. My siblings **Erik** and **Sara**, my closest and most loyal friends, standing by me through any and all challenges. Our new additions, **Mina** and **Mio**, mostly for cheering me up but also for giving me perspective, both of which are useful in stressful times.

Last but definitely not least, thank you **Marcus**, the love of my life, for always supporting me in all my more or less crazy endeavors, and for periodically convincing me that most things will work out eventually. Det löser sig.

References

1. Mackay AL: *A Dictionary of Scientific Quotations*. Bristol: Institute of Physics Publishing; 1991.
2. Socialstyrelsen. *Cancer i siffror 2018*. https://static-files.cancerfonden.se/NY_Cancer%20i%20siffror%202018.pdf.
3. World Health Organization. *Cancer*. https://www.who.int/health-topics/cancer (2021; date last accessed: September 13, 2021).
4. World Health Organization. *Cancer fact sheets: All cancers*. https://gco.iarc.fr/today/fact-sheets-cancers (December 2020; date last accessed: September 13, 2021).
5. Hanahan D, Weinberg RA: The hallmarks of cancer. *Cell* 2000, **100**(1):57-70.
6. Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation. *Cell* 2011, **144**(5):646-674.
7. Garber JE, Offit K: Hereditary Cancer Predisposition Syndromes. *J Clin Oncol* 2005, **23**(2):276-292.
8. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW: Cancer genome landscapes. *Science* 2013, **339**(6127):1546-1558.
9. Tomasetti C, Vogelstein B: Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 2015, **347**(6217):78-81.
10. Tomasetti C, Li L, Vogelstein B: Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* 2017, **355**(6331):1330-1334.
11. Kimsey IJ, Petzold K, Sathyamoorthy B, Stein ZW, Al-Hashimi HM: Visualizing transient Watson–Crick-like mispairs in DNA and RNA duplexes. *Nature* 2015, **519**(7543):315-320.
12. Goldsby RE, Lawrence NA, Hays LE, Olmsted EA, Chen X, Singh M, Preston BD: Defective DNA polymerase-δ proofreading causes cancer susceptibility in mice. *Nat Med* 2001, **7**(6):638-639.
13. Palles C, Cazier J-B, Howarth KM, Domingo E, Jones AM, Broderick P, Kemp Z, Spain SL, Guarino E, Salguero I *et al*: Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2013, **45**(2):136-144.
14. Chatterjee N, Walker GC: Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* 2017, **58**(5):235-263.
15. Waris G, Ahsan H: Reactive oxygen species: role in the development of cancer and various chronic conditions. *J Carcinog* 2006, **5**:14-14.

16. Lindahl T: Instability and decay of the primary structure of DNA. *Nature* 1993, **362**(6422):709-715.
17. Weinberg RA: Chapter 12: Maintenance of genomic integrity and the development of cancer. In: *The biology of cancer.* 2 edn.: Garland Science; 2013: 511-575.
18. Li X, Heyer W-D: Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res* 2008, **18**(1):99-113.
19. Chang HHY, Pannunzio NR, Adachi N, Lieber MR: Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* 2017, **18**(8):495-506.
20. Loeb LA: A Mutator Phenotype in Cancer. *Cancer Res* 2001, **61**(8):3230-3239.
21. Niersch J, Vega-Rubín-de-Celis S, Bazarna A, Mergener S, Jendrossek V, Siveke JT, Peña-Llopis S: A BAP1 synonymous mutation results in exon skipping, loss of function and worse patient prognosis. *iScience* 2021, **24**(3).
22. Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B: Synonymous mutations frequently act as driver mutations in human cancers. *Cell* 2014, **156**(6):1324-1335.
23. Lindeboom RGH, Supek F, Lehner B: The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet* 2016, **48**(10):1112-1118.
24. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA *et al*: Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012, **149**(5):979-993.
25. Burns MB, Temiz NA, Harris RS: Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* 2013, **45**(9):977-983.
26. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G *et al*: An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 2013, **45**(9):970-976.
27. Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, Policicchio E, Urso ED, Agostini M, Genuardi M *et al*: A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. *EBioMedicine* 2017, **20**:39-49.
28. Alkan C, Coe BP, Eichler EE: Genome structural variation discovery and genotyping. *Nat Rev Genet* 2011, **12**(5):363-376.
29. Rowley JD: A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature* 1973, **243**(5405):290-293.
30. Shtivelman E, Lifshitz B, Gale RP, Canaani E: Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* 1985, **315**(6020):550-554.
31. Davis RL, Konopka JB, Witte ON: Activation of the c-abl oncogene by viral transduction or chromosomal translocation generates altered c-abl proteins with similar in vitro kinase properties. *Mol Cell Biol* 1985, **5**(1):204-213.
32. Fujiwara T, Bandi M, Nitta M, Ivanova EV, Bronson RT, Pellman D: Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* 2005, **437**(7061):1043-1047.

33. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhsng CZ, Wala J, Mermel CH *et al*: Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013, **45**(10):1134-1140.
34. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA *et al*: Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 2011, **144**(1):27-40.
35. McClintock B: The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics* 1941, **26**(2):234-282.
36. Gisselsson D, Jonson T, Petersén A, Strömbeck B, Dal Cin P, Höglund M, Mitelman F, Mertens F, Mandahl N: Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proc Natl Acad Sci U S A* 2001, **98**(22):12683-12688.
37. O'Hagan RC, Chang S, Maser RS, Mohan R, Artandi SE, Chin L, DePinho RA: Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell* 2002, **2**(2):149-155.
38. Maher Christopher A, Wilson Richard K: Chromothripsis and Human Disease: Piecing Together the Shattering Process. *Cell* 2012, **148**(1):29-32.
39. Tubio JMC, Estivill X: When catastrophe strikes a cell. *Nature* 2011, **470**(7335):476-477.
40. Cortés-Ciriano I, Lee JJ, Xi R, Jain D, Jung YL, Yang L, Gordenin D, Klimczak LJ, Zhang CZ, Pellman DS *et al*: Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* 2020, **52**(3):331-341.
41. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium: Pan-cancer analysis of whole genomes. *Nature* 2020, **578**(7793):82-93.
42. Liu P, Erez A, Nagamani SC, Dhar SU, Kołodziejska KE, Dharmadhikari AV, Cooper ML, Wiszniewska J, Zhang F, Withers MA *et al*: Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* 2011, **146**(6):889-903.
43. Baca Sylvan C, Prandi D, Lawrence Michael S, Mosquera Juan M, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald Theresa Y, Ghandi M *et al*: Punctuated Evolution of Prostate Cancer Genomes. *Cell* 2013, **153**(3):666-677.
44. Pellestor F: Chromoanagenesis: cataclysms behind complex chromosomal rearrangements. *Mol Cytogenet* 2019, **12**:6.
45. Del Monte U: Does the cell number 10(9) still really fit one gram of tumor tissue? *Cell Cycle* 2009, **8**(3):505-506.
46. Stratton MR, Campbell PJ, Futreal PA: The cancer genome. *Nature* 2009, **458**(7239):719-724.
47. Sugihara E, Saya H: Complexity of cancer stem cells. *Int J Cancer* 2013, **132**(6):1249-1259.
48. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, Minden M, Paterson B, Caligiuri MA, Dick JE: A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 1994, **367**(6464):645-648.

49. Yang YC, Wang SW, Hung HY, Chang CC, Wu IC, Huang YL, Lin TM, Tsai JL, Chen A, Kuo FC *et al*: Isolation and characterization of human gastric cell lines with stem cell phenotypes. *J Gastroenterol Hepatol* 2007, **22**(9):1460-1468.
50. Schatton T, Murphy GF, Frank NY, Yamaura K, Waaga-Gasser AM, Gasser M, Zhan Q, Jordan S, Duncan LM, Weishaupt C *et al*: Identification of cells initiating human melanomas. *Nature* 2008, **451**(7176):345-349.
51. Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, Hoey T, Gurney A, Huang EH, Simeone DM *et al*: Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci U S A* 2007, **104**(24):10158-10163.
52. Chen X, Li Q, Liu X, Liu C, Liu R, Rycaj K, Zhang D, Liu B, Jeter C, Calhoun-Davis T *et al*: Defining a Population of Stem-like Human Prostate Cancer Cells That Can Generate and Propagate Castration-Resistant Prostate Cancer. *Clin Cancer Res* 2016, **22**(17):4505-4516.
53. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF: Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* 2003, **100**(7):3983-3988.
54. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM: Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 2010, **12**(5):R68.
55. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S *et al*: Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 2007, **8**(5):R76.
56. Paget S: The Distribution of Secondary Growths in Cancer of the Breast. *Lancet* 1889, **133**(3421):571-573.
57. 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.
58. Turajlic S, Swanton C: Metastasis as an evolutionary process. *Science* 2016, **352**(6282):169.
59. Welch DR, Hurst DR: Defining the Hallmarks of Metastasis. *Cancer Res* 2019, **79**(12):3011-3027.
60. Weinberg RA: Chapter 14: Moving out: Invasion and metastasis. In: *The biology of cancer.* 1 edn.: Garland Science; 2007: 587-654.
61. Coleman RE: Clinical Features of Metastatic Bone Disease and Risk of Skeletal Morbidity. *Clin Cancer Res* 2006, **12**(20):6243s-6249s.
62. Patel JK, Didolkar MS, Pickren JW, Moore RH: Metastatic pattern of malignant melanoma: A study of 216 autopsy cases. *Am J Surg* 1978, **135**(6):807-810.
63. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, Aas T, Alexandrov LB, Larsimont D, Davies H *et al*: Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* 2015, **21**(7):751-759.

64. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M *et al*: The life history of 21 breast cancers. *Cell* 2012, **149**(5):994-1007.

65. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A *et al*: Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012, **366**(10):883-892.

66. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA *et al*: Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 2010, **467**(7319):1114-1117.

67. Green MR, Gentles AJ, Nair RV, Irish JM, Kihira S, Liu CL, Kela I, Hopmans ES, Myklebust JH, Ji H *et al*: Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood* 2013, **121**(9):1604-1611.

68. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, Seth S, Chow CW, Cao Y, Gumbs C *et al*: Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science (New York, NY)* 2014, **346**(6206):256-259.

69. Harbst K, Lauss M, Cirenajwis H, Isaksson K, Rosengren F, Törngren T, Kvist A, Johansson MC, Vallon-Christersson J, Baldetorp B *et al*: Multiregion Whole-Exome Sequencing Uncovers the Genetic Evolution and Mutational Heterogeneity of Early-Stage Metastatic Melanoma. *Cancer Res* 2016, **76**(16):4765-4774.

70. Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, Kremeyer B, Butler A, Lynch AG, Camacho N *et al*: Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet* 2015, **47**(4):367-372.

71. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H *et al*: Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 2014, **512**(7513):155-160.

72. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D *et al*: Tumour evolution inferred by single-cell sequencing. *Nature* 2011, **472**(7341):90-94.

73. Cortot AB, Italiano A, Burel-Vandenbos F, Martel-Planche G, Hainaut P: KRAS mutation status in primary nonsmall cell lung cancer and matched metastases. *Cancer* 2010, **116**(11):2682-2687.

74. Brannon AR, Vakiani E, Sylvester BE, Scott SN, McDermott G, Shah RH, Kania K, Viale A, Oschwald DM, Vacic V *et al*: Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biol* 2014, **15**(8):454.

75. Bhullar DS, Barriuso J, Mullamitha S, Saunders MP, O'Dwyer ST, Aziz O: Biomarker concordance between primary colorectal cancer and its metastases. *EBioMedicine* 2019, **40**:363-374.

76. Meric-Bernstam F, Frampton GM, Ferrer-Lozano J, Yelensky R, Pérez-Fidalgo JA, Wang Y, Palmer GA, Ross JS, Miller VA, Su X *et al*: Concordance of genomic alterations between primary and recurrent breast cancer. *Mol Cancer Ther* 2014, **13**(5):1382-1389.

77. Dupont Jensen J, Laenkholm AV, Knoop A, Ewertz M, Bandaru R, Liu W, Hackl W, Barrett JC, Gardner H: PIK3CA mutations may be discordant between primary and corresponding metastatic disease in breast cancer. *Clin Cancer Res* 2011, **17**(4):667-677.

78. Grellety T, Lucchesi C, Hostein I, Auzanneau C, Khalifa E, Soubeyran I, Italiano A: High-depth sequencing of paired primary and metastatic tumours: Implications for personalised medicine. *Eur J Cancer* 2017, **84**:250-256.

79. Goswami RS, Patel KP, Singh RR, Meric-Bernstam F, Kopetz ES, Subbiah V, Alvarez RH, Davies MA, Jabbar KJ, Roy-Chowdhuri S *et al*: Hotspot mutation panel testing reveals clonal evolution in a study of 265 paired primary and metastatic tumors. *Clin Cancer Res* 2015, **21**(11):2644-2651.

80. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, Duyvesteyn K, Haidari S, van Hoeck A, Onstenk W *et al*: Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* 2019, **575**(7781):210-216.

81. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J *et al*: Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 2009, **461**(7265):809-813.

82. Ng CKY, Bidard FC, Piscuoglio S, Geyer FC, Lim RS, de Bruijn I, Shen R, Pareja F, Berman SH, Wang L *et al*: Genetic Heterogeneity in Therapy-Naïve Synchronous Primary Breast Cancers and Their Metastases. *Clin Cancer Res* 2017, **23**(15):4402-4415.

83. Lee SY, Haq F, Kim D, Jun C, Jo H-J, Ahn S-M, Lee W-S: Comparative Genomic Analysis of Primary and Synchronous Metastatic Colorectal Cancers. *PLoS One* 2014, **9**(3):e90459.

84. Ishaque N, Abba ML, Hauser C, Patil N, Paramasivam N, Huebschmann D, Leupold JH, Balasubramanian GP, Kleinheinz K, Toprak UH *et al*: Whole genome sequencing puts forward hypotheses on metastasis evolution and therapy in colorectal cancer. *Nat Commun* 2018, **9**(1):4782.

85. Makohon-Moore AP, Zhang M, Reiter JG, Bozic I, Allen B, Kundu D, Chatterjee K, Wong F, Jiao Y, Kohutek ZA *et al*: Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat Genet* 2017, **49**(3):358-366.

86. Toy W, Shen Y, Won H, Green B, Sakr RA, Will M, Li Z, Gala K, Fanning S, King TA *et al*: ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* 2013, **45**(12):1439-1445.

87. Jeselsohn R, Yelensky R, Buchwalter G, Frampton G, Meric-Bernstam F, Gonzalez-Angulo AM, Ferrer-Lozano J, Perez-Fidalgo JA, Cristofanilli M, Gomez H *et al*: Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res* 2014, **20**(7):1757-1767.

88. Diaz LA, Jr., Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, Allen B, Bozic I, Reiter JG, Nowak MA *et al*: The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 2012, **486**(7404):537-540.
89. Bozic I, Nowak MA: Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. *Proc Natl Acad Sci U S A* 2014, **111**(45):15964-15968.
90. Luebker SA, Koepsell SA: Diverse Mechanisms of BRAF Inhibitor Resistance in Melanoma Identified in Clinical and Preclinical Studies. *Front Oncol* 2019, **9**(268).
91. Wagle N, Emery C, Berger MF, Davis MJ, Sawyer A, Pochanard P, Kehoe SM, Johannessen CM, Macconaill LE, Hahn WC *et al*: Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol* 2011, **29**(22):3085-3096.
92. Zhao B, Wang L, Qiu H, Zhang M, Sun L, Peng P, Yu Q, Yuan X: Mechanisms of resistance to anti-EGFR therapy in colorectal cancer. *Oncotarget* 2017, **8**(3):3980-4000.
93. 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**:153-179.
94. Ross MH, Pawlina W: Histology: A text and atlas, 7 edn. Philadelphia, US: Lippincott Williams and Wilkins; 2015.
95. Figueroa JD, Pfeiffer RM, Patel DA, Linville L, Brinton LA, Gierach GL, Yang XR, Papathomas D, Visscher D, Mies C *et al*: Terminal duct lobular unit involution of the normal breast: implications for breast cancer etiology. *J Natl Cancer Inst* 2014, **106**(10):dju286.
96. Simpson ER: Sources of estrogen and their importance. *J Steroid Biochem Mol Biol* 2003, **86**(3):225-230.
97. Boron WF, Boulpaep EL (eds.): Medical physiology, Chapter 53, 56 pp 1090, 1146-1149,, 3 edn: Elsevier; 2017.
98. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021, **71**(3):209-249.
99. Socialstyrelsen: Statistik om nyupptäckta cancerfall 2019. In.; 2020
100. American Cancer Society: Breast cancer facts and figures 2019-2020. In. Atlanta: American Cancer Society, Inc.; 2019.
101. WHO Classification of Tumors Editorial Board: Breast tumors. Lyon (France): International Agency for Research on Cancer; 2019.
102. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018, **68**(6):394-424.

103. Antoniou A, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg Å *et al*: Average Risks of Breast and Ovarian Cancer Associated with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies. *Am J Hum Genet* 2003, **72**(5):1117-1130.
104. Chen S, Parmigiani G: Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol* 2007, **25**(11):1329-1333.
105. Mavaddat N, Peock S, Frost D, Ellis S, Platte R, Fineberg E, Evans DG, Izatt L, Eeles RA, Adlard J *et al*: Cancer Risks for BRCA1 and BRCA2 Mutation Carriers: Results From Prospective Analysis of EMBRACE. *J Natl Cancer Inst* 2013, **105**(11):812-822.
106. Shiovitz S, Korde LA: Genetics of breast cancer: a topic in evolution. *Ann Oncol* 2015, **26**(7):1291-1299.
107. van den Brandt PA, Ziegler RG, Wang M, Hou T, Li R, Adami HO, Agnoli C, Bernstein L, Buring JE, Chen Y *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.
108. Bjørge T, Häggström C, Ghaderi S, Nagel G, Manjer J, Tretli S, Ulmer H, Harlid S, Rosendahl AH, Lang A *et al*: BMI and weight changes and risk of obesity-related cancers: a pooled European cohort study. *Int J Epidemiol* 2019, **48**(6):1872-1885.
109. Boyle P, Boniol M, Koechlin A, Robertson C, Valentini F, Coppens K, Fairley LL, Boniol M, Zheng T, Zhang Y *et al*: Diabetes and breast cancer risk: a meta-analysis. *British Journal of Cancer* 2012, **107**(9):1608-1617.
110. Jones ME, Schoemaker MJ, Wright LB, Ashworth A, Swerdlow AJ: Smoking and risk of breast cancer in the Generations Study cohort. *Breast Cancer Res* 2017, **19**(1):118-118.
111. Romieu I, Scoccianti C, Chajès V, de Batlle J, Biessy C, Dossus L, Baglietto L, Clavel-Chapelon F, Overvad K, Olsen A *et al*: Alcohol intake and breast cancer in the European prospective investigation into cancer and nutrition. *Int J Cancer* 2015, **137**(8):1921-1930.
112. Opdahl S, Alsaker MD, Janszky I, Romundstad PR, Vatten LJ: Joint effects of nulliparity and other breast cancer risk factors. *Br J Cancer* 2011, **105**(5):731-736.
113. Kaaks R, Rinaldi S, Key TJ, Berrino F, Peeters PH, Biessy C, Dossus L, Lukanova A, Bingham S, Khaw KT *et al*: Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocr Relat Cancer* 2005, **12**(4):1071-1082.
114. James RE, Lukanova A, Dossus L, Becker S, Rinaldi S, Tjønneland A, Olsen A, Overvad K, Mesrine S, Engel P *et al*: Postmenopausal Serum Sex Steroids and Risk of Hormone Receptor–Positive and -Negative Breast Cancer: a Nested Case–Control Study. *Cancer Prev Res (Phila)* 2011, **4**(10):1626-1635.
115. Jensen EV, Desombre ER, Kawashima T, Suzuki T, Kyser K, Jungblut PW: Estrogen-binding substances of target tissues. *Science (New York, NY)* 1967, **158**(3800):529-530.

116. Jensen EV, Suzuki T, Kawashima T, Stumpf WE, Jungblut PW, DeSombre ER: A two-step mechanism for the interaction of estradiol with rat uterus. *Proc Natl Acad Sci U S A* 1968, **59**(2):632-638.

117. Gosden JR, Middleton PG, Rout D: Localization of the human oestrogen receptor gene to chromosome 6q24----q27 by in situ hybridization. *Cytogenet Cell Genet* 1986, **43**(3-4):218-220.

118. Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA: Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci U S A* 1996, **93**(12):5925-5930.

119. Enmark E, Pelto-Huikko M, Grandien K, Lagercrantz S, Lagercrantz J, Fried G, Nordenskjöld M, Gustafsson JA: Human estrogen receptor beta-gene structure, chromosomal localization, and expression pattern. *J Clin Endocrinol Metab* 1997, **82**(12):4258-4265.

120. Ogawa S, Inoue S, Watanabe T, Hiroi H, Orimo A, Hosoi T, Ouchi Y, Muramatsu M: The Complete Primary Structure of Human Estrogen Receptor β (hERβ) and Its Heterodimerization with ER α in Vivo and in Vitro. *Biochem Biophys Res Commun* 1998, **243**(1):122-126.

121. Taylor SE, Martin-Hirsch PL, Martin FL: Oestrogen receptor splice variants in the pathogenesis of disease. *Cancer Letters* 2010, **288**(2):133-148.

122. Carmeci C, Thompson DA, Ring HZ, Francke U, Weigel RJ: Identification of a gene (GPR30) with homology to the G-protein-coupled receptor superfamily associated with estrogen receptor expression in breast cancer. *Genomics* 1997, **45**(3):607-617.

123. Revankar CM, Cimino DF, Sklar LA, Arterburn JB, Prossnitz ER: A transmembrane intracellular estrogen receptor mediates rapid cell signaling. *Science* 2005, **307**(5715):1625-1630.

124. Thomas P, Pang Y, Filardo EJ, Dong J: Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. *Endocrinology* 2005, **146**(2):624-632.

125. Filardo EJ, Graeber CT, Quinn JA, Resnick MB, Giri D, DeLellis RA, Steinhoff MM, Sabo E: Distribution of GPR30, a seven membrane-spanning estrogen receptor, in primary breast cancer and its association with clinicopathologic determinants of tumor progression. *Clin Cancer Res* 2006, **12**(21):6359-6366.

126. Samartzis EP, Noske A, Meisel A, Varga Z, Fink D, Imesch P: The G protein-coupled estrogen receptor (GPER) is expressed in two different subcellular localizations reflecting distinct tumor properties in breast cancer. *PLoS One* 2014, **9**(1):e83296.

127. Sjöström M, Hartman L, Grabau D, Fornander T, Malmström P, Nordenskjöld B, Sgroi DC, Skoog L, Stål O, Leeb-Lundberg LM *et al*: Lack of G protein-coupled estrogen receptor (GPER) in the plasma membrane is associated with excellent long-term prognosis in breast cancer. *Breast Cancer Res Treat* 2014, **145**(1):61-71.

128. Ignatov A, Ignatov T, Weissenborn C, Eggemann H, Bischoff J, Semczuk A, Roessner A, Costa SD, Kalinski T: G-protein-coupled estrogen receptor GPR30 and tamoxifen resistance in breast cancer. *Breast Cancer Res Treat* 2011, **128**(2):457-466.

129. Mo Z, Liu M, Yang F, Luo H, Li Z, Tu G, Yang G: GPR30 as an initiator of tamoxifen resistance in hormone-dependent breast cancer. *Breast Cancer Res* 2013, **15**(6):R114.
130. Weikum ER, Liu X, Ortlund EA: The nuclear receptor superfamily: A structural perspective. *Protein Sci* 2018, **27**(11):1876-1892.
131. Kumar R, Zakharov MN, Khan SH, Miki R, Jang H, Toraldo G, Singh R, Bhasin S, Jasuja R: The dynamic structure of the estrogen receptor. *J Amino Acids* 2011, **2011**:812540-812540.
132. Cowley SM, Hoare S, Parker MG, Mosselman S: Estrogen Receptors alpha and beta form heterodimers on DNA. *J Biol Chem* 1997, **272**(32):19858-19862.
133. Kumar V, Chambon P: The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. *Cell* 1988, **55**(1):145-156.
134. Charn TH, Liu ET, Chang EC, Lee YK, Katzenellenbogen JA, Katzenellenbogen BS: Genome-wide dynamics of chromatin binding of estrogen receptors alpha and beta: mutual restriction and competitive site selection. *Mol Endocrinol* 2010, **24**(1):47-59.
135. Petz L, Nardulli A: Sp1 binding sites and an estrogen response element half-site are involved in regulation of the human progesterone receptor A promoter. *Mol Endocrinol* 2000, **14**:972-985.
136. Paech K, Webb P, Kuiper GG, Nilsson S, Gustafsson J, Kushner PJ, Scanlan TS: Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites. *Science (New York, NY)* 1997, **277**(5331):1508-1510.
137. Jakacka M, Ito M, Weiss J, Chien PY, Gehm BD, Jameson JL: Estrogen receptor binding to DNA is not required for its activity through the nonclassical AP1 pathway. *J Biol Chem* 2001, **276**(17):13615-13621.
138. Pietras RJ, Arboleda J, Reese DM, Wongvipat N, Pegram MD, Ramos L, Gorman CM, Parker MG, Sliwkowski MX, Slamon DJ: HER-2 tyrosine kinase pathway targets estrogen receptor and promotes hormone-independent growth in human breast cancer cells. *Oncogene* 1995, **10**(12):2435-2446.
139. Kato S, Endoh H, Masuhiro Y, Kitamoto T, Uchiyama S, Sasaki H, Masushige S, Gotoh Y, Nishida E, Kawashima H *et al*: Activation of the Estrogen Receptor Through Phosphorylation by Mitogen-Activated Protein Kinase. *Science (New York, NY)* 1995, **270**(5241):1491-1494.
140. Campbell RA, Bhat-Nakshatri P, Patel NM, Constantinidou D, Ali S, Nakshatri H: Phosphatidylinositol 3-kinase/AKT-mediated activation of estrogen receptor alpha: a new model for anti-estrogen resistance. *J Biol Chem* 2001, **276**(13):9817-9824.
141. Aronica SM, Katzenellenbogen BS: Stimulation of estrogen receptor-mediated transcription and alteration in the phosphorylation state of the rat uterine estrogen receptor by estrogen, cyclic adenosine monophosphate, and insulin-like growth factor-I. *Mol Endocrinol* 1993, **7**(6):743-752.
142. Tremblay A, Tremblay GB, Labrie F, Giguère V: Ligand-independent recruitment of SRC-1 to estrogen receptor beta through phosphorylation of activation function AF-1. *Mol Cell* 1999, **3**(4):513-519.

143. Sanchez M, Sauvé K, Picard N, Tremblay A: The hormonal response of estrogen receptor beta is decreased by the phosphatidylinositol 3-kinase/Akt pathway via a phosphorylation-dependent release of CREB-binding protein. *J Biol Chem* 2007, **282**(7):4830-4840.

144. Lam HM, Suresh Babu CV, Wang J, Yuan Y, Lam YW, Ho SM, Leung YK: Phosphorylation of human estrogen receptor-beta at serine 105 inhibits breast cancer cell migration and invasion. *Mol Cell Endocrinol* 2012, **358**(1):27-35.

145. Rhodes LV, Bratton MR, Zhu Y, Tilghman SL, Muir SE, Salvo VA, Tate CR, Elliott S, Nephew KP, Collins-Burow BM *et al*: Effects of SDF-1-CXCR4 signaling on microRNA expression and tumorigenesis in estrogen receptor-alpha (ER-α)-positive breast cancer cells. *Experimental cell research* 2011, **317**(18):2573-2581.

146. Sauvé K, Lepage J, Sanchez M, Heveker N, Tremblay A: Positive Feedback Activation of Estrogen Receptors by the CXCL12-CXCR4 Pathway. *Cancer Res* 2009, **69**(14):5793-5800.

147. Le Goff P, Montano MM, Schodin DJ, Katzenellenbogen BS: Phosphorylation of the human estrogen receptor. Identification of hormone-regulated sites and examination of their influence on transcriptional activity. *J Biol Chem* 1994, **269**(6):4458-4466.

148. Chen D, Washbrook E, Sarwar N, Bates GJ, Pace PE, Thirunuvakkarasu V, Taylor J, Epstein RJ, Fuller-Pace FV, Egly J-M *et al*: Phosphorylation of human estrogen receptor α at serine 118 by two distinct signal transduction pathways revealed by phosphorylation-specific antisera. *Oncogene* 2002, **21**(32):4921-4931.

149. Rogatsky I, Trowbridge JM, Garabedian MJ: Potentiation of human estrogen receptor alpha transcriptional activation through phosphorylation of serines 104 and 106 by the cyclin A-CDK2 complex. *J Biol Chem* 1999, **274**(32):22296-22302.

150. Carroll JS, Brown M: Estrogen Receptor Target Gene: An Evolving Concept. *Mol Endocrinol* 2006, **20**(8):1707-1714.

151. O'Lone R, Frith MC, Karlsson EK, Hansen U: Genomic Targets of Nuclear Estrogen Receptors. *Mol Endocrinol* 2004, **18**(8):1859-1875.

152. Welboren W-J, Stunnenberg HG, Sweep FCGJ, Span PN: Identifying estrogen receptor target genes. *Mol Oncol* 2007, **1**(2):138-143.

153. Ikeda K, Horie-Inoue K, Inoue S: Identification of estrogen-responsive genes based on the DNA binding properties of estrogen receptors using high-throughput sequencing technology. *Acta Pharmacol Sin* 2015, **36**(1):24-31.

154. Gruvberger S, Ringnér M, Chen Y, Panavally S, Saal LH, Borg A, Fernö M, Peterson C, Meltzer PS: Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 2001, **61**(16):5979-5984.

155. Charpentier AH, Bednarek AK, Daniel RL, Hawkins KA, Laflin KJ, Gaddis S, MacLeod MC, Aldaz CM: Effects of estrogen on global gene expression: identification of novel targets of estrogen action. *Cancer Res* 2000, **60**(21):5977-5983.

156. Leygue E, Gol-Winkler R, Gompel A, Louis-Sylvestre C, Soquet L, Staub S, Kuttenn F, Mauvais-Jarvis P: Estradiol stimulates c-myc proto-oncogene expression in normal human breast epithelial cells in culture. *J Steroid Biochem Mol Biol* 1995, **52**(4):299-305.

157. Lewis JS, Thomas TJ, Klinge CM, Gallo MA, Thomas T: Regulation of cell cycle and cyclins by 16alpha-hydroxyestrone in MCF-7 breast cancer cells. *J Mol Endocrinol* 2001, **27**(3):293-307.

158. Altucci L, Addeo R, Cicatiello L, Dauvois S, Parker MG, Truss M, Beato M, Sica V, Bresciani F, Weisz A: 17beta-Estradiol induces cyclin D1 gene transcription, p36D1-p34cdk4 complex activation and p105Rb phosphorylation during mitogenic stimulation of G(1)-arrested human breast cancer cells. *Oncogene* 1996, **12**(11):2315-2324.

159. Paruthiyil S, Parmar H, Kerekatte V, Cunha GR, Firestone GL, Leitman DC: Estrogen receptor beta inhibits human breast cancer cell proliferation and tumor formation by causing a G2 cell cycle arrest. *Cancer Res* 2004, **64**(1):423-428.

160. Matthews J, Wihlén Br, Tujague M, Wan J, Ström A, Gustafsson J-Ak: Estrogen Receptor (ER) β Modulates ERα-Mediated Transcriptional Activation by Altering the Recruitment of c-Fos and c-Jun to Estrogen-Responsive Promoters. *Mol Endocrinol* 2006, **20**(3):534-543.

161. Williams C, Edvardsson K, Lewandowski SA, Ström A, Gustafsson JÅ: A genome-wide study of the repressive effects of estrogen receptor beta on estrogen receptor alpha signaling in breast cancer cells. *Oncogene* 2008, **27**(7):1019-1032.

162. Clarke RB, Howell A, Potten CS, Anderson E: Dissociation between Steroid Receptor Expression and Cell Proliferation in the Human Breast. *Cancer Res* 1997, **57**(22):4987-4991.

163. Russo J, Ao X, Grill C, Russo IH: Pattern of distribution of cells positive for estrogen receptor alpha and progesterone receptor in relation to proliferating cells in the mammary gland. *Breast Cancer Res Treat* 1999, **53**(3):217-227.

164. Mallepell S, Krust A, Chambon P, Brisken C: Paracrine signaling through the epithelial estrogen receptor α is required for proliferation and morphogenesis in the mammary gland. *Proc Natl Acad Sci U S A* 2006, **103**(7):2196-2201.

165. Förster C, Mäkela S, Wärri A, Kietz S, Becker D, Hultenby K, Warner M, Gustafsson J-Å: Involvement of estrogen receptor β in terminal differentiation of mammary gland epithelium. *Proc Natl Acad Sci U S A* 2002, **99**(24):15578-15583.

166. Li S, Han B, Liu G, Li S, Ouellet J, Labrie F, Pelletier G: Immunocytochemical localization of sex steroid hormone receptors in normal human mammary gland. *J Histochem Cytochem* 2010, **58**(6):509-515.

167. Speirs V, Skliris GP, Burdall SE, Carder PJ: Distinct expression patterns of ER alpha and ER beta in normal human mammary gland. *J Clin Pathol* 2002, **55**(5):371-374.

168. Shaw JA, Udokang K, Mosquera JM, Chauhan H, Jones JL, Walker RA: Oestrogen receptors alpha and beta differ in normal human breast and breast carcinomas. *J Pathol* 2002, **198**(4):450-457.

169. Palmieri C, Saji S, Sakaguchi H, Cheng G, Sunters A, O'Hare MJ, Warner M, Gustafsson JA, Coombes RC, Lam EW: The expression of oestrogen receptor (ER)-beta and its variants, but not ERalpha, in adult human mammary fibroblasts. *J Mol Endocrinol* 2004, **33**(1):35-50.

170. Hawse JR, Carter JM, Aspros KGM, Bruinsma ES, Koepplin JW, Negron V, Subramaniam M, Ingle JN, Rech KL, Goetz MP: Optimized immunohistochemical detection of estrogen receptor beta using two validated monoclonal antibodies confirms its expression in normal and malignant breast tissues. *Breast Cancer Res Treat* 2020, **179**(1):241-249.
171. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001, **98**(19):10869-10874.
172. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: Molecular portraits of human breast tumours. *Nature* 2000, **406**(6797):747-752.
173. Skliris GP, Carder PJ, Lansdown MR, Speirs V: Immunohistochemical detection of ERbeta in breast cancer: towards more detailed receptor profiling? *Br J Cancer* 2001, **84**(8):1095-1098.
174. Fuqua SA, Schiff R, Parra I, Moore JT, Mohsin SK, Osborne CK, Clark GM, Allred DC: Estrogen receptor beta protein in human breast cancer: correlation with clinical tumor parameters. *Cancer Res* 2003, **63**(10):2434-2439.
175. Marotti JD, Collins LC, Hu R, Tamimi RM: Estrogen receptor-β expression in invasive breast cancer in relation to molecular phenotype: results from the Nurses' Health Study. *Mod Pathol* 2010, **23**(2):197-204.
176. Novelli F, Milella M, Melucci E, Di Benedetto A, Sperduti I, Perrone-Donnorso R, Perracchio L, Venturo I, Nisticò C, Fabi A *et al*: A divergent role for estrogen receptor-beta in node-positive and node-negative breast cancer classified according to molecular subtypes: an observational prospective study. *Breast Cancer Res* 2008, **10**(5):R74.
177. Oueslati M, Bittaieb I, Sassi N, Jemaa AB, Gamoudi A, Rahal K, Oueslati R: ERα and ERβ co-expression: An indicator of aggressive tumors and hormonal sensitivity. *Oncol Lett* 2017, **14**(2):1675-1682.
178. Reese JM, Suman VJ, Subramaniam M, Wu X, Negron V, Gingery A, Pitel KS, Shah SS, Cunliffe HE, McCullough AE *et al*: ERβ1: characterization, prognosis, and evaluation of treatment strategies in ERα-positive and -negative breast cancer. *BMC Cancer* 2014, **14**(1):749.
179. Andersson S, Sundberg M, Pristovsek N, Ibrahim A, Jonsson P, Katona B, Clausson CM, Zieba A, Ramström M, Söderberg O *et al*: Insufficient antibody validation challenges oestrogen receptor beta research. *Nat Commun* 2017, **8**:15840.
180. Mal R, Magner A, David J, Datta J, Vallabhaneni M, Kassem M, Manouchehri J, Willingham N, Stover D, Vandeusen J *et al*: Estrogen Receptor Beta (ERβ): A Ligand Activated Tumor Suppressor. *Front Oncol* 2020, **10**(2151).
181. Park B-W, Kim K-S, Heo M-K, Ko S-S, Lee KS, Hong SW, Yang W-I, Kim J-H, Kim GE: Expression of Estrogen Receptor-β in Normal Mammary and Tumor Tissues: is it Protective in Breast Carcinogenesis? *Breast Cancer Res Treat* 2003, **80**(1):79-85.

182. Roger P, Sahla ME, Mäkelä S, Gustafsson JA, Baldet P, Rochefort H: Decreased expression of estrogen receptor beta protein in proliferative preinvasive mammary tumors. *Cancer Res* 2001, **61**(6):2537-2541.

183. Iwao K, Miyoshi Y, Egawa C, Ikeda N, Noguchi S: Quantitative analysis of estrogen receptor-beta mRNA and its variants in human breast cancers. *Int J Cancer* 2000, **88**(5):733-736.

184. Sugiura H, Toyama T, Hara Y, Zhang Z, Kobayashi S, Fujii Y, Iwase H, Yamashita H: Expression of estrogen receptor beta wild-type and its variant ERbetacx/beta2 is correlated with better prognosis in breast cancer. *Jpn J Clin Oncol* 2007, **37**(11):820-828.

185. Elebro K, Borgquist S, Rosendahl AH, Markkula A, Simonsson M, Jirström K, Rose C, Ingvar C, Jernström H: High Estrogen Receptor β Expression Is Prognostic among Adjuvant Chemotherapy-Treated Patients-Results from a Population-Based Breast Cancer Cohort. *Clin Cancer Res* 2017, **23**(3):766-777.

186. Maehle BO, Collett K, Tretli S, Akslen LA, Grotmol T: Estrogen receptor beta--an independent prognostic marker in estrogen receptor alpha and progesterone receptor-positive breast cancer? *APMIS* 2009, **117**(9):644-650.

187. Gruvberger-Saal SK, Bendahl PO, Saal LH, Laakso M, Hegardt C, Edén P, Peterson C, Malmström P, Isola J, Borg A *et al*: Estrogen receptor beta expression is associated with tamoxifen response in ERalpha-negative breast carcinoma. *Clin Cancer Res* 2007, **13**(7):1987-1994.

188. Honma N, Horii R, Iwase T, Saji S, Younes M, Takubo K, Matsuura M, Ito Y, Akiyama F, Sakamoto G: Clinical importance of estrogen receptor-beta evaluation in breast cancer patients treated with adjuvant tamoxifen therapy. *J Clin Oncol* 2008, **26**(22):3727-3734.

189. Mann S, Laucirica R, Carlson N, Younes PS, Ali N, Younes A, Li Y, Younes M: Estrogen receptor beta expression in invasive breast cancer. *Hum Pathol* 2001, **32**(1):113-118.

190. Reese JM, Suman VJ, Subramaniam M, Wu X, Negron V, Gingery A, Pitel KS, Shah SS, Cunliffe HE, McCullough AE *et al*: ERβ1: characterization, prognosis, and evaluation of treatment strategies in ERα-positive and -negative breast cancer. *BMC Cancer* 2014, **14**:749.

191. Hamilton N, Márquez-Garbán D, Mah V, Fernando G, Elshimali Y, Garbán H, Elashoff D, Vadgama J, Goodglick L, Pietras R: Biologic roles of estrogen receptor-β and insulin-like growth factor-2 in triple-negative breast cancer. *Biomed Res Int* 2015, **2015**:925703.

192. Austin D, Hamilton N, Elshimali Y, Pietras R, Wu Y, Vadgama J: Estrogen receptor-beta is a potential target for triple negative breast cancer treatment. *Oncotarget* 2018, **9**(74):33912-33930.

193. Kim TJ, Lee A, Choi YJ, Song BJ, Yim HW, Kang CS: Prognostic Significance of High Expression of ER-beta in Surgically Treated ER-Positive Breast Cancer Following Endocrine Therapy. *J Breast Cancer* 2012, **15**(1):79-86.

194. Guo L, Zhang YU, Yilamu D, Liu S, Guo C: ERβ overexpression results in endocrine therapy resistance and poor prognosis in postmenopausal ERα-positive breast cancer patients. *Oncol Lett* 2016, **11**(2):1531-1536.

195. Guo L, Zhang Y, Zhang W, Yilamu D: Correlation between estrogen receptor β expression and the curative effect of endocrine therapy in breast cancer patients. *Exp Ther Med* 2014, **7**(6):1568-1572.

196. Tan W, Li Q, Chen K, Su F, Song E, Gong C: Estrogen receptor beta as a prognostic factor in breast cancer patients: A systematic review and meta-analysis. *Oncotarget* 2016, **7**(9):10373-10385.

197. Zhao L, Huang S, Mei S, Yang Z, Xu L, Zhou N, Yang Q, Shen Q, Wang W, Le X *et al*: Pharmacological activation of estrogen receptor beta augments innate immunity to suppress cancer metastasis. *Proc Natl Acad Sci U S A* 2018, **115**(16):E3673-E3681.

198. Hinsche O, Girgert R, Emons G, Gründker C: Estrogen receptor β selective agonists reduce invasiveness of triple-negative breast cancer cells. *Int J Oncol* 2015, **46**(2):878-884.

199. Veeraraghavan J, Tan Y, Cao XX, Kim JA, Wang X, Chamness GC, Maiti SN, Cooper LJ, Edwards DP, Contreras A *et al*: Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers. *Nat Commun* 2014, **5**:4577.

200. Li S, Shen D, Shao J, Crowder R, Liu W, Prat A, He X, Liu S, Hoog J, Lu C *et al*: Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep* 2013, **4**(6):1116-1130.

201. Lei JT, Shao J, Zhang J, Iglesia M, Chan DW, Cao J, Anurag M, Singh P, He X, Kosaka Y *et al*: Functional Annotation of ESR1 Gene Fusions in Estrogen Receptor-Positive Breast Cancer. *Cell Rep* 2018, **24**(6):1434-1444 e1437.

202. Pakdel F, Reese JC, Katzenellenbogen BS: Identification of charged residues in an N-terminal portion of the hormone-binding domain of the human estrogen receptor important in transcriptional activity of the receptor. *Mol Endocrinol* 1993, **7**(11):1408-1417.

203. Zhang QX, Borg A, Wolf DM, Oesterreich S, Fuqua SA: An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. *Cancer Res* 1997, **57**(7):1244-1249.

204. Robinson DR, Wu YM, Vats P, Su F, Lonigro RJ, Cao X, Kalyana-Sundaram S, Wang R, Ning Y, Hodges L *et al*: Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* 2013, **45**(12):1446-1451.

205. Dustin D, Gu G, Fuqua SAW: ESR1 mutations in breast cancer. *Cancer* 2019.

206. Toy W, Weir H, Razavi P, Lawson M, Goeppert AU, Mazzola AM, Smith A, Wilson J, Morrow C, Wong WL *et al*: Activating ESR1 Mutations Differentially Affect the Efficacy of ER Antagonists. *Cancer discovery* 2017, **7**(3):277-287.

207. Kuang Y, Siddiqui B, Hu J, Pun M, Cornwell M, Buchwalter G, Hughes ME, Wagle N, Kirschmeier P, Janne PA *et al*: Unraveling the clinicopathological features driving the emergence of ESR1 mutations in metastatic breast cancer. *NPJ Breast Cancer* 2018, **4**:22.

208. Razavi P, Dickler MN, Shah PD, Toy W, Brown DN, Won HH, Li BT, Shen R, Vasan N, Modi S *et al*: Alterations in PTEN and ESR1 promote clinical resistance to alpelisib plus aromatase inhibitors. *Nat Cancer* 2020, **1**(4):382-393.
209. Bardia A, Kaklamani V, Wilks S, Weise A, Richards D, Harb W, Osborne C, Wesolowski R, Karuturi M, Conkling P *et al*: Phase I Study of Elacestrant (RAD1901), a Novel Selective Estrogen Receptor Degrader, in ER-Positive, HER2-Negative Advanced Breast Cancer. *J Clin Oncol* 2021, **39**(12):1360-1370.
210. Weir HM, Bradbury RH, Lawson M, Rabow AA, Buttar D, Callis RJ, Curwen JO, de Almeida C, Ballard P, Hulse M *et al*: AZD9496: An Oral Estrogen Receptor Inhibitor That Blocks the Growth of ER-Positive and ESR1-Mutant Breast Tumors in Preclinical Models. *Cancer Res* 2016, **76**(11):3307-3318.
211. Kastner P, Krust A, Turcotte B, Stropp U, Tora L, Gronemeyer H, Chambon P: Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. *EMBO J* 1990, **9**(5):1603-1614.
212. Sartorius CA, Melville MY, Hovland AR, Tung L, Takimoto GS, Horwitz KB: A third transactivation function (AF3) of human progesterone receptors located in the unique N-terminal segment of the B-isoform. *Mol Endocrinol* 1994, **8**(10):1347-1360.
213. Meyer ME, Quirin-Stricker C, Lerouge T, Bocquel MT, Gronemeyer H: A limiting factor mediates the differential activation of promoters by the human progesterone receptor isoforms. *J Biol Chem* 1992, **267**(15):10882-10887.
214. Lim E, Palmieri C, Tilley WD: Renewed interest in the progesterone receptor in breast cancer. *Br J Cancer* 2016, **115**(8):909-911.
215. Kurozumi S, Matsumoto H, Hayashi Y, Tozuka K, Inoue K, Horiguchi J, Takeyoshi I, Oyama T, Kurosumi M: Power of PgR expression as a prognostic factor for ER-positive/HER2-negative breast cancer patients at intermediate risk classified by the Ki67 labeling index. *BMC Cancer* 2017, **17**(1):354.
216. Purdie CA, Quinlan P, Jordan LB, Ashfield A, Ogston S, Dewar JA, Thompson AM: Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. *Br J Cancer* 2014, **110**(3):565-572.
217. Schechter AL, Stern DF, Vaidyanathan L, Decker SJ, Drebin JA, Greene MI, Weinberg RA: The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. *Nature* 1984, **312**(5994):513-516.
218. Bargmann CI, Hung MC, Weinberg RA: The neu oncogene encodes an epidermal growth factor receptor-related protein. *Nature* 1986, **319**(6050):226-230.
219. Weiner DB, Kokai Y, Wada T, Cohen JA, Williams WV, Greene MI: Linkage of tyrosine kinase activity with transforming ability of the p185neu oncoprotein. *Oncogene* 1989, **4**(10):1175-1183.
220. Du Z, Lovly CM: Mechanisms of receptor tyrosine kinase activation in cancer. *Mol Cancer* 2018, **17**(1):58.

221. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001, **98**(19):10869-10874.
222. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003, **100**(14):8418-8423.
223. Balduzzi S, Mantarro S, Guarneri V, Tagliabue L, Pistotti V, Moja L, D'Amico R: Trastuzumab-containing regimens for metastatic breast cancer. *Cochrane Database Syst Rev* 2014, **2014**(6):Cd006242.
224. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L *et al*: The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006, **7**(1):96.
225. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al*: Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009, **27**(8):1160-1167.
226. Prat A, Pineda E, Adamo B, Galván P, Fernández A, Gaba L, Díez M, Viladot M, Arance A, Muñoz M: Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* 2015, **24**:S26-S35.
227. Lane DP: Cancer. p53, guardian of the genome. *Nature* 1992, **358**(6381):15-16.
228. The Cancer Genome Atlas N, Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L *et al*: Comprehensive molecular portraits of human breast tumours. *Nature* 2012, **490**:61.
229. Pereira B, Chin S-F, Rueda OM, Vollan H-KM, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut S-J *et al*: The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* 2016, **7**:11479-11479.
230. Kastenhuber ER, Lowe SW: Putting p53 in Context. *Cell* 2017, **170**(6):1062-1078.
231. Zhu G, Pan C, Bei J-X, Li B, Liang C, Xu Y, Fu X: Mutant p53 in Cancer Progression and Targeted Therapies. *Front Oncol* 2020, **10**(2418).
232. Arafeh R, Samuels Y: PIK3CA in cancer: The past 30 years. *Semin Cancer Biol* 2019, **59**:36-49.
233. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC *et al*: Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016, **534**(7605):47-54.
234. Brueffer C, Gladchuk S, Winter C, Vallon-Christersson J, Hegardt C, Häkkinen J, George AM, Chen Y, Ehinger A, Larsson C *et al*: The mutational landscape of the SCAN-B real-world primary breast cancer transcriptome. *EMBO Mol Med* 2020, **12**(10):e12118.
235. Saal LH, Holm K, Maurer M, Memeo L, Su T, Wang X, Yu JS, Malmstrom PO, Mansukhani M, Enoksson J *et al*: PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. *Cancer Res* 2005, **65**(7):2554-2559.

236. Zardavas D, Te Marvelde L, Milne RL, Fumagalli D, Fountzilas G, Kotoula V, Razis E, Papaxoinis G, Joensuu H, Moynahan ME *et al*: Tumor PIK3CA Genotype and Prognosis in Early-Stage Breast Cancer: A Pooled Analysis of Individual Patient Data. *J Clin Oncol* 2018, **36**(10):981-990.

237. Mosele F, Stefanovska B, Lusque A, Tran Dien A, Garberis I, Droin N, Le Tourneau C, Sablin MP, Lacroix L, Enrico D *et al*: Outcome and molecular landscape of patients with PIK3CA-mutated metastatic breast cancer. *Ann Oncol* 2020, **31**(3):377-386.

238. Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She QB, Maurer M, Koujak S, Ferrando AA, Malmström P, Memeo L *et al*: Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci U S A* 2007, **104**(18):7564-7569.

239. Saal LH, Gruvberger-Saal SK, Persson C, Lövgren K, Jumppanen M, Staaf J, Jönsson G, Pires MM, Maurer M, Holm K *et al*: Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nat Genet* 2008, **40**(1):102-107.

240. García JM, Silva J, Peña C, Garcia V, Rodríguez R, Cruz MA, Cantos B, Provencio M, España P, Bonilla F: Promoter methylation of the PTEN gene is a common molecular change in breast cancer. *Genes Chromosomes Cancer* 2004, **41**(2):117-124.

241. McCabe N, Turner NC, Lord CJ, Kluzek K, Bialkowska A, Swift S, Giavara S, O'Connor MJ, Tutt AN, Zdzienicka MZ *et al*: Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. *Cancer Res* 2006, **66**(16):8109-8115.

242. Chen M-T, Sun H-F, Zhao Y, Fu W-Y, Yang L-P, Gao S-P, Li L-D, Jiang H-l, Jin W: Comparison of patterns and prognosis among distant metastatic breast cancer patients by age groups: a SEER population-based analysis. *Sci Rep* 2017, **7**(1):9254.

243. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JGM, Foekens JA, Martens JWM: Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 2008, **68**(9):3108-3114.

244. Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, Cheang MC, Gelmon K, Nielsen TO, Blomqvist C *et al*: Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med* 2010, **7**(5):e1000279.

245. Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, Peto R, Pritchard KI, Bergh J, Dowsett M *et al*: 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *N Engl J Med* 2017, **377**(19):1836-1846.

246. Mittempergher L, Saghatchian M, Wolf DM, Michiels S, Canisius S, Dessen P, Delaloge S, Lazar V, Benz SC, Tursz T *et al*: A gene signature for late distant metastasis in breast cancer identifies a potential mechanism of late recurrences. *Mol Oncol* 2013, **7**(5):987-999.

247. Kim RS, Avivar-Valderas A, Estrada Y, Bragado P, Sosa MS, Aguirre-Ghiso JA, Segall JE: Dormancy signatures and metastasis in estrogen receptor positive and negative breast cancer. *PLoS One* 2012, **7**(4):e35569.

248. Simmons C, Miller N, Geddie W, Gianfelice D, Oldfield M, Dranitsaris G, Clemons MJ: Does confirmatory tumor biopsy alter the management of breast cancer patients with distant metastases? *Ann Oncol* 2009, **20**(9):1499-1504.

249. Lindström LS, Karlsson E, Wilking UM, Johansson U, Hartman J, Lidbrink EK, Hatschek T, Skoog L, Bergh J: Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J Clin Oncol* 2012, **30**(21):2601-2608.

250. Curtit E, Nerich V, Mansi L, Chaigneau L, Cals L, Villanueva C, Bazan F, Montcuquet P, Meneveau N, Perrin S *et al*: Discordances in estrogen receptor status, progesterone receptor status, and HER2 status between primary breast cancer and metastasis. *Oncologist* 2013, **18**(6):667-674.

251. Aurilio G, Monfardini L, Rizzo S, Sciandivasci A, Preda L, Bagnardi V, Disalvatore D, Pruneri G, Munzone E, Della Vigna P *et al*: Discordant hormone receptor and human epidermal growth factor receptor 2 status in bone metastases compared to primary breast cancer. *Acta Oncol* 2013, **52**(8):1649-1656.

252. Ibrahim T, Farolfi A, Scarpi E, Mercatali L, Medri L, Ricci M, Nanni O, Serra L, Amadori D: Hormonal receptor, human epidermal growth factor receptor-2, and Ki67 discordance between primary breast cancer and paired metastases: clinical impact. *Oncology* 2013, **84**(3):150-157.

253. Aurilio G, Disalvatore D, Pruneri G, Bagnardi V, Viale G, Curigliano G, Adamoli L, Munzone E, Sciandivasci A, De Vita F *et al*: A meta-analysis of oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 discordance between primary breast cancer and metastases. *Eur J Cancer* 2014, **50**(2):277-289.

254. Cardoso F, Senkus E, Costa A, Papadopoulos E, Aapro M, André F, Harbeck N, Aguilar Lopez B, Barrios CH, Bergh J *et al*: 4th ESO–ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4)†. *Ann Oncol* 2018, **29**(8):1634-1657.

255. Vichapat V, Gillett C, Fentiman IS, Tutt A, Holmberg L, Luchtenborg M: Risk factors for metachronous contralateral breast cancer suggest two aetiological pathways. *Eur J Cancer* 2011, **47**(13):1919-1927.

256. Spronk I, Schellevis FG, Burgers JS, de Bock GH, Korevaar JC: Incidence of isolated local breast cancer recurrence and contralateral breast cancer: A systematic review. *Breast* 2018, **39**:70-79.

257. Hartman M, Czene K, Reilly M, Adolfsson J, Bergh J, Adami H-O, Dickman PW, Hall P: Incidence and Prognosis of Synchronous and Metachronous Bilateral Breast Cancer. *J Clin Oncol* 2007, **25**(27):4210-4216.

258. Schaapveld M, Visser O, Louwman WJ, Willemse PHB, de Vries EGE, van der Graaf WTA, Otter R, Coebergh JWW, van Leeuwen FE: The impact of adjuvant therapy on contralateral breast cancer risk and the prognostic significance of contralateral breast cancer: a population based study in the Netherlands. *Breast Cancer Res Treat* 2008, **110**(1):189-197.
259. 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):2739-2750.
260. Regionala Cancercentrum i Samverkan. *Bröstcancer - nationellt vårdprogram 2020* https://www.cancercentrum.se/samverkan/cancerdiagnoser/brost/vardprogram/ (October 22, 2020; date last accessed: September 10, 2021).
261. Nationellt kvalitetsregister för bröstcancer. *Årsrapport 2020*. https://statistik.incanet.se/brostcancer/ (April 5, 2021; date last accessed: July 13, 2021).
262. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, Zackrisson S, Senkus E: Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2019.
263. Weigelt B, Geyer FC, Reis-Filho JS: Histological types of breast cancer: how special are they? *Mol Oncol* 2010, **4**(3):192-208.
264. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, Fox SB, Ichihara S, Jacquemier J, Lakhani SR *et al*: Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* 2010, **12**(4):207-207.
265. Amin MB, Edge, S., Greene, F., Byrd, D.R., Brookland, R.K., Washington, M.K., Gershenwald, J.E., Compton, C.C., Hess, K.R., Sullivan, D.C., Jessup, J.M., Brierley, J.D., Gaspar, L.E., Schilsky, R.L., Balch, C.M., Winchester, D.P., Asare, E.A., Madera, M., Gress, D.M., Meyer, L.R. (ed.): AJCC (American Joint Committee on Cancer) Cancer Staging Manual, 8 edn. Chicago: Springer; 2017.
266. Clarke M, Collins R, Darby S, Davies C, Elphinstone P, Evans V, Godwin J, Gray R, Hicks C, James S *et al*: Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005, **366**(9503):2087-2106.
267. Martinkovich S, Shah D, Planey SL, Arnott JA: Selective estrogen receptor modulators: tissue specificity and clinical utility. *Clinical Interv Aging* 2014, **9**:1437-1452.
268. Carpenter R, Miller WR: Role of aromatase inhibitors in breast cancer. *Br J Cancer* 2005, **93 Suppl 1**:S1-5.
269. Dowsett M, Haynes BP: Hormonal effects of aromatase inhibitors: focus on premenopausal effects and interaction with tamoxifen. *J Steroid Biochem Mol Biol* 2003, **86**(3-5):255-263.
270. Lee CI, Goodwin A, Wilcken N: Fulvestrant for hormone-sensitive metastatic breast cancer. *Cochrane Database Syst Rev* 2017(1).

271. Valachis A, Mauri D, Polyzos NP, Mavroudis D, Georgoulias V, Casazza G: Fulvestrant in the treatment of advanced breast cancer: a systematic review and meta-analysis of randomized controlled trials. *Crit Rev Oncol Hematol* 2010, **73**(3):220-227.
272. Nathan MR, Schmid P: A Review of Fulvestrant in Breast Cancer. *Oncol Ther* 2017, **5**(1):17-29.
273. Gao JJ, Cheng J, Bloomquist E, Sanchez J, Wedam SB, Singh H, Amiri-Kordestani L, Ibrahim A, Sridhara R, Goldberg KB *et al*: CDK4/6 inhibitor treatment for patients with hormone receptor-positive, HER2-negative, advanced or metastatic breast cancer: a US Food and Drug Administration pooled analysis. *Lancet Oncol* 2020, **21**(2):250-260.
274. Sharma R, Hamilton A, Beith J: LHRH agonists for adjuvant therapy of early breast cancer in premenopausal women. *Cochrane Database Syst Rev* 2008(4):Cd004562.
275. Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, Abraham M, Medeiros Alencar VH, Badran A, Bonfill X *et al*: Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* 2013, **381**(9869):805-816.
276. Gianni L, Pienkowski T, Im Y-H, Roman L, Tseng L-M, Liu M-C, Lluch A, Staroslawska E, de la Haba-Rodriguez J, Im S-A *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. *The Lancet Oncology* 2012, **13**(1):25-32.
277. Valachis A, Nearchou A, Lind P, Mauri D: Lapatinib, trastuzumab or the combination added to preoperative chemotherapy for breast cancer: a meta-analysis of randomized evidence. *Breast Cancer Res Treat* 2012, **135**(3):655-662.
278. Martin M, Holmes FA, Ejlertsen B, Delaloge S, Moy B, Iwata H, von Minckwitz G, Chia SKL, Mansi J, Barrios CH *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):1688-1700.
279. von Minckwitz G, Huang C-S, Mano MS, Loibl S, Mamounas EP, Untch M, Wolmark N, Rastogi P, Schneeweiss A, Redondo A *et al*: Trastuzumab Emtansine for Residual Invasive HER2-Positive Breast Cancer. *N Engl J Med* 2018, **380**(7):617-628.
280. Araki K, Miyoshi Y: Mechanism of resistance to endocrine therapy in breast cancer: the important role of PI3K/Akt/mTOR in estrogen receptor-positive, HER2-negative breast cancer. *Breast Cancer* 2018, **25**(4):392-401.
281. Turner NC, Ro J, André F, Loi S, Verma S, Iwata H, Harbeck N, Loibl S, Huang Bartlett C, Zhang K *et al*: Palbociclib in Hormone-Receptor–Positive Advanced Breast Cancer. *N Engl J Med* 2015, **373**(3):209-219.
282. Finn RS, Martin M, Rugo HS, Jones S, Im SA, Gelmon K, Harbeck N, Lipatov ON, Walshe JM, Moulder S *et al*: Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med* 2016, **375**(20):1925-1936.

283. André F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, Iwata H, Conte P, Mayer IA, Kaufman B *et al*: Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *N Engl J Med* 2019, **380**(20):1929-1940.

284. André F, Ciruelos EM, Juric D, Loibl S, Campone M, Mayer IA, Rubovszky G, Yamashita T, Kaufman B, Lu YS *et al*: Alpelisib plus fulvestrant for PIK3CA-mutated, hormone receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: final overall survival results from SOLAR-1. *Ann Oncol* 2021, **32**(2):208-217.

285. Di Leo A, Johnston S, Lee KS, Ciruelos E, Lønning PE, Janni W, O'Regan R, Mouret-Reynier MA, Kalev D, Egle D *et al*: Buparlisib plus fulvestrant in postmenopausal women with hormone-receptor-positive, HER2-negative, advanced breast cancer progressing on or after mTOR inhibition (BELLE-3): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* 2018, **19**(1):87-100.

286. Dent S, Cortés J, Im YH, Diéras V, Harbeck N, Krop IE, Wilson TR, Cui N, Schimmoller F, Hsu JY *et al*: Phase III randomized study of taselisib or placebo with fulvestrant in estrogen receptor-positive, PIK3CA-mutant, HER2-negative, advanced breast cancer: the SANDPIPER trial. *Ann Oncol* 2021, **32**(2):197-207.

287. Kim S-B, Dent R, Im S-A, Espié M, Blau S, Tan AR, Isakoff SJ, Oliveira M, Saura C, Wongchenko MJ *et al*: Ipatasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (LOTUS): a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Oncol* 2017, **18**(10):1360-1372.

288. Schmid P, Abraham J, Chan S, Wheatley D, Brunt M, Nemsadze G, Baird R, Park YH, Hall P, Perren T *et al*: AZD5363 plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (PAKT): A randomised, double-blind, placebo-controlled, phase II trial. *J Clin Oncol* 2018, **36**(15_suppl):1007-1007.

289. Chien AJ, Tripathy D, Albain KS, Symmans WF, Rugo HS, Melisko ME, Wallace AM, Schwab R, Helsten T, Forero-Torres A *et al*: MK-2206 and Standard Neoadjuvant Chemotherapy Improves Response in Patients With Human Epidermal Growth Factor Receptor 2–Positive and/or Hormone Receptor–Negative Breast Cancers in the I-SPY 2 Trial. *Journal of Clinical Oncology* 2019, **38**(10):1059-1069.

290. Lord CJ, Ashworth A: PARP inhibitors: Synthetic lethality in the clinic. *Science (New York, NY)* 2017, **355**(6330):1152-1158.

291. Wu Q, Jiang L, Li S-c, He Q-j, Yang B, Cao J: Small molecule inhibitors targeting the PD-1/PD-L1 signaling pathway. *Acta Pharmacol Sin* 2021, **42**(1):1-9.

292. Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, Iwata H, Diéras V, Hegg R, Im S-A, Shaw Wright G *et al*: Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *New England Journal of Medicine* 2018, **379**(22):2108-2121.

293. Emens LA, Adams S, Barrios CH, Dieras VC, Iwata H, Loi S, Rugo HS, Schneeweiss A, Winer EP, Patel S *et al*: LBA16 IMpassion130: Final OS analysis from the pivotal phase III study of atezolizumab + nab-paclitaxel vs placebo + nab-paclitaxel in previously untreated locally advanced or metastatic triple-negative breast cancer. *Ann Oncol* 2020, **31**:S1148-S1148.

294. Alkner S, Bendahl P-O, Fernö M, Manjer J, Rydén L: Prediction of outcome after diagnosis of metachronous contralateral breast cancer. *BMC Cancer* 2011, **11**(1):114.
295. Saal LH, Vallon-Christersson J, Hakkinen J, Hegardt C, Grabau D, Winter C, Brueffer C, Tang MH, Reutersward C, Schulz R *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.
296. Ryden L, Loman N, Larsson C, Hegardt C, Vallon-Christersson J, Malmberg M, Lindman H, Ehinger A, Saal LH, Borg A: 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-e168.
297. Brueffer C, Vallon-Christersson J, Grabau D, Ehinger A, Häkkinen J, Hegardt C, Malina J, Chen Y, Bendahl P-O, Manjer J *et al*: Clinical Value of RNA Sequencing–Based Classifiers for Prediction of the Five Conventional Breast Cancer Biomarkers: A Report From the Population-Based Multicenter Sweden Cancerome Analysis Network—Breast Initiative. *JCO Precis Oncol* 2018(2):1-18.
298. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988, **239**(4839):487-491.
299. Vogelstein B, Kinzler KW: Digital PCR. *Proc Natl Acad Sci U S A* 1999, **96**(16):9236-9241.
300. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL, Tewari M: Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods* 2013, **10**(10):1003-1005.
301. Olsson E, Winter C, George A, Chen Y, Howlin J, Tang MH, Dahlgren M, Schulz R, Grabau D, van Westen D *et al*: Serial monitoring of circulating tumor DNA in patients with primary breast cancer for detection of occult metastatic disease. *EMBO Mol Med* 2015, **7**(8):1034-1047.
302. International Human Genome Sequencing Consortium: Finishing the euchromatic sequence of the human genome. *Nature* 2004, **431**(7011):931-945.
303. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012, **486**:346.
304. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G *et al*: The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012, **486**:395.
305. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977, **74**(12):5463-5467.
306. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE: Fluorescence detection in automated DNA sequence analysis. *Nature* 1986, **321**(6071):674-679.
307. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM: Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005, **309**(5741):1728-1732.

308. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB: Characterizing and measuring bias in sequence data. *Genome Biol* 2013, **14**(5):R51.
309. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al*: Real-time DNA sequencing from single polymerase molecules. *Science (New York, NY)* 2009, **323**(5910):133-138.
310. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M: Improved data analysis for the MinION nanopore sequencer. *Nat Methods* 2015, **12**(4):351-356.
311. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT *et al*: Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 2018, **36**(4):338-345.
312. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C: The Third Revolution in Sequencing Technology. *Trends Genet* 2018, **34**(9):666-681.
313. Reuter JA, Spacek DV, Snyder MP: High-throughput sequencing technologies. *Mol Cell* 2015, **58**(4):586-597.
314. Head SR, Komori HK, LaMere SA, Whisenant T, Nieuwerburgh FV, Salomon DR, Ordoukhanian P: Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques* 2014, **56**(2):61-77.
315. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ: Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* 2009, **6**(4):291-295.
316. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A: Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic acids research* 2009, **37**(18):e123.
317. Metzker ML: Sequencing technologies - the next generation. *Nat Rev Genet* 2010, **11**(1):31-46.
318. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR *et al*: Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008, **456**(7218):53-59.
319. Risca VI, Greenleaf WJ: Beyond the Linear Genome: Paired-End Sequencing as a Biophysical Tool. *Trends Cell Biol* 2015, **25**(12):716-719.
320. Brenner S: Loose end. *Curr Biol* 1995, **5**(3):332.
321. Hakkinen J, Nordborg N, Mansson O, Vallon-Christersson J: Implementation of an Open Source Software solution for Laboratory Information Management and automated RNAseq data analysis in a large-scale Cancer Genomics initiative using BASE with extension package Reggie. *bioRxiv* 2016.
322. Church DM, Schneider VA, Steinberg KM, Schatz MC, Quinlan AR, Chin C-S, Kitts PA, Aken B, Marth GT, Hoffman MM *et al*: Extending reference assembly models. *Genome Biol* 2015, **16**(1):13-13.
323. Kim D, Langmead B, Salzberg SL: HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015, **12**(4):357-360.
324. Broad Institute. *Picard tools*. http://broadinstitute.github.io/picard/.

325. Parekh S, Ziegenhain C, Vieth B, Enard W, Hellmann I: The impact of amplification on differential expression analyses by RNA-seq. *Sci Rep* 2016, **6**(1):25533.
326. Faust GG, Hall IM: SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 2014, **30**(17):2503-2505.
327. Boeva V, Zinovyev A, Bleakley K, Vert JP, Janoueix-Lerosey I, Delattre O, Barillot E: Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics* 2011, **27**(2):268-269.
328. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP *et al*: BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 2009, **6**(9):677-681.
329. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T: Transcriptomics technologies. *PLoS Comput Biol* 2017, **13**(5):e1005457-e1005457.
330. Sahraeian SME, Mohiyuddin M, Sebra R, Tilgner H, Afshar PT, Au KF, Bani Asadi N, Gerstein MB, Wong WH, Snyder MP *et al*: Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat Commun* 2017, **8**(1):59.
331. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, Johnson J, Dougherty B, Barrett JC, Dry JR: VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* 2016, **44**(11):e108.
332. Pedersen BS, Layer RM, Quinlan AR: Vcfanno: fast, flexible annotation of genetic variants. *Genome Biol* 2016, **17**(1):118.
333. Ramaswami G, Li JB: RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res* 2014, **42**(Database issue):D109-113.
334. Kiran A, Baranov PV: DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics* 2010, **26**(14):1772-1776.
335. Sun J, De Marinis Y, Osmark P, Singh P, Bagge A, Valtat B, Vikman P, Spégel P, Mulder H: Discriminative Prediction of A-To-I RNA Editing Events from DNA Sequence. *PLoS One* 2016, **11**(10):e0164962.
336. Picardi E, D'Erchia AM, Lo Giudice C, Pesole G: REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res* 2017, **45**(D1):D750-d757.
337. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, Ainscough BJ, Ramirez CA, Rieke DT, Kujan L *et al*: CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* 2017, **49**(2):170-174.
338. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L *et al*: COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 2017, **45**(D1):D777-D783.
339. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA: The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* 2018, **18**(11):696-705.
340. Vanderbilt-Ingram Cancer Center. *My Cancer Genome (release March 2016)*. http://www.mycancergenome.org.

341. Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC, Griffith OL, Griffith M: DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res* 2018, **46**(D1):D1068-d1073.

342. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A, Lopez-Bigas N: IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods* 2013, **10**(11):1081-1082.

343. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, Viklund J, Kähäri AK, Lundin P, Che H *et al*: SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. *Eur J Hum Genet* 2017, **25**(11):1253-1260.

344. Maretty L, Jensen JM, Petersen B, Sibbesen JA, Liu S, Villesen P, Skov L, Belling K, Theil Have C, Izarzugaza JMG *et al*: Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. *Nature* 2017, **548**(7665):87-91.

345. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012, **6**(2):80-92.

346. Guttery DS, Page K, Hills A, Woodley L, Marchese SD, Rghebi B, Hastings RK, Luo J, Pringle JH, Stebbing J *et al*: Noninvasive detection of activating estrogen receptor 1 (ESR1) mutations in estrogen receptor-positive metastatic breast cancer. *Clin Chem* 2015, **61**(7):974-982.

347. Lefebvre C, Bachelot T, Filleron T, Pedrero M, Campone M, Soria JC, Massard C, Levy C, Arnedos M, Lacroix-Triki M *et al*: Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS Med* 2016, **13**(12):e1002201.

348. Shaw JA, Guttery DS, Hills A, Fernandez-Garcia D, Page K, Rosales BM, Goddard KS, Hastings RK, Luo J, Ogle O *et al*: Mutation Analysis of Cell-Free DNA and Single Circulating Tumor Cells in Metastatic Breast Cancer Patients with High Circulating Tumor Cell Counts. *Clin Cancer Res* 2017, **23**(1):88-96.

349. Yanagawa T, Kagara N, Miyake T, Tanei T, Naoi Y, Shimoda M, Shimazu K, Kim SJ, Noguchi S: Detection of ESR1 mutations in plasma and tumors from metastatic breast cancer patients using next-generation sequencing. *Breast Cancer Res Treat* 2017, **163**(2):231-240.

350. Yu M, Bardia A, Aceto N, Bersani F, Madden MW, Donaldson MC, Desai R, Zhu H, Comaills V, Zheng Z *et al*: Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* 2014, **345**(6193):216-220.

351. Merenbakh-Lamin K, Ben-Baruch N, Yeheskel A, Dvir A, Soussan-Gutman L, Jeselsohn R, Yelensky R, Brown M, Miller VA, Sarid D *et al*: D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res* 2013, **73**(23):6856-6864.

352. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008, **5**(7):621-628.

353. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X *et al*: A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016, **17**(1):13.
354. Zhao S, Ye Z, Stanton R: Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* 2020, **26**(8):903-909.
355. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010, **28**(5):511-515.
356. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L: Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 2011, **12**(3):R22.
357. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015, **43**(7):e47-e47.
358. Sergushichev AA: An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* 2016:060012.
359. *Today in Science History: Science quotes by Fredrick Mosteller*. https://todayinsci.com/M/Mosteller_Frederick/MostellerFrederick-Quotations.htm (August 7, 2021).
360. Thomas E: An introduction to medical statistics for health care professionals: Hypothesis tests and estimation. *Musculoskeletal Care* 2005, **3**(2):102-108.
361. Sterne JAC, Cox DR, Smith GD: Sifting the evidence—what's wrong with significance tests?Another comment on the role of statistical methods. *BMJ* 2001, **322**(7280):226-231.
362. Altman DG, Bland JM: Uncertainty and sampling error. *BMJ* 2014, **349**:g7064.
363. Gourgou-Bourgade S, Cameron D, Poortmans P, Asselain B, Azria D, Cardoso F, A'Hern R, Bliss J, Bogaerts J, Bonnefoi H *et al*: Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol* 2015, **26**(5):873-879.
364. Kaplan EL, Meier P: Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* 1958, **53**(282):457-481.
365. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG: Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer* 1977, **35**(1):1-39.
366. Cox DR: Regression Models and Life-Tables. *J R Stat Soc Series B Stat Methodol* 1972, **34**(2):187-220.
367. Schoenfeld D: Partial Residuals for The Proportional Hazards Regression Model. *Biometrika* 1982, **69**(1):239-241.
368. Jackson SE, Chester JD: Personalised cancer medicine. *Int J Cancer* 2015, **137**(2):262-266.

369. Klevebring D, Lindberg J, Rockberg J, Hilliges C, Hall P, Sandberg M, Czene K: Exome sequencing of contralateral breast cancer identifies metastatic disease. *Breast Cancer Res Treat* 2015, **151**(2):319-324.

370. Takeshita T, Yamamoto Y, Yamamoto-Ibusuki M, Tomiguchi M, Sueta A, Murakami K, Omoto Y, Iwase H: Analysis of ESR1 and PIK3CA mutations in plasma cell-free DNA from ER-positive breast cancer patients. *Oncotarget* 2017, **8**(32):52142-52155.

371. Wang P, Bahreini A, Gyanchandani R, Lucas PC, Hartmaier RJ, Watters RJ, Jonnalagadda AR, Trejo Bittar HE, Berg A, Hamilton RL *et al*: Sensitive Detection of Mono- and Polyclonal ESR1 Mutations in Primary Tumors, Metastatic Lesions, and Cell-Free DNA of Breast Cancer Patients. *Clin Cancer Res* 2016, **22**(5):1130-1137.

372. Bahreini A, Li Z, Wang P, Levine KM, Tasdemir N, Cao L, Weir HM, Puhalla SL, Davidson NE, Stern AM *et al*: Mutation site and context dependent effects of ESR1 mutation in genome-edited breast cancer cell models. *Breast Cancer Res* 2017, **19**(1):60.

373. Jeselsohn R, Bergholz JS, Pun M, Cornwell M, Liu W, Nardone A, Xiao T, Li W, Qiu X, Buchwalter G *et al*: Allele-Specific Chromatin Recruitment and Therapeutic Vulnerabilities of ESR1 Activating Mutations. *Cancer Cell* 2018, **33**(2):173-186 e175.