From the Department of Oncology and Pathology
Karolinska Institutet, Stockholm, Sweden

TRANSCRIPTIONAL GENE SIGNATURES: PASSING THE RESTRICTION POINT FOR ROUTINE CLINICAL IMPLEMENTATION

Arian Lundberg

Karolinska
Institutet

Stockholm 2019

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by E-Print AB 2019
© Arian Lundberg, 2019
ISBN 978-91-7831-278-8

Transcriptional gene signatures: Passing the Restriction point for routine clinical implementation

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Arian Lundberg

Principal Supervisor:
Associate Professor Nicholas P. Tobin
Karolinska Institutet
Department of Oncology and Pathology

Co-supervisor(s):
Professor Jonas Bergh
Karolinska Institutet
Department of Oncology and Pathology

Associate Professor Linda S. Lindström
Karolinska Institutet
Department of Biosciences and Nutrition

Associate Professor Johan Hartman
Karolinska Institutet
Department of Oncology and Pathology

Associate Professor Ola Larsson
Karolinska Institutet
Department of Oncology and Pathology

Opponent:
Hege Russnes, MD, PhD.
Oslo University Hospital
Department of Cancer Genetics
Division of Cancer Medicine

Examination Board:
Associate Professor Ann Rosendahl
Lund University
Department of Clinical Sciences
Division of Oncology and Pathology

Associate Professor Teresita Diaz De Ståhl
Karolinska Institutet
Department of Oncology and Pathology

Professor Lars-Gunnar Larsson
Karolinska Institutet
Department of Microbiology,
Division of Tumor and Cell Biology

To a real-life wonder woman, my beloved sister by choice, Maryam, for believing in me and her endless kindness and support.

“Don’t be satisfied with stories, how things have gone with others. Unfold your own myth.”

Rumi

ABSTRACT

Uncontrolled cell growth and cell division are central to the process of tumorigenesis and a number of gene expression signatures have been developed based on genes that are involved in the cell cycle. Notably, gene expression signatures are used extensively in breast cancer research to examine the disease at a molecular level to describe tumour progression, treatment response and patients' survival.
The subject of this thesis is to explore the potential prognostic capacity of gene expression signatures in breast cancer and additionally, determine the prognostic capacity of a transcriptomic cell cycle activity (CCS) signature within variety of cancer types.

Several breast cancer gene expression signatures have emerged and been validated over the past two decades in large retrospective clinical trials. Although the clinical impact of these signatures has been clearly demonstrated, breast cancer therapeutic guidelines are still established on the basis of immunohistochemical markers (IHC) such as estrogen (ER), progesterone (PR), human epidermal growth factor 2 (HER2) and the proliferation marker Ki67. In **Study I**, the additional prognostic information derived from the combination of gene expression signatures and IHC/Ki67 was investigated in two Swedish breast cancer cohorts. Cohort I is comprised of 621 individuals with primary breast cancer tumours diagnosed between 1997 and 2005 in Stockholm region of Sweden. Cohort II consists of 484 individuals with primary breast tumours who diagnosed and received primary therapy in the Uppsala region of Sweden between 1987 and 1989. In Cohort I, Recurrence score (RS) and PAM50 gene expression signatures added prognostic information beyond Ki67 and IHC subtypes while only IHC subtypes provided additional prognostic information to all gene expression signatures with the exception of PAM50 gene signature in this cohort. Similar results were observed in Cohort II.

The ability of gene expression signatures to provide prognostic and treatment predictive information has been tested in primary breast tumours; however, their capability to provide similar information in the metastatic breast cancer (MBC) patients has not been investigated. In **Study II**, the prognostic capacity of gene expression signatures in breast cancer was evaluated in the metastatic setting in a Swedish multicenter randomized clinical trial known as "TEX" with 304 patients diagnosed with advanced locoregional or distant breast cancer relapse. A large number of tumours were classified into intermediate or high-

risk groups by all gene expression signatures. PAM50 was the only gene expression signature that provided prognostic information from lymph node (LN) metastases.

In **Study III**, the prognostic and treatment-specific potential of *CCND1* amplification was assessed in two breast cancer cohorts with 1965 and 340 patients, respectively. In the combined cohort, patients with *CCND1*-amplified tumours show worse survival in ER+/HER2-/LN-, luminal A and luminal B subtypes. Moreover, luminal A subtype with *CCND1*-amplified tumours shared similar gene expression changes with and luminal B subtype.

In **Study IV**, the DNA mutations and chromosome arm-level aneuploidy within tumours with different cell cycle activity (CCS) were explored. We showed that cell cycle activity varied broadly among and within different cancer types. Two well-known oncogenes (*TP53* and *PIK3CA*) exhibit the highest rate of mutations within different CCS groups. Furthermore, chromosomal arm level aberrations present in all CCS groups with a higher number of gains in 7p, 20q whereas deletions were more frequent within 17p and 8p arms. In the survival analysis, patients with higher CCS tumours show worse Progression-free interval relative to low and intermediate CCS groups.

In conclusions, we have shown that PAM50 and RS gene expression signatures can add prognostic information to Ki67 and IHC subtypes; however, IHC subtypes did not add any prognostic information to PAM50 signature. Moreover, PAM50 gene expression signature can provide prognostic information from LN metastases in MBC patients. Additionally, *CCND1* gene amplification has the potential to stratify patients with worse survival outcome within good-prognosis luminal A subtype tumours. Finally, we have demonstrated that CCS can provide independent prognostic information across cancer types.

LIST OF SCIENTIFIC PAPERS

I. **Lundberg A**, Lindström LS, Harrell JC, Falato C, Carlson JW, Wright PK, *et al.* Gene Expression Signatures and Immunohistochemical Subtypes Add Prognostic Value to Each Other in Breast Cancer Cohorts.
Clinical Cancer Research. December 2017, Volume 23, Issue 24
doi: 10.1158/1078-0432.CCR-17-1535

II. Tobin NP, **Lundberg A**, Lindström LS, Harrell JC, Foukakis T, Carlsson L, *et al.* PAM50 Provides Prognostic Information When Applied to the Lymph Node Metastases of Advanced Breast Cancer Patients.
Clinical Cancer Research. December 2017, Volume 23, Issue 23
doi: 10.1158/1078-0432.CCR-17-2301

III. **Lundberg A**, Lindström LS, Li J, Harrell JC, Darai-Ramqvist E, Sifakis EG, Foukakis T, *et al.* The long-term prognostic and predictive capacity of cyclin D1 gene amplification in 2,305 breast tumours.
Manuscript

IV. **Lundberg A**, Bergh J, Tobin P.N. A Pan-Cancer analysis of the impact of DNA perturbations on cell cycle activity levels
Manuscript

CONTENTS

1 Background 14
  1.1 Hallmarks of cancer 14
  1.2 The Cell cycle 14
    1.2.1 Cyclins and Regulation of the cell cycle 15
    1.2.2 Cell cycle checkpoints 16
    1.2.3 Cyclin D1 17
2 Breast Cancer 20
  2.1 Biology of the female breast 20
  2.2 Epidemiology 20
  2.3 Risk factors 22
  2.4 Classification 22
    2.4.1 Non-Invasive and Invasive breast cancer 22
    2.4.2 Histological grade 23
    2.4.3 Clinical and pathological staging 23
    2.4.4 The TNM system 24
  2.5 Breast cancer subtypes 24
    2.5.1 Estrogen receptor alpha 25
    2.5.2 Progesterone receptor 25
    2.5.3 Human epidermal receptor 2 25
    2.5.4 Ki67 – a marker of proliferation 26
  2.6 Molecular Subtypes 27
  2.7 Intrinsic Subtypes 29
  2.8 Prognostic & predictive factors in early breast cancer 32
    2.8.1 Axillary lymph node status 32
    2.8.2 Tumour size 32
    2.8.3 Tumour histological grade 33
    2.8.4 Lymphovascular invasion 33
    2.8.5 Age 33
    2.8.6 Ethnicity 34
    2.8.7 Proliferation Marker 34
    2.8.8 ER/PR and HER2 status 35
  2.9 Prognostic & predictive factors in metastatic breast cancer 35
    2.9.1 Hormone receptor and HER2 status 35
    2.9.2 Initial metastatic sites 36
    2.9.3 Age at the time of relapse 36
    2.9.4 Prior Adjuvant treatment 37
    2.9.5 Metastasis-free interval 37
    2.9.6 PI3K pathway – alterations in PTEN and PIK3CA genes 37
    2.9.7 BRCA1/BRCA2 genes 38
    2.9.8 ESR1, ERBB2 genes 38
    2.9.9 Intrinsic subtypes of MBC 38
  2.10 Treatments 39
    2.10.1 Neo-Adjuvant therapies 39
    2.10.2 Surgery and radiotherapy 39
    2.10.3 Adjuvant systemic treatment 39
3 Microarray data analysis 42
  3.1 Single-channel 42
  3.2 Dual-channel 43
  3.3 Pre-processing 44
    3.3.1 Background correction 44

3.3.2 Normalization 45
3.3.3 Summarization 46
3.4 Differential gene expression analysis 46
3.5 Classification methods 47
3.5.1 Unsupervised classification 47
3.5.2 Supervised classification 48
4 Gene expression signatures 49
4.1 Bivariate Gene Expression Signatures 50
4.1.1 70-Gene signature (MammaPrint) 50
4.1.2 Genomic grade index (MapQuant DX) 51
4.2 Multivariate Gene Expression signatures 52
4.2.1 Recurrence score (Oncotype DX) 52
4.2.2 PAM50 (Prosigna) 53
4.2.3 Cell cycle score 54
4.2.4 Breast cancer Index 54
4.2.5 EndoPredict 55
4.3 Prognostic classification of breast tumours (Ki67 vs gene signatures) 55
5 Pan-Cancer Atlas 56
Aim of the thesis 58
6 Patients and methods 59
6.1 Data collection 59
6.2 Patient cohorts 60
6.2.1 Merck Cohort 60
6.2.2 Uppsala Cohort 60
6.2.3 TEX Cohort 61
6.2.4 METABRIC Cohort 61
6.2.5 PanCan Atlas 62
6.3 Normalization 62
6.4 Statistical analysis 62
6.4.1 Study I 63
6.4.2 Study II 63
6.4.3 Study III 63
6.4.4 Study IV 64
6.5 Further Analysis 64
6.5.1 Study I 64
6.5.2 Study III 65
6.5.3 Study IV 66
7 Results And Discussions 67
7.1 Study I 67
7.2 Study II 68
7.3 Study III 69
7.4 Study IV 70
8 Conclusions 72
8.1 Study I 72
8.2 Study II 72
8.3 Study III 72
8.4 Study IV 72
Acknowledgements 73
References 75

LIST OF ABBREVIATIONS

ABCSG	Austrian Breast and Colorectal Cancer Study Group (ABCSG
AC	Combination of Cyclophosphamide with anthracyclines
ACC	Adrenocortical carcinoma
AI	Aromatase inhibitor
AJCC	American Joint Committee on Cancer
ANOVA	Analysis of variance
AR	Androgen receptor
ASCO	The American Society of Clinical Oncology
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related protein Kinase
ASR	Age-Standardized Rate
BCI	Breast Cancer Index
Bcl	B-cell lymphoma protein
BCSS	Breast cancer specific survival
BIG	Breast International Group Trial
BLCA	Bladder Urothelial Carcinoma
BRCA	Breast invasive carcinoma
BRCA1/2	Breast Cancer Gene 1/2
CBS	Circular Binary Segmentation
CCS	Cell Cycle Score
CDC	Cell division cell
CDF	Affymetrix chip definition file
CDK	Cyclin Dependent kinase
cDNA	Complementary DNA
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHK	Checkpoint kinase
CHOL	Cholangiocarcinoma
CISH	Chromogenic in situ hybridization
CMF	Cyclophosphamide, methotrexate, and 5-fluorouracil
CNA	Copy number alteration
CNV	Copy number variation

COAD	Colon adenocarcinoma
cRNA	Complementary RNA
CSS	Cancer specific survival
CTS	Clinical Treatment Score
DAC	Docetaxel, Doxorubicin, and Cyclophosphamide
DCIS	Ductal carcinoma in situ
DFS	Disease-free survival
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
DMFS	Distant metastasis free survival
EBCTCG	Early Breast Cancer Trialists' Collaborative Group
EGA	European Genome-Phenome Archive
EGFR	Epidermal growth factor receptor
ER	Estrogen receptor
ERα	Estrogen receptor-α
ESCA	Esophageal carcinoma
ESR1	Estrogen receptor 1 gene
FAC	Cyclophosphamide with Epirubicin and 5-Flourouracil
FFPE	Formalin fixed paraffin embedded samples
FISH	Fluorescence in-situ hybridization
FNA	Fine-Needle aspiration
GBM	Glioblastoma multiforme
GEO	Gene Expression Omnibus
GGI	Genomic Grade Index
GISTIC	Genomic Identification of Significant Targets in Cancer
HC	Hierarchical clustering
HER2	Human epidermal growth factor recptor 2
HG	Histological grade
HGNC	HUGO gene nomenclature company
HK	Housekeeping genes
HNSC	Head and Neck squamous cell carcinoma
Hr	Hormone receptor
HR	Hazard ratio

IDC	Invasive ductal carcinoma
IHC	Immunohistochemistry
ILC	Invasive lobular carcinoma
KEGG	Kyoto Encyclopedia of Genes and Genomes
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
KM	Kaplan-Meier analysis
LAML	Acute Myeloid Leukemia
LCML	Chronic Myelogenous Leukemia
LGG	Brain Lower Grade Glioma
LIHC	Liver hepatocellular carcinoma
LN	Axillary lymph nodes
LRχ2	Likelihood ratio test
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
LVI	Lymphatic vessels and vascular invasion
MAF	Mutational Annotation format
MBC	Metastatic breast cancer
MESO	Mesothelioma
METABRIC	Molecular Taxonomy and Breast Cancer International Consortium
MFI	Metastasis-free interval
MM	Mismatch probe
MT	Kinetochore microtubule
NAC	Neo-Adjuvant Chemotherapy
NHGRI	National Human Genome Research Institute
NHS	Nottingham Histological Score
NIH	The National Institute of Health
NKI	Netherlands Cancer Institute
NSABP	National Surgical Adjuvant Breast and Bowel Project clinical trial
OR	Odds ratio
OS	Overall survival

OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PAM	Prediction Analysis of Microarray
PCA	Principle component analysis
PCPG	Pheochromocytoma and Paraganglioma
pCR	Pathological complete response
PFI	Progression Free Interval
PFS	Progression free survival
PGR	Progestrone receptor gene
PM	Perfect match probe
PR	Progestrone receptor
PRAD	Prostate adenocarcinoma
qRT-PCR	Quantitative real-time polymerase chain reaction
RB1	Retinoblastoma-associated protein
RBL	Retinoblastoma-like protein
READ	Rectum adenocarcinoma
RFS	Recurrence free survival
RMA	Robust Multi-Array Average
ROR	Risk of relapse
RR	Relative risk
Rr	Risk of recurrence
RS	Recurrence score
RSEM	RNA-seq by Expectiation
RSR	Relative survival rate
SARC	Sarcoma
SEER	Surveillance, Epidemiology, and End Results Program
SKCM	Skin Cutaneous Melanoma
SNP	Single-nucleotide polymorphism
STAD	Stomach adenocarcinoma
TAILORx	Trial Assigning Individualized Options for Treatment
TCGA	The Cancer Genome Atlas
TGCT	Testicular Germ Cell Tumors

THCA	Thyroid carcinoma
THYM	Thymoma
TMA	Tissue microarrays
UCEC	Uterine Corpus Endometrial Carcinoma
UCS	Uterine Carcinosarcoma
UVM	Uveal Melanoma
WHO	World Health Organization

1 BACKGROUND

1.1 HALLMARKS OF CANCER

Cancer is a series of diseases caused by uncontrolled cell growth and cell division with the potential to spread to the other organs in the body. There are more than 200 known types of cancer. The steps leading to the transformation of normal cells to cancer cells, and their subsequent development and progression to a metastatic lesion are best described by the landmark "hallmarks of cancer" publication by Hannan and Weinberg[1]. The authors describe six main steps governing disease progression including sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis[1]. In the most recent update by the same authors, four new hallmarks were added[2]: deregulating cell energetics, avoiding immune destruction, genome instability and mutation, tumour-promoting inflammation. Whilst the authors note that cancers develop as an abnormal mass of tissue known as a tumour or neoplasm in a solid or fluid form, not all tumours are cancerous and benign (not cancerous), premalignant (precancerous) and malignant (cancerous) tumours have also been described[3, 4].

1.2 THE CELL CYCLE

In a normal cell, the process of duplicating DNA in the chromosomes and segregating the copies into genetically identical daughter cells is the basic function of the cell cycle[5]. The cell cycle in Eukaryotes is comprised of 4 distinct phases, each of which is shown in Figure 1. G1 phase: In proliferative cells the cell cycle starts with the "G1" or Gap1 phase where cells prepare for DNA replication in the S phase. In contrast, in highly differentiated cells, cell cycle begins from "G0" or Gap0 phase where cells are quiescent (non-proliferative). S phase: The next phase is known as the synthesis or "S" phase, where the DNA material of the cell is replicated. G2 phase: The "G2" or Gap2 phase resides between the S and M phase of the cell cycle and is where cells prepare for mitosis - the process where cell division, occurs. M phase: The "M" or mitosis phase is the final phase of the cell cycle when the replicated DNA is divided into two nuclei[5, 6](Figure 1), after which the cell undergoes cytokinesis and two daughter cells are formed.

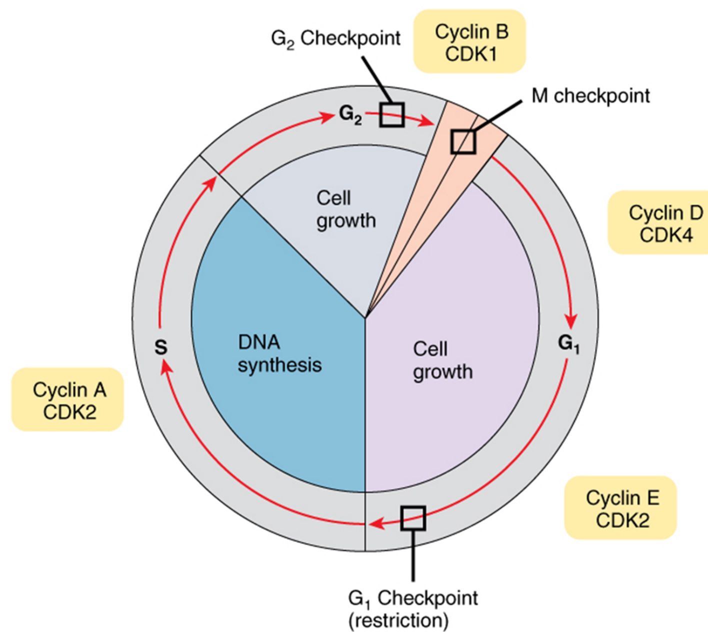


Figure 1. Schematic view of different phases of the cell cycle, cell cycle checkpoints and key components of the cell cycle regulation.

(Wikipedia under the license: CC BY-SA: Attribution -ShareAlike https://creativecommons.org/licenses/by-sa/4.0/)

1.2.1 Cyclins and Regulation of the cell cycle

Our knowledge on the cell cycle and cell cycle regulators has expanded over the past three decades. The cell division cycle (CDC) genes and cyclins were identified in yeast by Hartwell *et al.* for the first time[7] and following up on this discovery, Lee and Nurse *et al.* discovered a key regulator gene of the cell cycle "cdc2". Then, by isolating the corresponding cell cycle control genes in humans they demonstrated that these genes belong to the family of cyclin dependent kinases (CDKs)[8]. In 1982, Evans *et al.* discovered Cyclins in sea urchins and demonstrated their interaction with CDKs[9]. Hitherto, several cyclins and CDKs have been discovered. Cyclins in association with their binding partners CDKs govern multiple events including the progression of the cell cycle, cellular growth and DNA replication processes at different check-points of the cell cycle [5, 6].

Cyclin D: In the G1 phase of the cell cycle, Cyclin D1, D2 and D3 in combination with CDK4 and 6 phosphorylates retinoblastoma-associated protein (*RB1*) and retinoblastoma-like protein 1 and 3 (*RBL1*, *RBL3*); this process triggers Cyclin E1 and E2 expression and further phosphorylation of *RB1*, *RBL* and *RBL3*.

Cyclin E: E-type Cyclins are required in the early stages of DNA synthesis and the combination CDK2 and Cyclin E regulate G1 to S phase transition of the cell cycle[10, 11].
Cyclin A: In the G2 phase of the cell cycle, CDK2-Cyclin A2 complex initiates transition of S phase to mitosis in the M phase[10, 11].
Cyclin B: Cyclin B has an essential role for the progression of M phase in the cell cycle. Cyclin B-CDK1 complex initiates transition of G2 to M phase and the subsequent degradation of Cyclin B is necessary for the cell to exit the cell cycle.

Given the central role of Cyclins in regulating the cell cycle and by extension, cell proliferation, it is unsurprising that perturbations effecting the function and expression of Cyclin D[6, 12], E[13, 14], A[15, 16], B[17, 18], have all been linked to cancer. A schematic view of the cell cycle and involvement of Cyclins-CDKs complexes is shown in Figure 1.

1.2.2 Cell cycle checkpoints

The cell cycle machinery is highly regulated by several Onco- and tumour suppressor genes and other key regulatory elements such as the Cyclins-CDK complexes. There are three main restriction/checkpoints that occur during the cell cycle in order to stop its progression when it is not permitted[19]. These checkpoints take place at G1/S, G2 and M phases of the cell cycle (Figure 1).

G1/S checkpoint: In the G1/S checkpoint, the DNA bases will be examined for damages and prevented from being copied if any errors are recognized. After error identification, the DNA damage or stress signals have to be repaired in order to let cells pass this checkpoint[19]. These mechanisms are mediated by the ataxia telangiectasia mutated (ATM), ataxia telangiectasia and Rad3-related (ATR) protein kinases and their downstream targets checkpoint kinases (CHK) 1/2[19].

G2 checkpoint: In the G2 checkpoint, DNA strands are checked for any damages or breaks after the replication in the S phase. Any mistakes in the DNA replication detected by the checkpoint mechanisms at this point will induce cell cycle arrest. All the errors should be solved in order to permit cells passing G2 checkpoint[19].

M checkpoint: During the M phase of the cell cycle, chromosome segregation happens with the help of Kinetochore microtubules (MT)[20]. MTs will be attached to the center of chromosomes (centromeres), pulling them to the polar ends of the cell. At this checkpoint, all the prerequisites for chromosomal division will be checked ensuring proper progression of the M phase[19]. The M checkpoint prevents mitosis if the chromosomes are not properly attached to MTs or not correctly aligned with mitotic spindles[19].

In the case of a damage cell evading restriction points, a tumour suppressor gene (*P53*) can activate the CDK inhibitor 1 gene (*P21*) subsequently preventing Cyclin and CDKs from forming Cyclin-CDK complexes[21]. If DNA damages in the genome are not repairable, *P53* induces apoptosis by interacting with proapoptotic proteins such as B-cell lymphoma (Bcl)-2 protein family[22].

1.2.3 Cyclin D1

Cyclin D1 along with its binding partners CDK4 and CDK6 promotes progression from G1 to S phase of the cell cycle via the mechanism that has described in the previous section. In addition to the involvement of Cyclin D1-CDK4/6 complex in the cell cycle machinery, this complex contributes to several cellular functions including cell migration, cell growth, cell adhesion, cytoskeletal modeling and centrosome duplication as shown in Figure 2A[6, 23–26].

Furthermore, Cyclin D1 is independently associated in DNA repair and sequestration of CDK inhibitors promoting CDK2 activity. Subsequently, CDK2 binds to Cyclin E and forms the Cyclin E-CDK2 complex[27–29]. Cyclin D1 can also bind to P21 and P27 in the absence of CDK4/6, thus triggering the DNA damage response independently. Cyclin D1 is also involved in adipogenesis by interacting with transcription factors such as estrogen receptor-α (ERα) and Androgen receptor (*AR*) and their activators [6, 27, 28, 30].

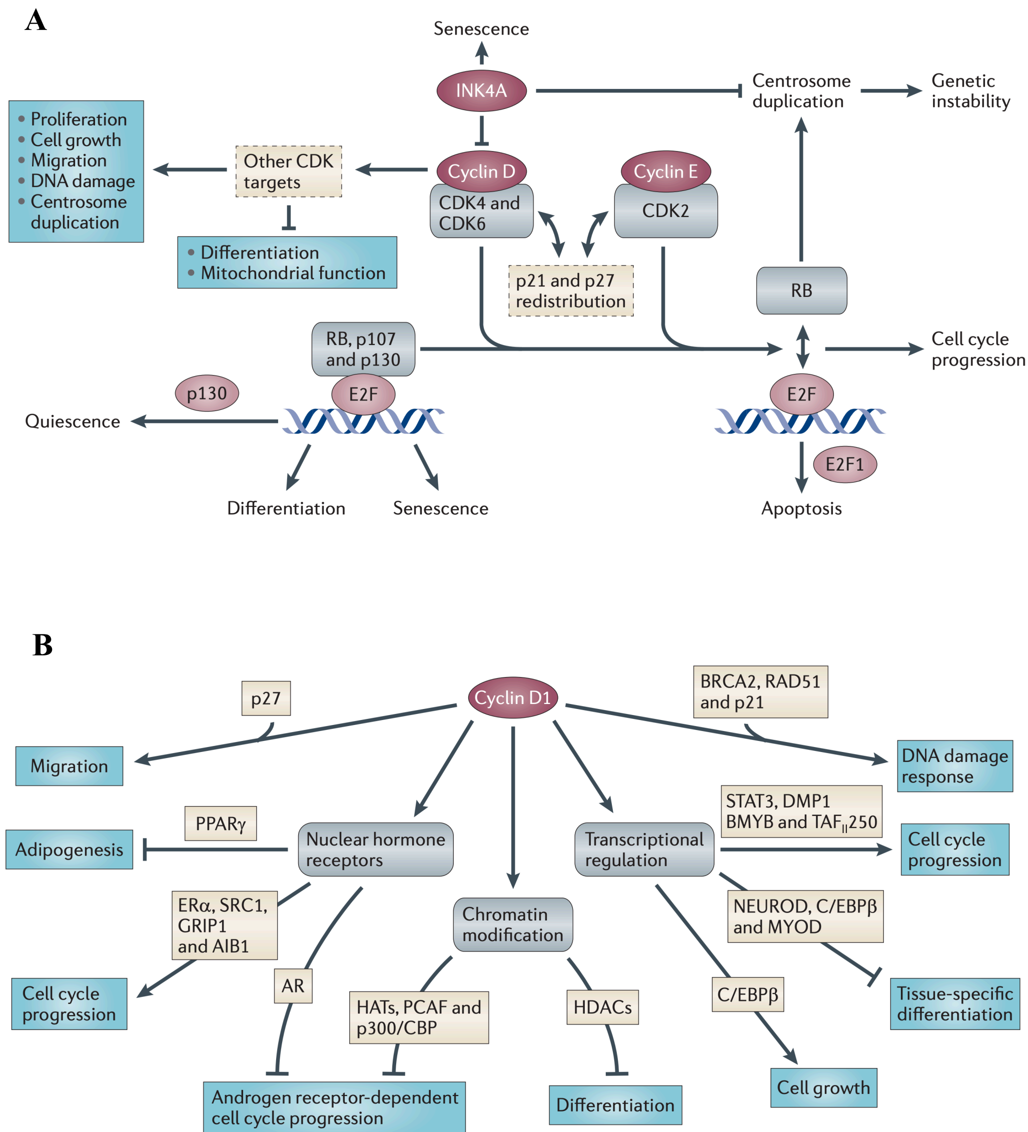


Figure 2. Functions of Cyclin D1 in (A) CDK-dependent and (B) CDK-independent manner (Musgrove et al. 2011)

Different genetic alterations in *CCND1* oncogene have been frequently observed in several cancer types (Figure 3). The overexpression and amplification of this gene have been reported to be correlated to poor prognosis in breast cancer and lung cancer patients[31–34]. Despite good correlation between *CCND1* gene amplification and Cyclin D1 protein expression, the overexpression of Cyclin D1 protein has contradictory impact on the prognosis of breast cancer patients. Some studies have shown its association with good prognosis in breast cancer patients[35, 36] whereas others have reported poor prognosis in these patients[37, 38]. Ortiz *et al.* have demonstrated that in a subgroup of patients with luminal A tumours, higher nuclear protein expression is associated with shorter disease-free

survival (DFS) (P = 0.029)[37]. Moreover, in a retrospective study with primary breast cancer tumours, patients whose tumours exhibited overexpression of Cyclin D1 protein have shown resistance to Tamoxifen treatment[38].

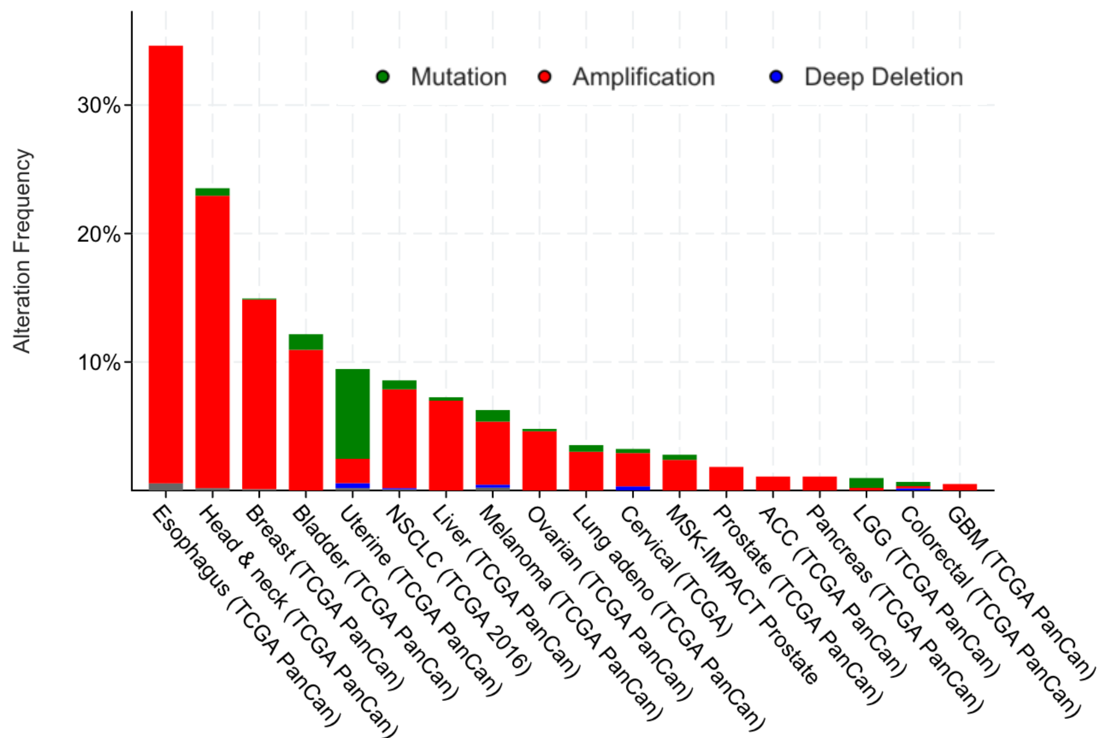


Figure 3. Frequency of genetic alterations in *CCND1* across different cancer types. Mutations, Gene amplification and deletions are shown in green, red and blue respectively.
ACC = Adenocarcinoma, LGG = Brain Lower Grade Glioma, GBM = Glioblastoma multiforme, NSCLC = non-small cell lung carcinoma; taken from cBioPortal (Gao et al. Sci. Signal. 2013 & Cerami et al. Cancer Discov. 2012)

2 BREAST CANCER

2.1 BIOLOGY OF THE FEMALE BREAST

The female breast is a very complex organ, comprised of adipose tissue with fibrous, connective tissues, nerves, lymph and blood vessels, as well as mammary glands which form around 15-20 lobules in a normal breast. During the process of breast tumorigenesis mammary cells undergo several genomic aberrations to bypass the cell cycle check points and tissue homeostasis. Interruption in cellular mechanisms such as cell migration or suppression of the immune system and genetic instability contribute to the development of cancer and metastatic progression. Several oncogenes including estrogen receptor 1 (*ESR1*), progesterone receptor (*PGR*), Human epidermal receptor gene 2 (*ERBB2*), Breast Cancer Gene (*BRCA*) 1 and 2 have been shown to be important prognostic and response to therapies predictors for breast cancer patients. More detailed information regarding the aforementioned risk factors will be provided in the following sections.

2.2 EPIDEMIOLOGY

Breast cancer is the most common cancer diagnosed in women (24.2%) with more than 2 million incidences each year[39]. According to World Health Organization (WHO) around 627,000 women died from breast cancer in 2018 that is around 15% of all cancer related deaths among women (Figure 4)[39].

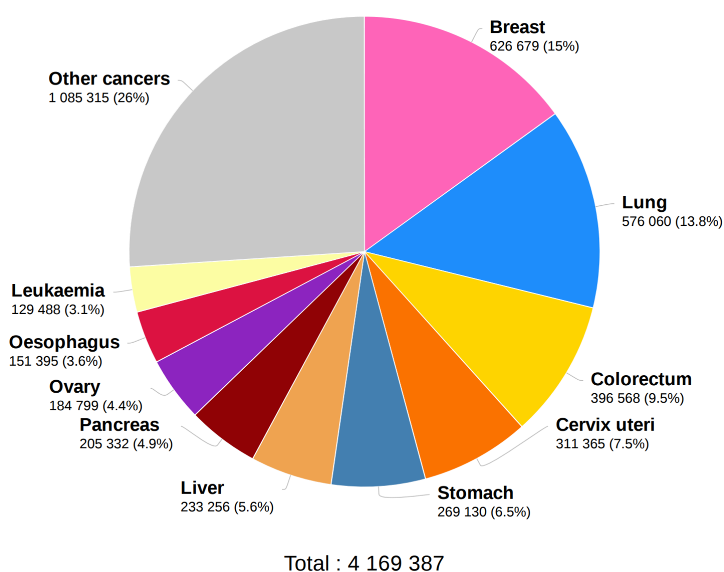


Figure 4. Estimated number of cancer-related deaths in 2018 among women around the world. (GLOBOCAN2018)

However, due to better prognosis, breast cancer is only the fifth leading cause of death in the world[40]. As shown in Figure 5A, despite a higher rate of breast cancer incidents in developed countries the survival rate is around 80% in North America and Europe. Whereas in less-developed countries the survival rate is significantly lower (40%), due to late detection of the disease, diagnosis and treatments in those regions[39]. Although incidence of breast cancer has increased exponentially in Scandiavian countries after 1960s, the number of breast cancer-related deaths has decreased in recent years; owing the population-based mammography screening in combination with adjuvant chemotherapy, endocrine and targeted therapies in the region (Figure 5B)[41]. According to the latest report from Swedish National Board of Health and Welfare (Socialstyrelsen)[42], around 5% of total 7,500 women who were diagnosed with breast cancer in 2016 died because of the disease[43].

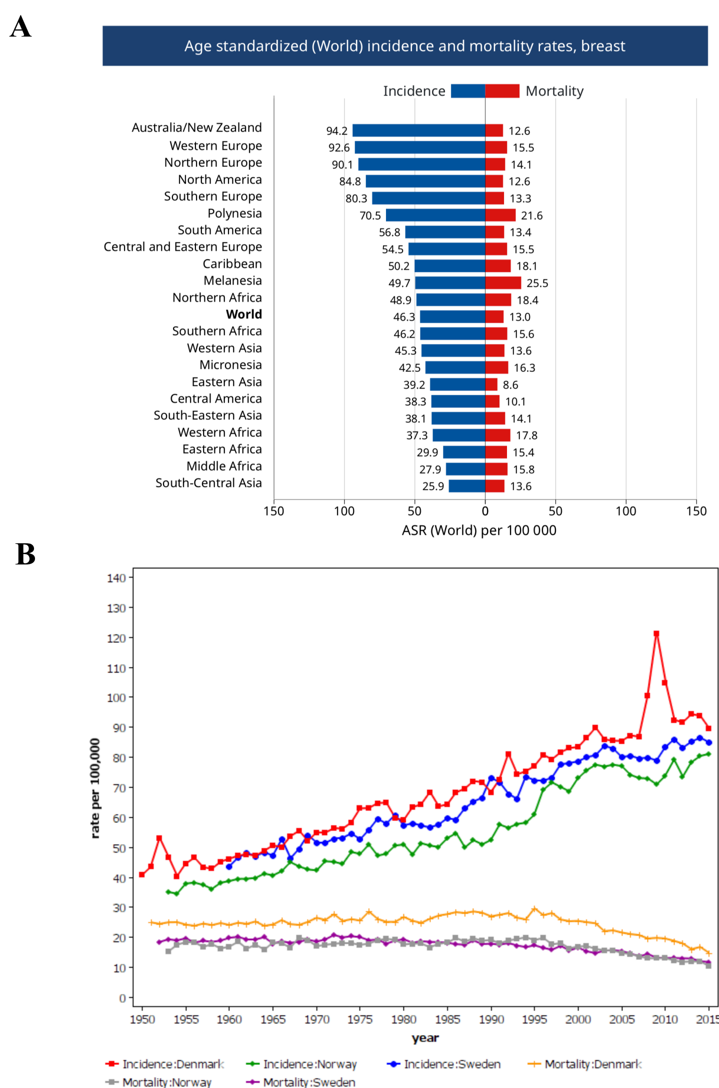


Figure 5. Age standardized incidence and mortality rates of breast cancer (A) around the world/per 100000 cases (WHO) and (B) in Scandinavian countries* between 1950-2015 (Nordic cancer registries 2018). *Incidences: Denmark in Red, Norway in Green, Sweden in Blue; Mortality rate: Denmark in Orange, Norway in Dark-Gray, Sweden in Purple; ASR = Age-Standardized Rate

2.3 RISK FACTORS

Cancer is also called a disease of ageing which leads to the malfunction of multiple cells and tissues. Apart from age, several other factors mostly associated with lifestyle have been identified as risk factors for cancer development including smoking, alcohol consumption, physical activities and body weight[44–46]. Family history of cancer and viral infections are also considered as risk factors in some cancer types[47]. However, some factors are directly associated with progression of specific cancer types. Risk factors for breast cancer include an early menstrual period, late or no pregnancies, menopause after age of 55, having a dense breast, using combination hormonal therapy and taking oral contraceptives have all been shown to be involved in increasing the risk of the disease[48].

2.4 CLASSIFICATION

2.4.1 Non-Invasive and Invasive breast cancer

Breast cancer can be divided into different types based on the cellular origin of cancer cells. A traditional pathological-driven classification of breast cancer divides the disease into three common types of carcinomas where the cancer spreads from epithelial cells of the breast tissue: ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC)[49]. DCIS is typically a non-invasive breast cancer; however, several studies have reported on the propensity of DCIS to become invasive (~ 40%)[50, 51]. IDC spreads from the milk ducts and invades into the fat tissue of the breast and has the ability to metastasize through the bloodstream or lymphatic system. ILC is another breast cancer type that starts in the lobules of the breast[52]; ILC is characterized by the loss of "E-cadherin" protein which is normally expressed in the epithelial tissue of breast[53].
Based on the presence of hormone receptors, invasive breast cancers can also be sub-divided into groups of those that are hormone receptor positive or negative. Hormone receptor positive tumours express the estrogen (ER), progesterone (PR) or human epidermal growth factor receptor 2 (HER2) receptors. In hormone receptor negative tumours, the growth of cancer cells is not dependent on systemic hormone levels and as such hormonal treatments (endocrine therapies) are not effective for this type of tumour which tends to be more aggressive and thus, grows faster than hormonal receptor positive

tumours[52, 54]. More detailed information regarding these receptors is provided in the following sections.

Apart from the hormonal receptor status of cancer cells, other traditional pathology-based classifications include tumour size, lymph nodal involvement and histological grade, all of which should be considered to determine the most appropriate treatment[55]. It should be noted that these factors still have limitations in classifying breast cancer patients accurately because some tumours with similar pathological characteristics may respond differently to similar treatments[55].

2.4.2 Histological grade

The histological grade of breast tumours provides important prognostic information in the management of the disease by aiding clinicians to make treatment decisions. Higher grade tumours tend to be more aggressive as opposed to lower grade tumours. One of the most validated scoring systems is the Nottingham Histological Score (NHS) also known as Elston-Ellis grading system. NHS classifies tumours into three different groups by considering three pathological factors of tubular formation, nuclear differentiation and mitotic counts which represents mitotic activity of the tumour cells. The combination of these features (scored 1-3 for each factor) makes a final score representing the final grade of the tumour from Grade I with 3-5, Grade II with 6-7 and Grade III with 8-9 score, respectively[56].

2.4.3 Clinical and pathological staging

Clinical staging is an estimation of a cancer spread, based on tumour biopsies taken from patients and a physical examination or tumour images. This approach is essential to choose initial treatment strategies especially before surgical interventions. Pathological or surgical stage of cancer is based on the reports of clinical staging. Additionally, pathological stage provides more precise information about cancer that has been obtained after the surgery; thus, it could be used to predict treatment response or prognosis in cancer patients[56].

2.4.4 The TNM system

The American Joint Committee on Cancer (AJCC) has proposed the TNM staging system as a standardized approach to classify cancer patients. TNM defines the prognosis and treatment recommendations of breast cancer patients by considering pathological and clinical staging of the tumour[56].

The TNM system classifies cancers into different stages based on the information about the primary tumour size “T”, the lymph node involvement “N” and metastasis spread “M”. The cancer could be categorized based on different features of the primary tumour into seven groups of Tx, T0, Tis and T1-T4; from unmeasurable tumour (Tx) to larger tumours that invaded the nearby tissues (T4). Lymph node involvement is another factor considered in TNM system. N is categorized according to the size, location and number of involved lymph nodes into five different groups of Nx, N0-N3 and the higher number indicates more aggressive cancer that has spread to the nearby lymph nodes. M0 and M1 represent positive or negative the status of metastatic tumour, respectively. Detailed information regarding The TNM system is accessible in the 8th edition of the AJCC cancer staging manual[57].

2.5 BREAST CANCER SUBTYPES

Clinical management of breast cancer greatly depends on the clinicopathological factors of tumours. These factors have strong association with patients’ prognosis and survival outcome. In routine clinical practice, three biomarkers of hormone receptor (Hr) status, including ER, PR receptors in addition to HER2 and a marker of proliferation “Ki67” are assessed to form the base of the adjuvant therapy given to patients. Assessment of Hr in addition to HER2 provides significant prognostic and predictive information regarding the response to hormonal therapy and HER2 targeted treatment, respectively. These information have improved our understanding about molecular features and heterogeneity of breast cancer. The ER, PR, HER2 and Ki67 biomarkers are explained in further detail in the following sections.

2.5.1 Estrogen receptor alpha

ERα is a transcription factor expressed by the *ESR1* gene in several organ and overexpression of ERα is commonly observed in patients with early stage breast cancer[58]. The expression of ERα is a strong prognostic and predictive biomarker in breast cancer patients and is routinely assessed in clinics by immunohistochemistry (IHC) staining in order to give treatment recommendations to breast cancer patients[58, 59].

Patients with ERα positive (ER+) tumours account for more than half of breast cancer incidences and these patients have better outcome relative to ERα negative (ER-) patients due to better response to anti-hormonal therapies such as Tamoxifen[60]. Recent recommendation guidelines state that tumours should be considered ER+ if more than 1% of tumour cells show positive nuclear staining by IHC tests[61]. However, some retrospective studies have shown that different threshold would be required for the ER status of the tumours since patients with 1-9% ER positivity have shown worse outcome relative to patients with tumours that are ≥ 10% ER+ in which the former group do not benefit from endocrine treatments[62, 63].

2.5.2 Progesterone receptor

PR status is used as a prognostic biomarker for ERα function in breast cancer patients[64]. PR is an ERα associated protein that controls ERα chromatin binding and transcriptional activity[64]. Additionally, PR has been shown to be a prognostic and predictor biomarker for early-stage breast cancer patients who have received adjuvant treatment in which, patients with ER+/PR+ tumours benefit more from hormonal therapy compared to ER+/PR-[65]. Furthermore, the importance of PR in relation to response to endocrine treatment has been studied[66]. Nevertheless, very little known about PR relative to ER and no targeted therapies have been proposed for the patients who might benefit from combined PR targeted treatment with standard ER targeted therapies.

2.5.3 Human epidermal receptor 2

HER2 also known as HER2/neu or *ERBB2* is a transmembrane receptor tyrosine kinase belonged to epidermal growth factor receptor family (EGFR). *ERBB2* plays an important role in several human malignancies. Overexpression of HER2 protein and

amplification of *ERBB2* gene have been observed in 15-30% of invasive breast tumours. The majority of patients with *ERBB2* gene overexpression had worse overall survival (OS) and poor prognosis until the *ERBB2* targeted therapy with Trastuzumab was approved to treat these patients[67].

Trastuzumab is a monoclonal antibody that targets *ERBB2* gene and is routinely used as a targeted therapy for patients whose tumours show overexpression of HER2 protein or amplification of *ERBB2* gene. There are two routine clinical tests available to stratify patients based on their HER2 receptor status: either by assessing the overexpression of HER2 protein using IHC or by fluorescence in-situ hybridization (FISH) which is used to determine *ERBB2* gene amplification[68].

2.5.4 Ki67 – a marker of proliferation

Uncontrolled proliferation is a distinct feature of different malignancies. There are several methods to assess proliferation including mitosis counting, flow cytometric or assessment of Ki67 antigen[69]. Ki67 is present in all active cell cycle phases (G1-S-G2 and M phases) while it is absent in the G0 phase of the cell cycle[70]. Ki67 staining is routinely used in a clinical setting to measure proliferation. Furthermore, several clinical studies have demonstrated Ki67 as an independent prognostic factor for OS in breast cancer patients[71, 72]. Ki67 protein prevents chromosomes from breaking down into a single bulk chromatin after the prophase phase of mitosis by interacting with the mitotic spindle leading to a normal genome replication[73]. Ki67 status carries a very important therapeutic implication in breast cancer patients; it can be used to stratify low-risk patients with luminal A tumours from more proliferative luminal B tumours since the latter respond better to adjuvant chemotherapy[74].

Although the clinical significance of Ki67 has been proven in several studies for breast cancer patients[71, 72, 74], there is still a lack of systematic methods to assess Ki67 in routine clinical practice. The international Ki67 in Breast Cancer Working Group proposed detailed guidelines to address this problem[75]. In this guideline, all issues regarding handling of tumour samples, fixation methods and recommended antibodies have been addressed[75]. According to the guideline, both core-biopsies and whole tumour sections are acceptable for the Ki67 scoring. In order to perform a prognostic assessment, the invasive area of a tumour should be considered to score Ki67. Additionally, the scoring result should be presented as a percentage relative to the total number of invasive cells in the scoring area[75].

The cut-off for Ki67 has changed in recent years; however, in the 14th Saint Galen Consensus Meeting, the "≥ 20%" cut-off was defined as "high Ki67" status to differentiate luminal A from luminal B tumours[76].

2.6 MOLECULAR SUBTYPES

As noted in previous sections, ER, PR, and HER2 receptor status along with Ki67 expression level are used as IHC surrogates for the classification of breast cancer tumours into 4 distinct subtypes including: luminal subtypes (A/B), *ERBB2*+/HER2-positive and Triple negative subtypes. All of these subtypes display different prognosis, response to treatment and distinct pattern of metastatic spread and survival time[77–82].
Luminal subtypes (ER+, PR+, HER2-): luminal tumours are the most common subtypes of breast cancer. These tumours show similar gene expression pattern to the genes in inner cells (luminal epithelial cells) of mammary ducts of a normal breast, including cytokeratin 8/18, ER and genes associated with ER activation pathways. luminal tumours can be divided into two subtypes.

Luminal A (ER+, PR+, HER2-/low Ki67): 30-70 percent of breast cancers are luminal A. According to the St. Gallen International Expert Consensus 2013, luminal A tumours are low proliferation tumours as measured by Ki67 with high expression of PR, moderate or high expression of ER, no expression of HER2 receptor and low pathological grade (1 or 2). They tend to have the best prognosis relative to other breast cancer subtypes with low recurrence incidences and high survival rates[77, 81]. luminal A tumours are highly responsive to endocrine treatment and less responsive to (neo) adjuvant chemotherapy[77].

Luminal B (ER+, PR+, HER2-/high Ki67): 10-20 percent of breast cancers are of luminal B subtype. These tumours are highly proliferative and show lower expression of PR related genes relative to luminal A tumours. luminal B tumours are ER+, PR+ and tend to have poor prognosis with higher tumour grade, larger tumour size and they often have lymph node metastasis[54, 83]. luminal B tumours positively respond to endocrine treatment and are relatively more sensitive to (neo) adjuvant chemotherapy than luminal A tumours[77].

HER2-positive (ER-, PR-, HER2+): 5-15 percent of breast cancers are HER2-positive subtype tumours. These tumours are characterized by amplified *ERBB2* or high HER2-protein expression and are highly proliferative; they show low or no expression of ER/PR related genes. These tumours often have higher pathological grade and high rate of liver metastases. HER2-enriched tumours response to (neo) adjuvant Trastuzumab treatment in combination with chemotherapy. They are also responsive to adjuvant anthracyclines and taxanes[78–80].

Triple-negative (ER-, PR-, HER2-): Around 15-20 percent of breast cancer tumours are triple negative. These tumours are commonly diagnosed in younger age women with African ancestry. They have the worst prognosis compared to other breast cancer subtypes with high metastasis rates[84–86]. These tumours do not show any expression of ER, PR and HER2.

It has been demonstrated that 45% of patients with triple-negative tumours have a pathological complete response (pCR) to anthracycline or anthracycline and taxane-based (neo) adjuvant chemotherapy[77] and 17% of them have shown a pCR to neoadjuvant platinum-based chemotherapy[82].

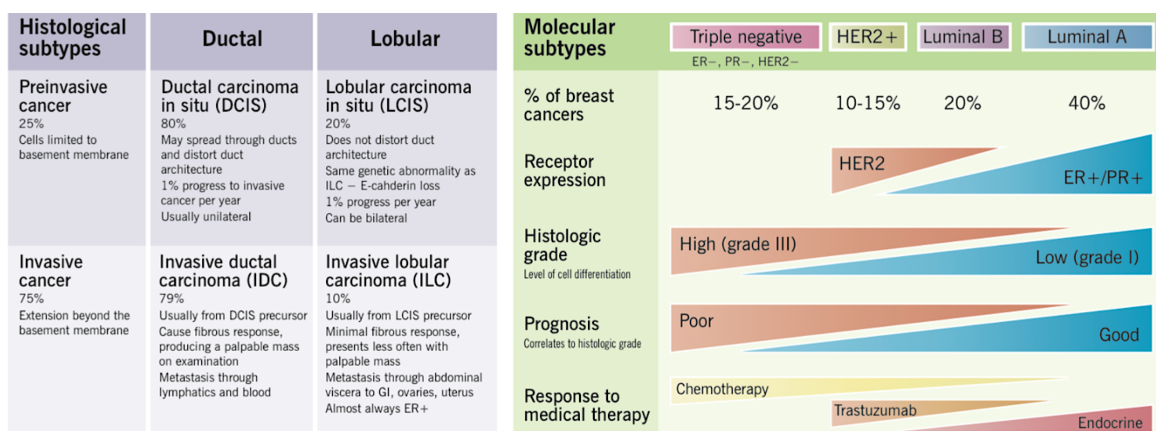


Figure 6. Histological and molecular subtypes of breast cancer.
(Adapted from McMaster Pathophysiology review - www.pathophys.org, Eric Wong, Jenna Rebelo and Sultan Chaudry.)

2.7 INTRINSIC SUBTYPES

Breast cancer is a heterogenous disease composed of multiple subtypes with unique morphological features and clinical outcomes. Technological advancement in the field of cancer discovery with high throughput platforms has helped us to understand the heterogeneity of the disease at a molecular level. Almost two decades ago, Perou *et al.* has demonstrated distinct gene expression patterns of breast cancer subtypes using the microarray-based classification technique. They proposed an initial list of 8102 genes involved in breast cancer progression[87], which was subsequently reduced to an intrinsic subset of 496 genes using a hierarchical clustering method. The expression pattern of intrinsic genes divided tumour samples into two main groups of ER+ and ER- cancers[87]. In 2001, Sørlie *et al.* described a distinctive "molecular portrait" of breast cancer using 456 cDNAs in which tumours were classified as five intrinsic subtypes of luminal A, luminal B, HER2-enriched, basal-like and normal-like tumours. Patients with different subtypes have different clinical outcomes[88]. The intrinsic subtypes exhibit distinguishable gene expression profiles among tumour samples (Figure 7).

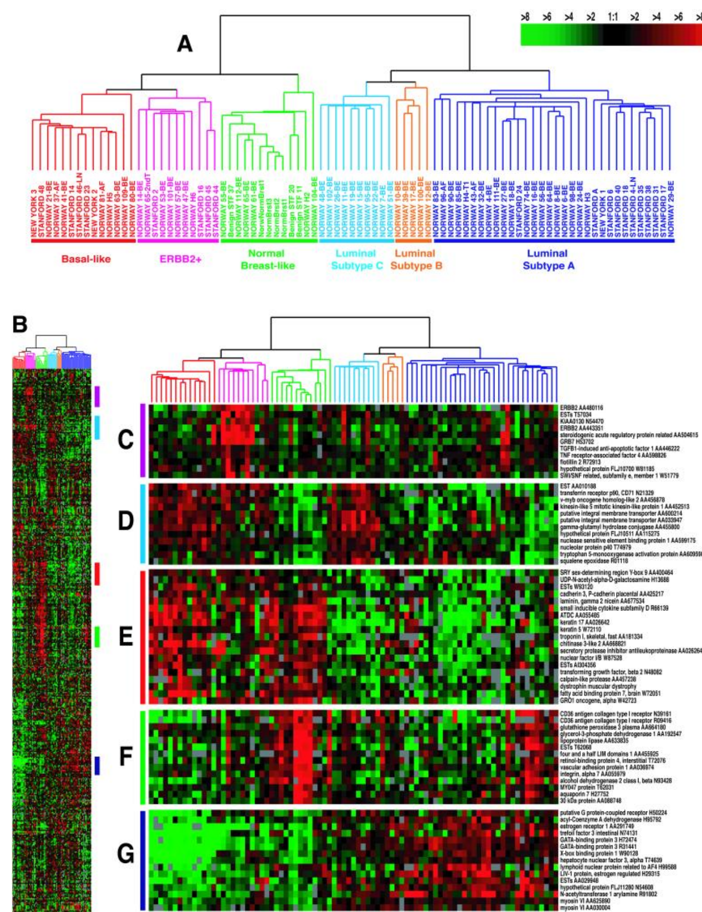


Figure 7. Intrinsic subtypes of breast cancer: A) Five subtypes were identified in 85 samples based on their gene expression pattern. Luminal A/B/C in dark-blue, yellow and light blue, respectively; normal-like in green, basal-like in red and ERBB2+ in pink. B) C. ERBB2 amplicon cluster, D. Novel unknown cluster, E. Basal epithelial cell-enriched cluster, F. Normal-like cluster and G. Luminal epithelial gene cluster enriched with ER. Green represents down-regulation, Red represents up-regulation (Taken from Sorlie et al 2001).

In a furtherance of exploring the intrinsic subtypes of breast cancer, Hu *et al.* tried to refine the list of genes included in the intrinsic subtypes. They proposed a new list of 306 genes which were associated with significant differences regarding OS and relapse free survival[89]. Following up on the Hu discovery, Parker *et al.*, used a supervised clustering method known as Prediction Analysis of Microarray (PAM) and decreased the number of genes to 50. This new classification method is known as PAM50.

PAM50 is comprised of Hr genes, proliferation genes and the genes that show basal and myoepithelial features (Figure 8)[90]. The significant prognostic and predictive value of the PAM50 signature has been demonstrated in several retrospective clinical studies[91–94]. Differences between PAM50 subtypes are presented in table 1.

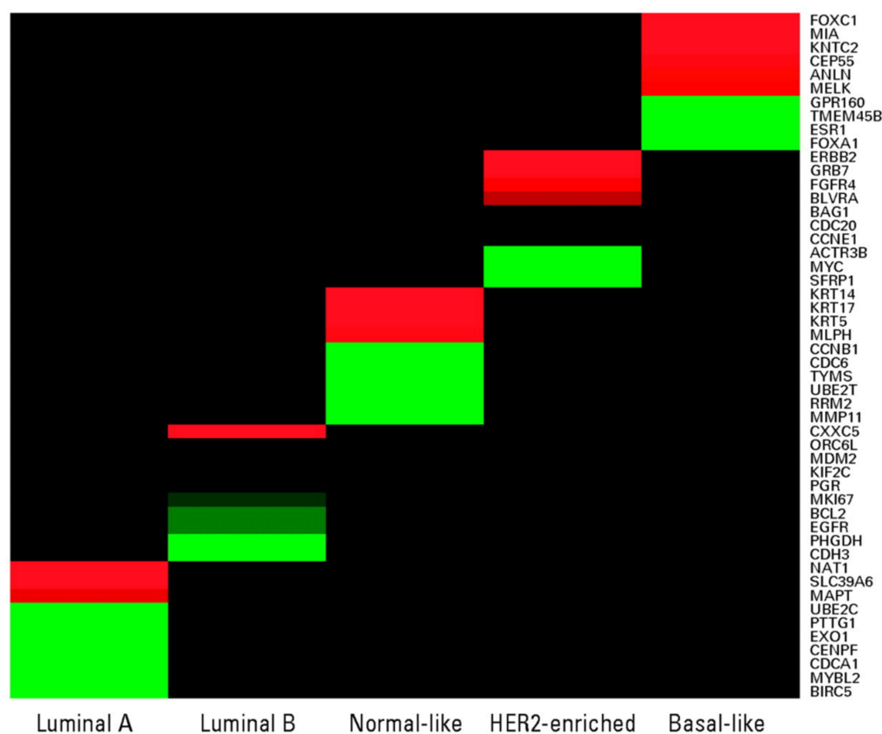


Figure 8. 50 genes included in PAM50 signature; the relative expression of genes included in each subtype are shown in red/green colors whereas black color indicates that genes were not chosen for the given subtype (Taken from Parker et al. 2009).

Table 1. Clinicopathological characteristics of PAM50 subtypes in breast cancer

	Luminal A	**Luminal B**	**HER2-Enriched**	**Basal-like**
Variables	30 - 70 %	10 - 20 %	5 - 15 %	15 - 20 %
ER*	Positive	Positive	Positive/Negative	Negative
PR*	Positive/negative (High)	Positive/negative (Low)	Negative	Negative
HER2*	Negative	Positive/negative	Positive	Negative
Ki67*	Low	High	High	High
Tumor grade	1 or 2 / low	High	High	High
LN status	Negative	Positive	Positive	Positive
Common mutations	TP53 - FOXA1 - PIK3CA	TP53 - PTEN - PIK3CA	TP53- APOBEC - PIK3CA	BRCA1 - TP53
Adjuvant therapy	Chemotherapy	(Neo) - Chemotherapy	Chemotherapy	Chemotherapy
	Endocrine therapy	Endocrine therapy	Anti-HER2 (Trastuzumab)	
Prognosis	High survival rate Low recurrence Best prognosis	Low survival rate Poor prognosis	Early recurrence Metastatic Poor prognosis	Aggressive Metastatic Worst prognosis
Metastases sites	Bone	Bone	Liver	Lung Brain

* = ER+ ≥ 1%, PR+ ≥1, HER2+ ≥10%/IHC3+, Ki67 ≥ 20% ; LN = Lymph node metastasis

2.8 PROGNOSTIC & PREDICTIVE FACTORS IN EARLY BREAST CANCER

After diagnosis of an invasive breast cancer, the first challenge is to identify patients who may benefit from adjuvant or neo-adjuvant treatments followed by choosing a proper therapy or combined treatment type for the diagnosed patients. Prognostic and predictive biomarkers address these challenges and help physicians to accurately make therapeutic recommendations. Prognostic factors are features representing a disease or a patient's characteristics at the time of surgery which can be correlated to natural history of the disease. These factors are mostly associated with OS or DFS. Predictive factors are indicative measurements exhibiting the likelihood that a patient will respond to a given treatment. In the following sections several prognostic and predictive factors involved in early and advanced breast cancer will be discussed in detail.

2.8.1 Axillary lymph node status

In patients diagnosed with an early stage breast cancer, involvement of axillary lymph nodes (LN) and increased number of engaged LN are highly associated with poorer outcome[95, 96]. In a previous report, LN+ patients with > 3 lymph nodes involved show lower 5-year median DFS relative to LN+ with ≤ 3 lymph nodes involved and therefore (5.4 years vs. 11.1 years , $P < 0.001$)[96] due to the high risk of distance metastases in LN+ patients with > 3 lymph nodes involved, adjuvant therapy should be given to these patients[97]. The strong prognostic value of LNs has consistently been shown in several studies[98–101]; however, some studies reported that LNs may not directly involved in the establishment distance metastases[102, 103].

2.8.2 Tumour size

The size of a primary tumour is another independent prognostic factor correlated with worse survival outcome in breast cancer patients[58, 104]. A study conducted under the Surveillance, Epidemiology, and End Results (SEER) Program has shown that patients with tumours between 2 to 5 or > 5 cm have worse 5-year breast cancer relative survival compared to patients with smaller tumours < 2 cm, respectively (79.8% and 62.7% vs. 91.3%) [105]. A similar trend has been demonstrated in another study with 20 years follow up where patients with larger tumours show shorter recurrence free survival (RFS) (75.5% +/-

2.3% for ≤ 2 cm vs. 63.2% +/- 3.1% for 2-5 cm ; $P < 0.001$, respectively)[106]. Additionally, in a recent population-based study, tumour size was reported as an independent prognostic factor for OS, breast cancer specific survival (BCSS), locoregional recurrence and distance metastasis(≤ 2.0 vs. > 2.0cm, Hazard ratio (HR) = 2.3, 95% CI, 1.13 – 4.78)[107].

2.8.3 Tumour histological grade

As previously mentioned, tumour grade is an important factor for the patients' prognosis in breast cancer. The NHS is a widely used grading system due to its higher reproducibility compare to other methods[108, 109]. The NHS grading system has been described in section 2.4.2.

2.8.4 Lymphovascular invasion

The lymphatic vessel and vascular invasion (LVI) have been demonstrated as an independent prognostic factor for BCSS (HR = 1.7 , $P =$ <0.0001)[110] and DFS (Relative risk (RR) = 2.489, 95% CI, 1.147-5.398)[111] in breast cancer patients. However, in a population-based study, LVI lost its strength as an independent prognostic factor to identify high-risk patients ($P = 0.65$)[112]. LVI is mainly considered as an additional guide to make therapy decisions for LN- patients who have borderline tumour sizes. In a retrospective study with 31 months follow-up, patients whose tumours exhibiting LVI have shown shorter progression free survival (PFS) and OS after neo-adjuvant chemotherapy(PFS:HR = 3.76 95% CI, 2.07-6.83, and OS:HR = 5.70 95% CI, 2.08-15.64, respectively)[113].

2.8.5 Age

Age at the time of diagnosis is considered as an important prognostic factor to identify patients with higher risk of recurrence (Rr). In several studies, the age of older patients has been shown to have a minor influence on their prognosis (Rr = 2.25, 95% CI, 1.66, 3.06)[114] (≥ 75 years, Relative survival rate (RSR) = 0.59, 95% CI, 0.57 - 0.61 vs. ≤ 34 years, RSR = 0.62, 95% CI, 0.58 – 0.66)[115]; additionally, younger women have shown to have worse 5-year cancer specific survival (CSS) compared to matched older patients (> 80 years vs. ≤ 41 years, CSS: 71.45% vs. 69.7%)[116].

2.8.6 Ethnicity

The ethnicity of patients has been investigated in relation to prognosis among different populations. Investigators suggested that several other factors may have been involved in the survival outcomes of patients with different ethnical background. Influential factors including limited access to healthcare and treatment facilities, tumour types, obesity or low rate of mammography screening screen were included in the report[117]. However, African-American women with early stage breast cancer who have gone under conservative surgery followed by radiation therapy have shown higher regional or distance recurrence rates along with lower RFS and OS (P = 0.01 and P = 0.0002, respectively)[118]. In another study, south-Asian women with Indian heritage have been compared to non-Asian (British-native) women where the former group showed higher 10-year RSR after the diagnosis independent from age, socioeconomic status and disease stage (South Asian:RSR = 72.6%, 95% CI: 69.0, 75.9% vs. non-Asian:RSR = 65.2% , 95% CI:64.5, 65.8%)[119].

2.8.7 Proliferation Marker

The prognostic capacity of Ki67, a proliferation marker in breast cancer has been tested in a large meta-analysis with more than 12,000 patients with early stage breast cancer in which, patients with high Ki67 had higher risk of relapse and worse DFS (HR = 1.93, 95% CI, 1.74 – 2.14, P = 0.001) and OS (HR = 1.95, 95% CI, 1.70 – 2.24; P = 0.001)[71]. The predictive capacity of Ki67 has also been reported as a surrogate marker for clinical response to hormonal treatment in breast cancer patients[120–122]. In the Breast International Group Trial (BIG) 1-98, Ki67 was used as a predictive marker in the adjuvant setting where treatment with Tamoxifen was compared with Letrozole. In this study, patients with higher proliferative tumours measured by Ki67 responded better to Letrozole over Tamoxifen and had longer DFS (HR[Letrozole: Tamoxifen] = 0.53; 95% CI, 0.39 to 0.72)[122]. Other studies have examined the predictive role of Ki67 in the neo-adjuvant setting[120, 121]. Both studies have shown a similar trend in which, patients with high Ki67 expression responded better to the treatments (Clinical response, P = 0.02 and Pathological response, P = 0.045)[121](P < 0.04)[120]. More information regarding Ki67 has been provided in section 2.5.4.

2.8.8 ER/PR and HER2 status

The assessment of the prognostic capacity of ER and PR is difficult due to the administration of adjuvant endocrine therapy; however, in a randomized clinical trial where none of the participants received systemic adjuvant treatment, ER+ patients have shown longer 5-year DFS relative to ER- patients ($P = 0.005$)[123]. Nevertheless, in a longer follow up study ER+ patients only exhibited better prognosis in the first 3 years of the follow-up while ER lost its prognostic significance after 3 years[124]. Despite limited prognostic strength of ER and PR, they provide significant predictive value.

As noted in section 2.5.3, overexpression and amplification of *ERBB2* were correlated with poorer prognosis before the routine administration of anti-HER2 treatments. Nevertheless, HER2+ patients treated with Trastuzumab have shown better prognosis when compared with HER2- patients, owing the therapeutic advantage of anti-HER2 regimen (OS:HR = 0.67, 95% CI: 0.48 - 0.92)[125]. In addition to prognostic capacity of HER2, several studies have demonstrated predictive capacity of this receptor which will be explored in later sections.

2.9 PROGNOSTIC & PREDICTIVE FACTORS IN METASTATIC BREAST CANCER

Most of prognostic and predictive factors of early stage breast cancer are well known however the knowledge is not well defined in metastatic breast cancer tumours (MBC). Regardless of the improvement in efficacy of treatments for early breast cancer patients, around one third of these patients will eventually develop metastasis[126]. The survival time for MBC patients varies between 24 - 36 months after the first relapse[127]. Therefore, identifying prognostic and predictive factors in MBC are essential for better treatment recommendation. In the following sections some of these factors will be discussed.

2.9.1 Hormone receptor and HER2 status

Different breast cancer subtypes have been shown to be correlated to different outcomes in MBC patients each of which has distinct pattern of metastatic root[128]. Kennecke *et al.* have demonstrated that MBC patients with luminal subtype tumours have

longer survival time relative to MBC patients with HER2-enriched or basal-like tumours[128]. Additionally, patients with luminal A/B subtypes showed lower rate of brain metastasis relative HER2-enriched and basal-like subtype tumours[128]. Several studies reported a relatively better outcome in MBC patients with hormonal positive (Hr+) /HER2+ primary tumours while triple negative tumours have the shortest survival rate among the patients[129]. In a retrospective study conducted before administration of anti-HER2 drugs, MBC patients with HER2+ subtype tumours show worse survival outcome compared to Hr+/HER2- tumours regardless of Hr status[128]. However, MBC patients with HER2+ tumours who were treated with Trastuzumab have shown a better prognosis compared to HER2- MBC (HR = 0.56; 95% CI, 0.45 - 0.69; $P < 0.0001$)[130]. Some recent reports the changes in the subtype of primary tumours have been compared to MBC in some patients[131–133] in which the metastatic tumours have lost/gained HER2 or Hr receptor status in MBC patients[134, 135]. In latter study, patients with ER+ MBC have better prognosis compared to ER- MBC, regardless of ER status of their primary tumours[135]; however, MBC patients with Hr-/HER2- tumours in relapse sites have shown poorer prognosis (HR = 1.48; 95% CI, 1.08 - 2.05)[135].

2.9.2 Initial metastatic sites

Several studies have been demonstrated the significance of the site of metastasis in prognosis of MBC patients. OS was shown to be better among MBC patients whose tumours have relapsed to the bone compared to the visceral or brain metastases[136–139]; similarly, metastases to the liver or distant lymph nodes reduced the survival rate in MBC patients[136, 140].

2.9.3 Age at the time of relapse

The age of MBC patients at the time of relapse has been poorly explored. Although in a few reports, the higher age and shorter OS and BCSS have shown to be directly correlated; the poorer outcome in elderly women could be related to the higher treatment related mortality rate in these patients [139, 141, 142]. Sabiani *et al.* have reported that younger women often have more aggressive tumours and poorer DFS (≥35years, Triple negative = 22.2% and HER2 tumours = 22.1%, respectively, $P < 0.01$, DFS:HR = 1.995)[143]. Also, these patients show higher risk of local and/or distance metastases as well as higher incidences of lymph node

metastases([>36 vs. > 35 years], 21.5%,15.4% ; [>36 vs 40 years], 21.8%,12.6%, $P < 0.01$) [143].

2.9.4 Prior Adjuvant treatment

Despite early detection of the breast cancer and administration of adjuvant treatments, the impact of the treatments on OS in MBC patients is very small and controversial. Some population-based studies reported longer survival time in Hr+ MBC patients who have received hormonal treatments in the primary setting [144, 145] while others stated small, no significant difference or decrease patients survival[146–148]. This could be due to the appearance of more aggressive and treatment-resistant tumour cells in MBC patients who were given adjuvant treatment at the early stage of the disease[148].

2.9.5 Metastasis-free interval

Metastasis-free interval (MFI), defined as the time of diagnosis of a primary breast cancer to the first distance metastasis has shown to be an independent prognostic factor in MBC patients[149]. A report shows that a survival outcome in patients with MFI < 2 years was significantly lower than the ones with MFI ≥ 2 years in a univariate analysis[150]. Another study, reported MFI > 10 years is associated with longer survival time in MBC patients and lower risk of metastatic BCSS mortality (HR = 0.77, 95% CI, 0.65–0.90; P = 0.191 and < 0.001, respectively)[151].

2.9.6 PI3K pathway – alterations in PTEN and PIK3CA genes

PTEN and *PIK3CA* oncogenes play key roles in activation of the PI3K pathways and are involved in resistance to Trastuzumab[152]. Mutations in *PIK3CA* and low expression of *PTEN* have been reported to have direct association with poor prognosis in MBC patients who have gone under Trastuzumab therapy. Also, patients who have alterations in both *PTEN* and *PIK3CA* have higher risk of disease progression compared to patients with only *PTEN* alteration[153]. Similar results have been shown in another study in which HER2+ MBC patients with mutated *PIK3CA* had poorer response to Trastuzumab subsequently had shorter survival time($P = 0.047$ and $P = 0.015$, respectively)[152].

2.9.7 BRCA1/BRCA2 genes

In a phase I trial of advanced breast cancer tumours, MBC patients with germline mutations in *BRCA*1/*BRCA*2 reported to have better response to Talazoparib[154]; additionally, in a randomized study HER2- MBC patients with germline *BRCA* mutations responded to the treatment with Olaparib[155]. Moreover, *BRCA2* mutation carriers with sporadic breast cancer have reported to have higher risk of distant recurrence in univariate analysis(HR = 1.81; 95% CI, 1.15 to 2.86; $P = 0.01$)[156].

2.9.8 ESR1, ERBB2 genes

ESR1 mutations in the ligand binding domain of ERα receptor are known to be involved in resistance to aromatase inhibitors (AI) by forming a ligand-independent interaction with ER [157, 158]. The majority of MBC patients with *ESR1* mutations exhibit poorer survival outcome and shorter PFS in metastatic tumours after given AI therapies[159, 160].

Therefore, *ESR1* mutation could be used as a predictive biomarker to select patients who may benefit from combination or selective endocrine therapies. In a prospective-retrospective phase III randomized study, *ESR1* mutated patients have shown better PFS after receiving Fulvestrant combined with Exemestane compared to the patients who did not carry this mutation.[161] In another study patients with *ESR1* mutated gene treated by Exemestane in combination with Placebo or Everolimus, had shorter OS in comparison to wild-type *ESR1* gene[162].

MBC Patients who acquired mutations in *ERBB2* in their relapsed tumours may not response to Lapatinib[163]. However, MBC patients with mutated *ERBB2* show response to Neratinib alone or in combination with Capecitabine suggesting that these patients are good candidates for Neratinib or Neratinib combined treatment[164].

2.9.9 Intrinsic subtypes of MBC

Several studies have reported changes in clinicopathological characteristics of early breast cancer to metastatic lesions during cancer progression[135, 165, 166]. The intrinsic subtypes of MBC tumours have shown to provide prognostic information regarding post-relapse survival for MBC patients where HER2-enriched and basal-like subtype tumours had worst survival outcome relative to luminal A subtype tumours[166].

2.10 TREATMENTS

2.10.1 Neo-Adjuvant therapies

Neo-adjuvant treatments are mostly offered as chemotherapy (NAC) to early stage breast cancer patients to downstage their tumour. Large clinical trials reported no benefits from NAC regarding DFS (RR = 0.95; 95% CI = 0.88 - 1.10; $P = 0.50$)[167], OS (HZ = 1.16; $P = 0.38$), PFS (HZ = 1.15; $P = 0.27$) and loco-regional recurrence (LRR) (HZ = 1.13; $P = 0.61$)[168] between pre or post-surgical treatments. Additionally, some concerns have been reported regarding the risks of neo-adjuvant treatments. However, neo-adjuvant therapies increase the opportunity for breast conservation treatment by lowering the tumour stage[169]. Although NAC reduces need of axilla lymph node dissection or mastectomy[170] it significantly reduces the chance for fertility preservation in the young patients[171]; additionally, an increased rate of loco-regional recurrences have been reported in patients who have undergone neo-adjuvant treatment prior to surgery[172].

2.10.2 Surgery and radiotherapy

Surgical removal of the tumour is an initial treatment for primary invasive breast cancer patients. Surgical techniques vary depending on the size and location of the tumour and include mastectomy or breast conserving surgery followed by radiotherapy. In a collaborative study conducted by Early Breast Cancer Trialists' Collaborative Group (EBCTCG) radiotherapy reduced loco regional recurrences by 10-years and breast cancer death after breast conserving surgery by 15-years[173]. In another report from EBCTCG group, the10-year risk of any first recurrences (Absolute reduction = 15,7%, 95% CI, 13.7–17.7, $P < 0.00001$) and 15-year risk of breast cancer death (Absolute reduction = 3.8%, 1.6–6.0, $P = 0.00005$) have been decreased in patients who have received radiotherapy after mastectomy[174].

2.10.3 Adjuvant systemic treatment

Adjuvant systemic treatment is an effective therapy given to patients with primary invasive breast cancer to reduce risk of loco-regional or distance recurrences and include chemotherapy, hormonal therapy and anti-HER2 therapy depending on tumour subtypes.

2.10.3.1 Chemotherapy

Several studies have shown that adjuvant chemotherapy reduces Rr in early breast cancer patients[175–177]. Adjuvant! Online is a web-based tool that is commonly used in clinics in order to assess the benefits of adjuvant treatments in breast cancer patients. Based on this tool adjuvant chemotherapy could be divided into three different regimens. These adjuvant regimens are generally comprised of two active cytotoxic agents including anthracyclines and taxanes. In the first combination adjuvant chemotherapy initiated by Bonadonna *et al.* in 1973, Cyclophosphamide, methotrexate, and 5-fluorouracil (CMF) were compared with the control population in a randomized study with LN+ patients who have undergone radical mastectomy. Patients who were given the combined treatment showed significantly improved DFS and OS[178]. Another study conducted by Mansour *et al.* reported a reduction in the Rr in LN- patients[175].
In a meta-analysis reported by EBCTCG, patients who have received CMF showed reduced Rr for 10-years and in overall mortality risk[176]. Combination of Cyclophosphamide with anthracyclines (AC) such as Doxorubicin has been tested in several trials[179–181]. In the National Surgical Adjuvant Breast and Bowel Project clinical trial (NSABP) B-15 trial, DFS and OS were similar between LN+ patients who have given AC compared to the patients who have received CMF[180]. Similar results were observed in LN- patients[181]. Several randomized trials tested the efficacy of combination therapy of Cyclophosphamide with Epirubicin and 5-Flourouracil (FAC) in ER+/LN+ patients[182]. Similar combination was tested in MBC patients in a tailored dose study[183]. An increase in DFS time by 9-years has been reported in ER+/LN+ patients while OS rate was the same[182].
In a metastatic cohort, increased DFS (HR = 0.63; P = 0.02) and OS (HR = 0.45; P = 0.005) was reported in the higher dose compared to lower dose treatment[183]. Efficacy of FAC was compared with Docetaxel, Doxorubicin, and Cyclophosphamide (DAC) in several randomized trials[184, 185]. Both trials showed the improvement in DFS (HR = 0.80; P = 0.0043) and OS (HR = 0.74; P = 0.002) in the first, and DFS (HR = 0.68; P = 0.01) in the second trial, respectively in favor of DAC treatment arm over FAC. In general, adjuvant chemotherapy reduces Rr in the early stage and MBC patients irrespective of age, axillary lymph node, Hr status or tumour grade and prior Tamoxifen treatment[176].

2.10.3.2 Endocrine therapy

As mentioned in the previous sections, only Hr+ patients respond to endocrine treatments. The benefit of 5-years Tamoxifen administration in Hr+ patients has been previously

reported[186]. Early stage breast cancer patients who have received 5 years adjuvant Tamoxifen treatment show lower Rr up to 15 years compared to patients who have only received the treatment for 2 years[186]. Furthermore, results from a randomized ATLAS trial show that ER+ patients who have received Tamoxifen for 10 years show a further decrease in Rr and mortality rate in comparison to those were given Tamoxifen for 5 years[60]. AI was given to postmenopausal patients as an additional treatment to endocrine therapy. Results from Austrian Breast and Colorectal Cancer Study Group (ABCSG)-16 showed 5 years AI treatment after adjuvant Tamoxifen therapy did not increase DFS compared to patients who received this combination therapy (HR = 1.007; $P = 0.947$)[187].

2.10.3.3 Anti-HER2 therapy

Adjuvant anti-HER2 treatment with Trastuzumab improves survival in patients overexpressing HER2[125] in combination with anthracycline-based regimens[188]. One-year Trastuzumab treatment is a standard duration given to HER2+ patients however, this combination therapy shows higher cardiac toxicity among treated patients. Therefore, some studies explored shorter duration of the treatment with 1-year standard Trastuzumab treatment[189, 190], the results of a meta-analysis study have shown that patients who have treated for 1-year show improved OS (HR = 1.28, p=0.04), and DFS (HR = 1.24, p=0.005) compared to shorter durations[190]. However, investigators of a large phase III randomized treatment "Persephone" recommended at the latest annual American Society of Clinical Oncology (ASCO)-2018 meeting to reduce the Trastuzumab treatment for early breast cancer patients to 6 months since majority of these patients show cardiac or other toxicities between 7-12 months after the treatment. The results from Persephone trial show comparable benefits for the patients who have been treated for a shorter duration compared to one-year standard regimen[191]. Pertuzumab is another FDA (U.S. Food and Drug Administration) approved anti-HER2 monoclonal antibody drug to treat HER+ breast cancer patients. Pertuzumab could be admitted in both neo-adjuvant and adjuvant settings[192, 193]. Additionally, patients with metastatic HER2+ tumours in the CLEOPATRA trial showed an increase in OS (median 56.5 months) under combination treatment of Trastuzumab and Docetaxel in addition to Pertuzumab, compared with Trastuzumab and Docetaxel alone (HR = 0.68, 95% CI, 0.56 - 0.84, $P < 0.001$)[194].

3 MICROARRAY DATA ANALYSIS

Microarrays were introduced for the first time in mid 1990s[195]. Since then this technology became one of the most extensively used methods in cancer research. Microarrays are employed to study in a large-scale the expression of genes. A microarray slide has numerous spots in defined locations on the slide that contain many single-strand DNA or RNA sequence fragments so-called "probes" which are complementary to a gene of interest. The probes labelled with fluorescent dyes are attached in different locations on a solid substrate of the array and bind to their complementary DNAs (cDNA) on the slide by a process known as "hybridization". Subsequently, a detector measures the intensity of fluorescent light emitted from the bound matched probes that signifies the expression of a gene of interest. Microarrays can be divided into two main groups of Single- and Two-channel detection, based on the method used to spot the labelled probes[196, 197].

3.1 SINGLE-CHANNEL

Single-channel microarrays can only be used to detect labelled probes rather than direct comparison between samples since only one sample is hybridized per each probe. Therefore, separate hybridizations are required to compare the conditions for the same gene. It should be noted that the signal intensity in the single-channel microarray is an absolute value, representing expression level of a predetermined probe/gene. However, it could be seen as a relative expression value when the predetermined probe/gene is compared to another condition or a probe. Affymetrix GeneChips are a one-channel microarray platform which is commonly used in cancer research. In total around 900,000 oligonucleotides are directly placed on the glass surface of the array. A combination of different oligonucleotide probes represents different genes.
The Affymetrix microarray probes are 25-mer long and consist of two different probes: Perfect match (PM) probes contain the complimentary sequence of a gene of interest thus representing the expression level of that particular gene. Mismatched probes (MM) have a single base mismatched sequence in the middle of the probes that are used to distinguish nonspecific hybridizations and/or background noises in the experiment from PMs[198].

3.2 DUAL-CHANNEL

Dual- or two-channel microarrays are comprised of two cDNA samples that are labelled individually with two different fluorescent dyes “Cy3” and “Cy5” with different wavelength, respectively. Two labelled samples will be hybridized together allowing investigators to have a quantitative and direct comparison between two probe samples on the same array. The signal intensity in the dual-channel method represents a relative gene expression within and between samples that are located on the array. Therefore, dual-channel method is very efficient in comparative experiments. A schematic view of a two-channel microarray experiment is shown in Figure 9.

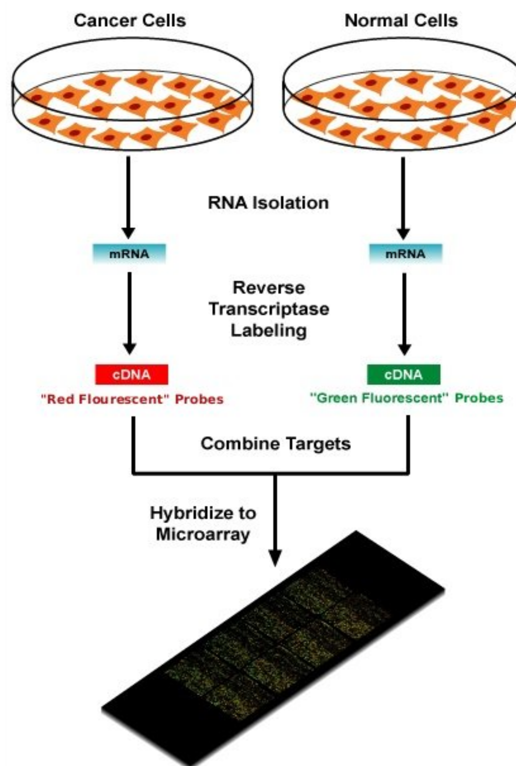


Figure 9. Different steps of two-channel microarray experiment. Two samples are labelled with different florescent dyes and hybridized on a same probe. Subsequently, the signal intensities of the samples are measured based on emitted lights from the probes (Taken from Wikimedia).

3.3 PRE-PROCESSING

The output from the microarray is often noisy and highly dimensional. Therefore, "trimming" the microarray data is essential for further analysis in order to extract relevant biological information from the experiment. In the preprocessing step, systematic errors caused by labelling, hybridization and scanning are removed, thus making the data comparable across multiple arrays. Several preprocessing methods have been proposed and yet there is no consensus on the choice of ideal preprocessing strategy[199–201]. Despite availability of different preprocessing methods based on microarray platforms, all of them consist of three main steps: Background correction, normalization and summarization.

3.3.1 Background correction

Background noise in microarray data impacts the signal intensity detected by the scanner. Background noise can be generated by non-hybridized probes a non-specific hybridization. As a result, background correction is needed to reduce the bias and increase the measurement accuracy of the signals. Various methods are available to estimate and adjust the background noise in microarray experiments. One of the most common algorithms for background correction in one-channel arrays is Affymetrix MAS 5.0[202]. In this algorithm, each chip breaks to a defined equal sized area "*k*". Then, a weighted average of lowest 2% of each region is chosen to calculate the local background of that explicit area B_k on the chip. Furthermore, MM probes are used as a complementary measurement of the background noise for PM probes in the *k* areas N_k. Then, the background noise of each specific probe in each particular geometrical location "(x , y)" is calculated by summing up of all B_k and N_k, respectively "B_a", "N_a". Moreover, for each *k* area, the local background B_a is subtracted from the raw intensity in that particular (x , y) location of the probe; unless this leads to a value lower than the noise value N_a, in that location. In that case, the probe intensity will be replaced by N_a[203].

Another commonly used background correction method is Robust Multi-Array Average (RMA)[204]. RMA estimates the signal intensity across all arrays by exclusively using PM probes while ignoring MM probes which may reduce the precision of the measurement system for the RMA method[205]. Including or excluding the MM probes in the background correction step subsequently impacts the choice of normalization and summarization techniques in the downstream analysis of the microarray data[206, 207].

Background correction of dual-channel microarray is usually achieved by subtracting values derived from the local background signals from the foreground ones. This approach greatly reduces the bias in the analysis. However, variance of lower signals can subsequently be magnified, causing new systematic artifacts. Therefore, there is a debate whether one should perform or ignore the background correction on the data derived from dual-channel microarrays[208–210].

3.3.2 Normalization

Several methods have been proposed to normalize microarray data and all of them attempt to remove systemic variations and balance the intensities of individual hybridizations in microarray experiments.

In the following sections, some of common normalization methods for microarray data including a) Standardization, b) Housekeeping gene-based, c) Quantile normalization will be briefly discussed.

3.3.2.1 Z-Score - Standardization

In the Z-Score method, gene expression data is normalized using the logarithmic value of signal intensities "log(s)"; log(s) values are normalized in order to have mean "μ" or median value "*M*" of 0 and standard deviation "*σ*" = 1. The standardized values known as "Z-score" are calculated by subtracting the μ or *M* from log (s) and then, divided by *σ*:

$$\underline{Z} = \log(s) - (\mu \text{ or } M) / \sigma$$

3.3.2.2 Housekeeping gene-based

In this method, the data are scaled based on the expression of the housekeeping genes (HK). The HK genes are the most common genes on a microarray and are assumed to have identical expression levels across all probes. However, this assumption has been proven to be false in several situations, especially across wider spectrum or several tissue types[211], or when some genes are not expressed relatively constantly in one tissue sample[212].

3.3.2.3 *Quantile normalization*

Quantile normalization aims to normalize probes' intensities in order to have a same distribution across arrays in which the intensity of each array will be replaced by the mean intensity of its level[200]. A significant limitation of this method is related to a necessity of having the same number of values in each sample, thus making it inappropriate for normalizing expression data derived from different array families.

3.3.3 Summarization

In some microarray platforms such as Affymetrix, a gene of interest could be represented by multiple probes. Therefore, a summarization step is needed in order to generate a single expression value for each probe sets (~20 pairs of probes in Affymetrix) that are representative of a single gene. Different summarization algorithms are available such as Tukey Biweight used in MAS 5.0 and Medianpolish (as a part of RMA normalization pipeline)[203, 204]. In the Affymetrix platform, each individual chip has a definition file (CDF) that contains annotation information. CDFs are used to map probe signals to genes. However, several studies have shown that the accuracy of Affymetrix microarray experiments can be improved by updating the annotations of the probes provided in the CDFs. As a result, it can increase the reproducibility of microarray data across different platforms[213, 214].

3.4 DIFFERENTIAL GENE EXPRESSION ANALYSIS

Microarray technology has been available for almost three decades and is still widely used because of its lower costs compared to other methods. It is applicable in several aspects of cancer discovery such as identifying single-nucleotide polymorphisms (SNPs) or mutations and classifying tumour types. One common application of microarray technology is gene and transcript differential analysis in which the pattern of gene expressions between cancer patients and normal patients are compared; thereby, tumour suppressor or oncogenes associated with a particular cancer can be identified.

3.5 CLASSIFICATION METHODS

Since microarray experiments generate a large amount of information, classification is essential to interpret the data in order to address relevant biological questions from the experiment. Classification methods are used to group samples/genes that share common features and distinguish them from other group of samples/genes e.g. a group of tumour samples vs normal samples. Additionally, classification can lead to the discovery of new groups or samples in the data e.g. identifying new subgroups of tumours within a specific cancer type. Unsupervised and supervised methods are two main approaches in microarray data analysis to classify samples. In the following sections these methods will briefly be discussed.

3.5.1 Unsupervised classification

In the unsupervised classification method, samples or genes are clustered with no previous knowledge regarding their class or group orders. Since no prior assumptions are involved in the clustering, unsupervised classification is impartial. Therefore, it represents a natural relationship between the samples e.g. genes that may have been involved in a similar pathway cluster together. Hierarchical clustering is one of the most common algorithms used in unsupervised classification.

3.5.1.1 Hierarchical clustering

Hierarchical clustering (HC) is used to group samples that are most similar into the same group known as cluster; each cluster has its distinct features that are shared within the cluster and are different from the other clusters. In the HC algorithm, samples are considered as individual clusters initially. Then, the two adjacent clusters will be identified using a distance metric method. In the final step, the most two similar clusters will be joined into one cluster. These steps will be executed repeatedly until all samples are clustered based on their similarities.

In order to perform HC, other factors such as measure of distance and linkage criteria should also be considered:

1) Measures of distance: represent the similarities between the samples. Several distance methods have been developed; therefore, the appropriate distance method should be chosen

based on a research hypothesis[215, 216]. As an example, in a simple Euclidean distance method, the distance between two clusters is based on the length of a straight line drawn from one cluster to one another.

2) Linkage criteria: is another factor in the HC algorithm that defines the position where distance should be calculated. Most common linkage criteria are a) Single-linkage, is the point between the two most similar clusters which is used as the starting point of the measurement, b) Average-linkage, in which the central point of all clusters is selected as the starting point of the measurement and c) Complete-linkage, the distance between the two least similar members of a cluster will be selected as the base of the distance measurement[216].

A partitional method is another clustering method whereby each sample is considered as one subset such that samples are divided into non-overlapping partitions.

Algorithms such as principle component analysis (PCA) have been introduced to reduce dimensionality of data[217]. In gene expression analysis, PCA initially considers genes as variables and then, generates a new variable that has linear correlation to the expression of those genes. Furthermore, variables/genes with the lowest variance compared to the new variable will be removed; subsequently, PCA reduces the complexity of the data, and emphasizing variation. It also reveals the interpretable pattern of gene expression in the dataset while keeping the most valuable information of the original data[218].

3.5.2 Supervised classification

Supervised classification is based on a predictor with known features associated with an outcome which could be used to predict the class of a new/unknown sample. In microarray analysis, supervised classification is used to generate a classifier or a signature that can be validated in a blinded dataset; additionally, supervised classification can identify a gene list that is important for the classification. Different predictors can be used in supervised classification of microarray data e.g. ER/PR status of a tumour or histopathological tumour grade are common predictors used to predict an outcome in cancer patients. Several supervised classification methods have been developed including: support vector machines (SVM) and logistic regression or neural networks and their utility have been previously demonstrated[219–222]. The predictive power of a supervised classifier should be validated with a new independent dataset to prove its independency from the features that are available in a test dataset.

4 GENE EXPRESSION SIGNATURES

In the context of breast cancer research and treatment, most of the prognostic and predictive factors noted here are known as clinicopathological characteristics of tumours and are used by clinicians to make therapeutic recommendations for breast cancer patients. It is appropriate to issue certain limitations regarding these factors. First and foremost, these factors are only partially independent from each other e.g. an ER- tumour is highly proliferative and tends to have high (II/III) tumour grade. Second, whilst assessment methods of clinical biomarkers such as ER, PR and HER2 have been standardized in recent years, Ki67 measurement still has high interlaboratory variability with a questionable accuracy. Third, these factors are measured with different scales, e.g. ER/PR are measured as binary variables while Ki67 and patients' age are continuous variables.

Thereupon, a multivariate predictor or classifier which incorporates all these factors is required in order to achieve an accurate prediction for patients. In this regard, gene expression signatures/classifiers have been proposed and their clinical utilities have been studied[93, 223–225]. In the context of cancer classification, a gene signature can be defined as a gene or combination of genes whose expression profiles are explicitly associated with a particular diagnosis, prognosis or response to a specific treatment in cancer patients[226]. In general, gene expression signatures in breast cancer can be divided in two main groups of first-generation and second-generation signatures. The first-generation signatures are capable of predicting recurrences in cancer patients within 5 years after initial diagnosis whereas second-generation signatures can make an accurate prediction both in early and late (more than 10-year) recurrences.

In following sections some common signatures will be described in detail.

4.1 BIVARIATE GENE EXPRESSION SIGNATURES

4.1.1 70-Gene signature (MammaPrint)

The 70-Gene signature is based on a DNA microarray platform and was initially developed on fresh frozen samples derived from 117 breast cancer LN- patients with < 5cm tumours. Despite not receiving systemic adjuvant treatment, half of the patients in the study were free from distance metastasis for at least 5 years. Therefore, investigators tried to generate a signature that could stratify the subgroup of patients with a low-risk of distance metastasis. Through the use of a supervised classification method, 70 out of 5000 initial genes were selected that showed the highest correlation with breast cancer prognosis.

The 70-Gene signature classifies LN- patients with stage I/II tumours into low or high-genomic Rr within 5-years after diagnosis. The signature has shown to have a significant prognostic capacity after being tested in a multivariate model, adjusting for all classical clinical prognostic factors (Odds ratio (OR) = 18.00, 95% CI, 3.3 - 94). The final 70 genes included in the signature are mostly involved in proliferation signaling, apoptosis resistance, invasion and metastasis well as angiogenesis[227].

The prognostic capacity of the 70-Gene signature has been validated in several retrospective studies. The predictive capacity of the signature has also been assessed where the 70-Gene signature predicts survival outcome for LN+ patients, in which patients in the low-genomic risk group show longer distant metastasis free survival times (DMFS) (98% at 5-years, 91% at 10 years) relative to patients in the high-genomic group metastasis (80% at 5-years, 76% at 10-years) respectively, when 1-3 nodes were involved[228]. Similar trends have been shown in patients with 4-9 lymph nodes involved[229].

Although, the 70-Gene signature has significant prognostic strength within 5-years after initial time of diagnosis, the prognostic capacity of this signature declined within a longer period of time[230]. Finally, the 70-Gene signature has also been validated in a prospective manner and it can be used to make treatment decisions in an adjuvant setting for early-stage breast cancer patients[223, 231].

In summary, 70-Gene signature has been validated in several retrospective, prospective and phase III randomized clinical trials; 70-Gene signature has been approved by FDA and is commercially known as MammaPrint Test.

4.1.2 Genomic grade index (MapQuant DX)

Histological grade of tumours (HG) is an important prognostic factor in breast cancer. Patients with high grade III tumours (HG3) show worse prognosis relative to low grade I tumours (HG1)[232]. However, the majority of tumours are classified as grade II tumours (HG2) with an intermediate Rr (30-60% of the cases) which makes it difficult to plan an effective treatment strategy for these patients. The Genomic Grade Index (GGI) signature was generated to address this problem by reclassifying HG2 tumours into GGI-1 and GGI-3 in which the latter group has a higher tendency for relapse (HR = 3.61, 95% CI, 2.25 to 5.78)[224].

The GGI signature is based on the Affymetrix microarray platform. In the original study investigators performed differential gene expression analysis on an initial training set of 65 ER+ tumours, comparing HG3 with HG1 tumours. The final number of 97 genes that had the highest association with HG were identified and used to generate the GGI signature. Since most of these genes are related to cell cycle and proliferation signaling, GGI could be seen as a representative signature of proliferation wherein GGI-1 tumours represents low proliferation and GGI-3 represents high proliferation level, respectively.

Furthermore, Sotiriou *et al.* have shown that GGI has the prognostic capacity to reclassify HG2 tumours into two groups of high and low Rr[224].

The GGI signature was shown to be able to predict response to combination neo-adjuvant chemotherapy with paclitaxel, fluorouracil, doxorubicin and cyclophosphamide, in which patients with GGI-3 were more sensitive to the treatment compared to GGI-1 patients[233, 234]

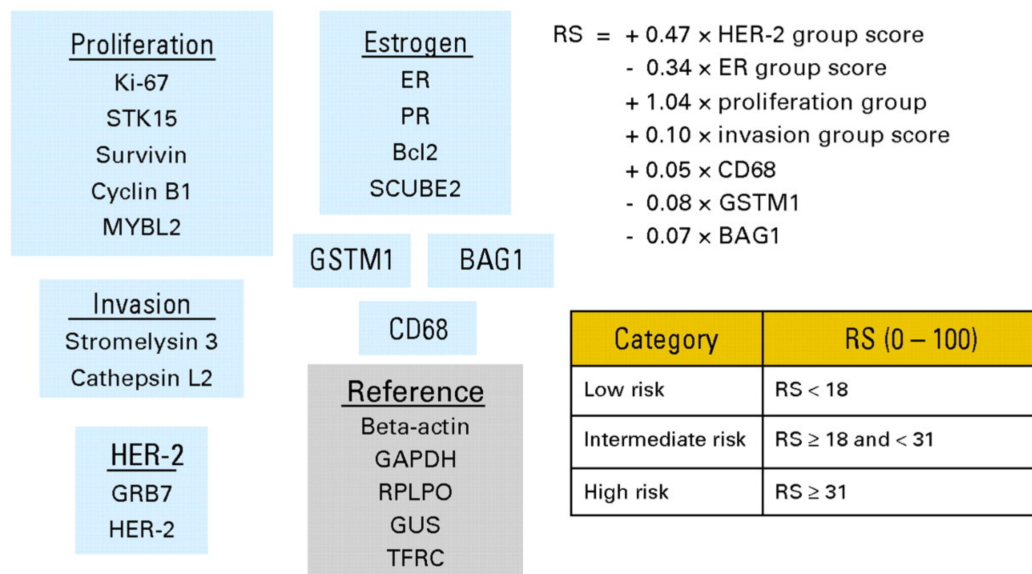
In a follow up study, the GGI signature was modified to run on quantitative real-time polymerase chain reaction (qRT-PCR) assay on formalin fixed paraffin embedded samples (FFPE). The PCR-GGI is comprised of 4 genes (*MYBL2*, *KPNA2*, *CDC2* and *CDC10*) associated with cell cycle progression and proliferation that were included in the GGI signature as well as 4 housekeeping genes (*GUS*, *TBP*, *RPLPO* and *TFRC*). The PCR-GGI signature has been shown have reliable prognostic performance in comparison to GGI, particularly in ER+ patients[235].

GGI signature is commercially known as MapQuant DX genomic grade test.

4.2 MULTIVARIATE GENE EXPRESSION SIGNATURES

4.2.1 Recurrence score (Oncotype DX)

The recurrence score (RS) is a RT-PCR based assay signature run on FFPE tissue samples in a central laboratory. This signature was initially tested on the FFPE samples derived from LN- patients who participated in the NSABP B-14 trial and have been treated only with Tamoxifen. Although chemotherapy has shown to reduce the RRs in early breast cancer patients[236, 237], the benefit of additional chemotherapy may be minimal in some patients[237]. RS is developed to identify patients who benefit from chemotherapy in addition to hormonal therapy in an adjuvant setting[238]. The RS is comprised of 21 genes associated with tumour development and cancer progression in addition to 5 reference genes (Figure 10). The risk score of recurrence is in a range from 0-100 (RS < 18 = low, 18 ≤ RS < 31 = intermediate and RS ≥ 31 = high Rr, respectively). RS classifies patients with ER+/LN-, stage I or II tumours into three different risk groups of low, intermediate and high Rr within 5-10 years after the initial time of diagnosis.



Category	RS (0 – 100)
Low risk	RS < 18
Intermediate risk	RS ≥ 18 and < 31
High risk	RS ≥ 31

Figure 10. Genes that are included in RS and their contribution in the risk score of recurrences. (Taken from Sparano et al. 2008.)

Higher expression of the genes belonging to the ER group are associated with lower risk of recurrence while higher expression of the genes belonging to the Proliferation group, HER2 group, invasion group and CD68 are considered to worsen patients' prognosis[238]. Several retrospective studies have shown the prognostic and predictive capacity of RS[239, 240].

The RS has also been validated in a prospective clinical trial with more than 10,000 patients known as the Trial Assigning Individualized Options for Treatment (TAILORx). The early

results of TAILORx study showed that patients with ER+/HER-/LN- tumours and low RS (0-10) have low risk of distance recurrences within 5 years[241]. Sparano *et al.* has recently reported follow-up results on TAILORx study, demonstrating that patients with ER+/HER-/LN- tumours and intermediate RS (11-25) do not benefit from adjuvant chemotherapy (HR = 1.08; 95% CI, 0.94 to 1.24; P=0.26, [Hormonal vs. combination therapy (Endocrine + chemotherapy)])[242]. The RS signature is the most common prognostic assays used in the United States for ER+ patients and is commercially known as Oncotype DX.

4.2.2 PAM50 (Prosigna)

The PAM50 signature was generated using a qRT-PCR protocol that classifies tumours into five intrinsic subtypes and is comprised of 50 cancer-related genes. Detailed information regarding intrinsic subtypes and development of PAM50 signature are provided in section 2.7. Parker *et al.* made a supervised continuous risk of relapse (ROR) score based on the PAM50 signature. ROR models were generated using clinicopathological characteristic of the tumours along with predefined variables to optimize the score[90]. These models include ROR-S, ROR-P and ROR-C.

ROR-S: is developed by incorporating a Pearson correlation score comparing the expression of genes in test samples with the expression level of prototypical genes included in PAM50 intrinsic subtypes. ROR-S was initially tested on a training set of 189 breast cancer samples and the risk models were trained independently on the Netherlands Cancer Institute (NKI) dataset with 141 LN-, untreated patients[90]. ROR-P: is generated by combining the ROR-S with a proliferation score derived from the expression of proliferation genes included in the PAM50 signature. ROR-P has been shown to have greater prognostic strength compared to standard clinical variables in a cohort with ER+ patients who received hormonal treatment[243]. ROR-C: is made by combining ROR-S and the tumour size[90].

The prognostic value and clinical impact of the PAM50 signature and ROR models have also been compared to other signatures[91, 243–246] where ROR provided more prognostic information than RS in ER+/LN- patients who received hormonal therapy[244]. Additionally, the ROR score added prognostic information to the Clinical Treatment Score (CTS) and could predict distance recurrences after 5 years in Hr+ patients who received hormonal therapy[246].

ProSigna is an FDA approved commercial assay that adapted the PAM50 signature using the NanoString nCounter Dx system to generate risk models for cancer patients. The Prosigna ROR score ranges between 0-100 indicating the probability of distance recurrence in patients. The Prosigna ROR score has been shown to be able to predict early and late recurrences in a clinical trial setting[244, 247, 248]. This test is strongly recommended for use as an additional guide on making treatment decisions only for breast cancer patients with luminal LN- tumours[249]. According to the Nanostring technologies, Inc. the Prosigna assay is not suitable to use for LN+ patients with more than 4 lymph nodes involved[250].

4.2.3 Cell cycle score

The prognostic value of gene expression signatures tends to rely heavily on genes associated with the cell cycle and proliferation[251–253]. Thus, the cell cycle score signature (CCS) was constructed on this notion: to reflect cell cycle activity of breast cancer tumours by employing genes that are directly involved in cell cycle progression[225]. A list of 463 cell cycle genes was assembled using the HUGO gene nomenclature company (HGNC), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Cyclebase databases[254–256]. CCS classifies tumours into three risk groups of low, intermediate and high cell cycle activity in which the latter group has the shortest BCSS relative to the low and intermediate group[225].

4.2.4 Breast cancer Index

The Breast Cancer Index (BCI) is another gene expression assay based on a RT-PCR method using FFPE samples performed by a central laboratory[257]. The BCI signature predicts distance recurrences in early stage breast cancer patients with ER+/LN- or LN+ tumours (1-3 nodes involved). BCI is developed by incorporating the expression of 5 genes associated with proliferation (*BUB1B, CENPA, NEK2, RACGAP1, RRM2*), 4 housekeeping genes (*ACTB, HMBS, SDHA*, and *UBC*) and a ratio between two genes (*HOXB13* and *IL17BR*)[258]. The *HOXB13* and *IL17BR* genes have been shown to be prognostic biomarkers in untreated or early-stage breast cancer patients who have received endocrine therapy[257, 258].

The prognostic power of BCI to predict response to extended hormonal therapy has been demonstrated in previous studies where this signature has the ability to predict late distance recurrences in breast cancer patients[258]. BCI has outperformed Oncotype DX and IHC

markers in predicting a late distance recurrences[259]. Additionally, BCI can identify ER+ patients with higher risk of distance recurrences who may benefit from extended adjuvant treatments[258, 260].

4.2.5 EndoPredict

Endopredict is another prognostic multivariate assay based on qRT-PCR; it is used to predict late distance recurrences in early stage ER+/LN- breast cancer patients who have been treated with Tamoxifen[261]. The Endopredict assay is applied, by using the expression level of 8 cancer related genes (*BIRC5, UBE2C, DHCR7, RBBP8, IL6ST, AZGP1, MGP*, and *STC2*) and 3 reference genes (*CALM2, OAZ1* and *RPL37A*). The combination of aforementioned genes, tumour size and LN status of the tumour generate a score calls EPclin. EPclin categorizes patients into low (EP score < 3.3 – lower than 10% of risk) and high risk (EPclin ≥ 3.3 – higher than 10% of risk) of distance recurrence within 10 years after diagnosis[261, 262]. The Endopredict signature outperformed common clinical parameters in terms of prognosis in both early and late distance recurrences in two adjuvant phase III clinical trials (ABCSG6 and ABCSG8) with ER+/HER2- patients who have been treated only with the endocrine therapy[262]. EPclin score has been proven to be a superior prognostic signature compared to the ROR-C and ROR-PT (ROR-P in combination with tumour size) score tested in a randomized phased III trial[263]. Buus *et al.* compared the prognostic strength of EPclin and EndoPredict with RS using Likelihood ratio test (LRχ2) in which both Endopredict and EPclin score outperformed the RS (LRχ2: EP = 49.3; LRχ2: EPclin = 139.3; LRχ2: RS = 29.1)[264].

4.3 PROGNOSTIC CLASSIFICATION OF BREAST TUMOURS (KI67 VS GENE SIGNATURES)

It has been shown that Ki67 measured by IHC can provide similar prognostic information as bivariate gene expression signatures[265]. Additionally, Ki67 and bivariate gene signatures have been reported to retain high correlation in terms of prognostic classification of breast tumours[265–267]. However, this correlation decreased when similar comparisons were performed between Ki67 and multilevel gene expression signatures[244, 265, 268, 269].

5 PAN-CANCER ATLAS

High-throughput platforms and multiple omics approaches are incorporated into the routine cancer research resulting a large amount of biological and clinical data. Vast resources have been put into initiating cancer projects to exploit these data derived from a large number of tumour samples at DNA, RNA, Protein and epigenetic levels. The Cancer Genome Atlas (TCGA) Research Network was initiated by the National Institute of Health (NIH) in 2005 with the aim of exploring genomic alterations in human cancers. Subsequently, a large number of systemic cancer genomic projects have been conducted by TCGA starting with glioblastoma, ovarian and breast cancers[270–272].TCGA projects have helped to identify novel oncogenes, biomarkers, cellular pathways and molecular subtypes with significant clinical implications[270, 272–275].

In 2013, TCGA announced a report from the first Pan-Can analysis project in which investigators have explored 12 different tumour types[276] in 6 different genomic, proteomic, transcriptomics and epigenomics platforms.

In a continuation of the first Pan-Can project, TCGA in collaboration with NIH and National Human Genome Research Institute (NHGRI) initiated a new Pan-Can Atlas project to study more than 11,000 samples across 33 cancer types (PanCan33) in similar platforms as the initial PanCan project in addition to whole genome sequencing and protein expression analysis. PanCan33 projects were conducted to extensively study cancer at a molecular level[277] by investigating the role of genetic variants and mutational loads in a variety of tumour types[278] and identifying signaling pathways involved in malignant transformation and cancer progression[279]; the tumour types studied in PanCan33 are Acute Myeloid Leukemia (LAML), Adrenocortical carcinoma (ACC), Bladder Urothelial Carcinoma (BLCA), Brain Lower Grade Glioma (LGG), Breast invasive carcinoma (BRCA), Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), Cholangiocarcinoma (CHOL), Chronic Myelogenous Leukemia (LCML), Colon adenocarcinoma (COAD), Esophageal carcinoma (ESCA), Glioblastoma multiforme (GBM), Head and Neck squamous cell carcinoma (HNSC), Kidney Chromophobe (KICH), Kidney renal clear cell carcinoma (KIRC), Kidney renal papillary cell carcinoma (KIRP), Liver hepatocellular carcinoma (LIHC), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC), Mesothelioma (MESO), Ovarian serous cystadenocarcinoma (OV), Pancreatic adenocarcinoma (PAAD), Pheochromocytoma and Paraganglioma (PCPG), Prostate adenocarcinoma (PRAD), Rectum adenocarcinoma (READ), Sarcoma (SARC), Skin

Cutaneous Melanoma (SKCM), Stomach adenocarcinoma (STAD), Testicular Germ Cell Tumors (TGCT), Thymoma (THYM), Thyroid carcinoma (THCA), Uterine Carcinosarcoma (UCS), Uterine Corpus Endometrial Carcinoma (UCEC) and Uveal Melanoma (UVM).

AIM OF THE THESIS

Study I: To examine if combination of Ki67 or IHC markers with gene expression signatures could provide more prognostic information than either of classifier alone.

Study II: To assess the ability of clinically relevant gene expression signatures to predict post-relapse survival in biopsies taken from MBC patients.

Study III: To clarify the prognostic and predictive capacity of *CCND1* gene amplification in breast cancer patients through long-term survival analysis.

Study IV: To explore the gene expression, DNA mutation and chromosomal arm-level alterations across and among PanCan Altas tumours with low, intermediate and high levels of cell cycle activity.

6 PATIENTS AND METHODS

6.1 DATA COLLECTION

Study I: Cohort 1: ER, PR, LN status and tumour size information of primary breast tumours were collected from the pathology reports. HER2 and Ki67 status were assessed on tissues microarrays (TMA) using chromogenic in situ hybridization (CISH) and MIB1 antigen, respectively. The gene expression microarrays data were obtained from Gene Expression Omnibus (GEO)[280] under the accession number of GSE48091.
Cohort 2: Clinicopathological information was derived from the patient records and routine clinical measurements at the time of diagnosis. ER and PR status were collected using biochemical assays. HER2 and Ki67 status were assessed on whole tumour sections. Tumour grades were derived from the pathology report based on the Elston-Ellis grading system. The gene expression microarray data can be retrieved from GEO under accession number of GSE3494. Survival data for both cohorts were retrieved from the Stockholm-Gotland Breast cancer registry and Socialstyrelsen[42].

Study II: All gene expression data were produced using on the Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray[166]. The gene expression microarray data are available on GEO under accession number GSE56493.

Study III: Cohort 1: The genomic and clinical data for this cohort (known as the METABRIC cohort) including, Hr and HER2 receptor status of the tumours, normalized copy number variation (CNV) and PAM50 molecular subtypes as well as normalized microarray gene expression data were taken from the online repositories accessible at European Genome-Phenome Archive (EGA) database[281], and are available for research purposes.
Cohort 2: is the same as Cohort 1 of Study I.

Study IV: Data from the PanCan33 was used in this study. The mRNA, sequencing data, mutational and chromosomal arm-level alteration data as well as tumour aneuploidy scores were downloaded from the NIH genomic data commons database. All data is anonymized and is freely available for all to use for research purposes[282].

All gene expression studies included in this thesis were approved by the ethics committee at Karoliniska Institutet (Stockholm, Sweden).

6.2 PATIENT COHORTS

6.2.1 Merck Cohort

The Merck cohort consists of 621 patients (768 study subjects) who were diagnosed with primary breast cancer in Sweden from January 1st 1997 to December 31st 2005 in the Stockholm health care area. This cohort was built upon a nested case-control design where the patients who were developed distant metastatic disease under the study period were selected as cases and the patients free from metastatic spread during the same period of time, were selected as controls. Then, the controls were randomly matched to each case in the following fashion: assigned to adjuvant treatment including chemotherapy, hormonal therapy or a combined treatment; the age of patients at primary breast cancer diagnosis were divided into (<45 years of age, 45-54 years, and >55 years).

The median follow-up of Merck cohort is 14.4 years until January 10th 2015. The Merck cohort was used in Study I (as Cohort 1) and Study III (as Cohort 2). However, the nested case-control design of this cohort was ignored in both studies and the individual patient data were included instead. The exclusion reasons for each study were as follows: Study I: Of total 621 number of patients in the cohort, 379 were included in the analysis. The reasons for exclusions were: bilateral tumours (n = 2), unclassified tumours (n = 13), missing ER (n = 13), PR (n = 147), HER2 (n = 96) and Ki67 information (n = 55). Study III: From original 621 number of patients, 340 were included in the analysis. The exclusions were due to: bilateral tumours (n = 2), unclassified tumours (n = 14, ER-/PR+/HER2- tumours), no matching SNP array (n = 68) and missing clinicopathological characteristics information (n = 197).

6.2.2 Uppsala Cohort

This cohort contains 468 patients with invasive breast cancer tumours who were treated with primary therapy in the Uppsala region of Sweden between January 1987 to December 31 1989. The age of patients was between 28 and 94 years. A subset of 97 patients had LN+ metastases who received systemic adjuvant therapy. Premenopausal women received chemotherapy whereas the postmenopausal patients received hormonal treatment. Median BCSS in this cohort was 12.5 years. The Uppsala cohort was used in Study I (as Cohort 2).
The exclusion reasons for the patients’ population in this study are as stated below:

Patients were excluded due to missing gene expression microarray information and lack of fresh frozen material (n = 251) as well as unclassified tumours (n = 24).

6.2.3 TEX Cohort

The TEX cohort, used in Study II is a Swedish multicentral randomized phase III trial where the combination of Epirubicin and Paclitaxel with/without Capecitabine were offered as a first line treatment. A total number of 304 patients with loco-regional advanced or metastatic breast tumours participated in the trial under a period of 5 years (December 2002 to June 2007). Tumour biopsies were taken from the most accessible metastatic site by Fine-needle aspiration (FNA). RNA was extracted from these samples for expression array profiling. The gene-expression array data from whole genome of 109 patients with complete follow-up were available at the end of TEX trial[283] and used for gene expression studies.

The reasons for exclusions in Study II were as follows: Of total number of 304 participants in the trial, 155 patients did not provide biopsies. 40 tumour biopsies did not pass the quality control. 4 biopsies classified as normal-like by PAM50 signature were excluded. The final inclusion was 105 samples in which 40 of them were LN biopsies and the rest of biopsies (n = 65) were taken from other metastatic sites.

6.2.4 METABRIC Cohort

Molecular Taxonomy and Breast Cancer International Consortium (METABRIC) cohort is a Canadian-British cohort consists of 1992 primary invasive breast cancer tumours. Only ER+/LN- patients received chemotherapy. Since this cohort study was initiated before approval of Trastuzumab, none of the HER2+ patients received anti-HER2 therapy[284]. DNA and RNA taken from samples were applied to Affymetrix SNP 6.0 and Illumina HT-12 v3 platforms to obtain gene expression and mRNA expression profiling, respectively. Median follow-up for censoring was 10.2 years. Of the original 1992 patients of METABRIC cohort, 1965 were included in Study III (as Cohort 1).

The reasons for exclusions were as follows: Duplicated samples (n=12) and unclassified samples (n=15, ER-/PR+/HER2- tumours).

6.2.5 PanCan Atlas

The PanCan Atlas project has been described in section 5. Of the total number of 11,071 patients in the PanCan33, 9,561 patients were included in Study IV (number of samples included in the study IV within each cancer type, ACC = 76, BLCA = 398, BRCA = 1038, CESC = 291, CHOL = 36, COAD = 428, DLBC = 47, ESCA = 161, GBM = 151, HNSC = 503, KICH = 65, KIRC = 481, KIRP = 280, LGG = 507, LIHC = 355, LUAD = 498, LUSC = 479, MESO = 81, OV = 289, PAAD = 158, PCPG = 160, PRAD = 471, READ = 154, SARC = 242, SKCM = 458, STAD = 404, TGCT = 133, THCA = 462, THYM = 103, UCEC = 514, UCS = 56 and UVM = 80). The reasons for exclusions are missing clinical data (n = 797), missing gene expression data (n = 212) and unmatched amplification data (n = 501).

6.3 NORMALIZATION

The normalization of microarray data in Study I, Study II and Study III (Cohort 2) was done in R statistical software version 3.3.3[285] using aroma.affymetrix package[286]. RMA method was chosen to normalize the data. The available data used in Study III (Cohort 1) has been normalized as described in the original study (supplemental information) using quantile normalization method[284]. The downloaded data used in Study IV (PanCan33) has been normalized as described in the original study[277] using RNA-seq by Expectation (RSEM) method[287].

6.4 STATISTICAL ANALYSIS

All the statistical analyses in the studies were performed using R statistical software version 3.3.3[285]. The individual analyses for each study are as listed below.

6.4.1 Study I

In this study, research versions of five gene signatures including PAM50, RS, GGI, CCS and 70-Gene signatures were tested in combination with Ki67 and IHC subtypes. The LRχ2 test was used to calculate the additional prognostic information added by the gene signatures to Ki67 or IHC subtypes. The same test was used in the reverse analysis. Results from LRχ2 test can be seen as a goodness of fit of models which let us compare the combination of Ki67/IHC subtypes with the gene signatures. We used coxph package in R to assess the LRχ2 test with BCSS as clinical endpoint.

6.4.2 Study II

The same gene expression signatures (PAM50, RS, GGI, CCS and 70-Gene signatures) as in Study I were evaluated in this study. Kaplan-Meier (KM) and multivariate Cox regression analyses with the post-relapse BCSS survival as a clinical end-point were performed in this study using survival package in R. The multivariate models were adjusted for age at diagnosis, diagnosis date and treatment received in the study. The prognostic capacity of the gene signatures was evaluated using LRχ2 test.

6.4.3 Study III

Differences between clinicopathological variables and *CCND1* amplified or non-amplified tumours, were evaluated using appropriate statistical tests based on the class of variables being compared. χ2 test was used when nominal variables were compared to nominal; Mann-Whitney test was used when ordinal variable was compared to nominal. The mean of gene expression of *CCND1* amplified/non-amplified was tested within different PAM50 subtypes using Student's t-test. The Analysis of variance (ANOVA) test with post-hoc Tukey was chosen to perform a similar comparison for gene expression levels among PAM50 subtypes. Multivariate Cox regression models and KM analysis with BCSS were employed in this study and all multivariate models were adjusted for tumour size, tumour grade, LN status, hormonal therapy and PAM50 subtypes (Cohort 1 and 2). In the combined cohort, multivariate models were also adjusted for the cohorts included in the study in addition to aforementioned factors.

6.4.4 Study IV

The same approach to Study III was taken to select appropriate statistical tests to assess differences among clinicopathological characteristics and inter-/intra-CCS subgroups. KM analysis was performed for inter-/intra-CCS subgroups. Due to the short follow-up time in the study, Progression Free Interval (PFI) was chosen as the clinical endpoint[288]. PFI is defined as the period during or after the treatment given to the patients in which the disease does not progress until loco-regional recurrences and/or other malignancies occur or the patients die from any cause. Multivariate Cox regression models were generated to evaluate the prognostic capacity of the inter- and intra-CCS subgroups. The multivariate models were adjusted for cancer types, age and gender of the patients, radiation therapy and pathological stage.

6.5 FURTHER ANALYSIS

In the following sections, individual methods used in each study will be discussed.

6.5.1 Study I

6.5.1.1 Hormonal receptor, HER2 receptor and Ki67 status

In Cohort 1, the ER and PR of tumours were collected from the pathological reports which was provided as a continuous variable. A ≥ 10 cut-off was used for the Hr positivity. Ki67, HER2 status were evaluated using MIB-1 antibody (1:100 dilution, DAKO) and CISH on TMAs, respectively. In Cohort 2, ER and PR were determined using ligand-binding assay. Ki67 expression was assessed by a pathologist at the invasive edge of a tumour and represented as a continuous variable. The same cut-offs as in Cohort 1 was used for the Hr positivity in Cohort 2. However, HER2 and Ki67 were evaluated on whole tumour sections in this cohort.

6.5.1.2 IHC subtyping

The IHC subtypes were generated using ER/PR/HER2 and Ki67 status of the tumours in the following manner: luminal A like: ER+ and PR+/- (considered as Hr+), HER2- and Ki67 low; luminal B like: Hr+, HER2-, Ki67 high; HER2+: Hr+/-, HER2+, Ki67 low/high;

and triple negative: Hr-, HER2-, Ki67 low/high. We chose the median value of Ki67 expression from Cohort 1 (low and high, <16 and ≥16) as the Ki67 cut-off.

6.5.2 Study III

6.5.2.1 Hormonal receptor, HER2 receptor

Cohort 1, ER and HER2 status of the tumours were taken from the clinicopathological data provided in the original study which were based on IHC. PR status was based on a gene expression classification as stated in the original report[284]. Cohort 2 of this study is the same as Cohort 1 in Study I (more information is provided in 6.5.1.1 section).

6.5.2.2 Genome profiling

Cohort 1, matched DNA and RNA extracted from patients were analysed using Illumina HT12-v3 platform and SNP arrays were run on Affymetrix microarray 6.0 platforms as described in the original study[284]. Cohort 2, RNA gene expression was carried out on customized HRSTA-2.0 Affymetrix array GPL10379[225]. DNA genotyping was performed on the Human1M-Duo BeadChip platform.

6.5.2.3 Copy number analysis

Cohort 1, data used in this study was downloaded from available online repositories as described in section 6.1 (Study III). Cohort 2, CNVs were generated using CNVpartition plugin within GenomeStudio software provided by Illumina (version 2011.1, Illumina, California, USA)[289]. The Circular Binary Segmentation (CBS) algorithm[290] was employed to generate CNV calls using DNAcopy[291] package in R statistical program. We used the Genomic Identification of Significant Targets in Cancer 2.0 (GISTIC) module to generate Copy Number Alteration (CNA) for *CCND1* gene in both cohorts. The GISTIC module gives an estimation about regions in the genome of samples that exhibit amplification and deletions. The G-Score provided in output of the algorithm shows the range and frequency of alterations across all samples[292].

6.5.3 Study IV

6.5.3.1 Mutation analysis

As described in section 6.1 (Study IV), all the data were derived from available online resources[282]. The mutational data was downloaded as a Mutational Annotation format (MAF) file. MAFtools[293] package in R program was used to extract mutational counts within the PanCan33 dataset. Additionally, the mutational burden was calculated in the following manner: the total number of mutations divided by the total number of tumours within each cancer type to avoid biased analysis that would have been caused by a large number of mutations in a single cancer type.

6.5.3.2 Chromosome arm-level alterations

Aneuploidy score is a numeric value representing the alterations along chromosomal arms in which, values of +1, 0 and -1 represent the gains, non-aneuploidy and losses, respectively[294]. Tumour samples were divided into subgroups of amplified and deleted cases using the Aneuploidy score. The final counts for alterations were calculated by dividing the total number of losses/gains by the total number of tumours within each cancer type to avoid the same biases mentioned in the previous section.

7 RESULTS AND DISCUSSIONS

7.1 STUDY I

Comparable prognostic capacity of Ki67 measured by IHC and bivariate gene expression signatures has been demonstrated in several studies[265–267]. However, this correlation becomes weak between Ki67 and multivariate gene expression signatures since some tumour samples may be differentially classified using Ki67 or multivariate prognostic gene signatures[265], respectively. Considering the discordance between tumours being classified as good prognosis by Ki67 or poor prognosis by multivariate gene expression signatures, we examined if the combination of Ki67/IHC subtypes, and the research version of gene expression signatures (GGI, 70-Gene, RS, CCS and PAM50) can provide more prognostic information than either classifier alone.

This assessment was performed on four clinically relevant subgroups of patients including All patients (n = 379, n = 209), ER+/LN- (n = 104, n = 115), ER+/LN+ (n = 167, n = 65) and ER- patients (n = 103, n = 24) of Cohort 1 and 2, respectively. However, we expected to have more aggressive tumours in Cohort 1 due to its "nested case-control design" nature, as described in section 6.2.1. Notably, we observed approximately 14-22% (in Cohort 1) and 11-28% (in Cohort 2) discordance in tumour prognostic classification using Ki67 vs gene expression signatures.

We took two approaches to examine our hypothesis using LRχ2 testing: First, we combined Ki67/IHC subtypes with gene expression signatures to evaluate additional prognostic capacity of the gene expression signatures beyond Ki67/IHC subtypes. Next, we reversed the combination and assessed the additional prognostic capacity of the Ki67/IHC subtypes beyond gene expression signatures. In all patients of Cohort 1, RS and PAM50 signatures provided statistically significant prognostic information beyond Ki67 (ΔLRχ2 RS = 12.8 and PAM50 = 20.7; P = 0.001 and $P < 0.001$, respectively) and IHC subtypes (ΔLRχ2 RS = 12.9 and PAM50 = 11.7; P = 0.001 and P = 0.020, respectively). A similar trend was observed in the ER+/LN- subgroup. In the ER+/LN+ subgroup, all gene expression signatures added prognostic information beyond Ki67 and IHC subtypes. This may indicate that gene expression signatures can capture additional biological information in this subgroup of patients which was not provided by the clinical IHC markers. Moreover, all gene expression signatures lost their prognostic power in the ER- subgroup of patients.

In the reverse analysis, IHC subtypes added prognostic information on the top of all gene expression signatures except PAM50 in All and ER- patients (All patients, PAM50/IHC subtypes ΔLRχ2 = 7.1, P = 0.068, ER- patients ΔLRχ2 = 2.1, P = 0.146), respectively.

Similar results were observed in Cohort 2 despite a lower number of patients compared to Cohort 1.
The results from Study I showed the prognostic superiority of PAM50 signature compared to IHC subtypes. We have also shown that IHC subtypes can compete well against most of prognostic gene expression classifiers.

7.2 STUDY II

Although gene expression signatures have been demonstrated to provide prognostic and treatment predictive information in primary breast tumours, their prognostic capacity has not been tested in MBC patients. In **Study II** we explored the prognostic strength of some clinically relevant gene expression signatures in the metastatic setting.
First, we tested the ability of GGI, 70-Gene, CCS and PAM50 gene expression signatures to predict post-relapse BCSS survival using KM and Cox regression analyses. Across all gene expression signatures, only PAM50 provided significant prognostic information when all metastatic sites were included in the analyses (univariate analysis, Short-term BCSS, HER2-enriched:HR = 3.0, 95% CI, 1.2 – 7.4, basal-like:HR = 3.2, 95% CI, 1.3 – 8.0, Long-term BCSS, HER2-enriched:HR = 1.9; 95% CI, 1.1 – 3.4). Our results remained significant in multivariate analysis after adjusting our models for age at diagnosis, diagnosis date and clinical treatment received, in which luminal A subtype tumours had the best prognosis relative to luminal B, HER2-enriched and basal-like subtype tumours (Short-term BCSS, HER2-enriched:HR = 3.2, 95% CI, 1.3 – 7.9, basal-like:HR = 3.0, 95% CI, 1.2 – 7.5, Long-term BCSS, HER2-enriched:HR = 1.9; 95% CI, 1.1 – 3.3). Next, we examined the capability of PAM50 gene expression signature to predict post-relapse from LN biopsies. In multivariate analyses basal-like and HER2-enriched (long-term BCSS) subtypes provided statistically significant information (Short-term BCSS, basal-like:HR = 2.6, 95% CI, 2.7 – 247.5; Long-term BCSS, HER2-enriched:HR = 3.7; 95% CI, 1.2 – 11.6; basal-like:HR = 7.9; 95% CI, 2.2 – 28.2).
Finally, LRχ2 test was used to determine the capacity of gene expression signatures to provide prognostic information in MBC patients. The PAM50 was the only signature that provided prognostic information in the metastatic setting (LRχ2: long-term = 20.0 and 10.4; $P < 0.001$ and $P = 0.015$ in LN and Other metastatic sites, respectively).

In **Study II**, we showed that the majority of gene expression signatures classify metastatic tumours as poor prognosis. Notably, PAM50 can provide prognostic information from LN metastases in MBC patients.

7.3 STUDY III

Several biomarkers have been identified that are capable of selecting breast cancer patients with poor long-term survival[232, 265]. Additionally, the clinical utility of gene expression signatures has been proven in recent studies[225, 264].

Cyclin D1 protein has been demonstrated to be overexpressed in more than half of breast cancer tumours in which higher level of cyclin D1 protein expression has been correlated to both good and poor prognosis[36, 295–297]. On the other hand, the amplification of its corresponding gene *CCND1* has consistently been associated with poor prognosis in breast cancer patients[36, 37, 298]. In this study, we aimed to clarify the prognostic and predictive capacity of *CCND1* gene amplification, thus promoting its use as a clinically relevant biomarker for long term BCSS.

Twenty-two (426/1965 22%) and thirty-five percent of tumours in <u>Cohort 1</u> and <u>Cohort 2</u>(119/340, 35%) were found to be *CCND1*-amplified, respectively. A higher number of *CCND1*-amplified tumours was observed in the more aggressive <u>Cohort 2</u>. In general, the majority of the *CCND1*-amplified tumours were ER+ or luminal subtype tumours relative to non-amplified tumours (Cohort 1 ER+ = 88% vs 73%, luminal A or B = 73% vs 58%, Amp/non-Amp, respectively).

In <u>Cohort 1</u>, patients with *CCND1*-amplified tumours showed worse long-term BCSS in subgroups of ER+/LN-/HER2- and ER+/LN+/HER2- (log rank $P < 0.001$ and $P = 0.016$, respectively) as well as luminal A subtype tumours in addition to endocrine treated and untreated patients (log rank $P = 0.019$, 0.007 and 0.014, respectively). In the multivariate Cox regression analyses, the results remained statistically significant in ER+/LN-/HER2- subgroup (HR = 1.72, 95% CI, 1.14 – 2.59) with comparable trends for luminal A patients.

We have observed similar results combining both cohorts 1 and 2. In the combined cohort, patients were found to have poorer 15-years BCSS in multivariate analysis (ER+/LN-/HER2-, HR = 1.66, 95% CI, 1.14 – 2.41; luminal A, HR = 1.68, 95% CI, 1.15 – 2.46; luminal B, HR = 1.37, 95% CI, 1.01 – 1.86, respectively). These results highlight the prognostic potential of *CCND1*-amplification status to stratify patients with poorer 15-years BCSS survival. In the next step, the expression of genes related to the cell cycle and cell proliferation was examined within PAM50 subtypes in <u>Cohort 1</u>.

We demonstrated that luminal A/*CCND1*-Amplified tumours have higher expression of *MKI67* (proliferation marker gene) and lower *PGR* expression relative to luminal A/non-Amplified tumours; this could be the reason that luminal A/*CCND1*-Amplified tumours show poorer 15-years BCSS relative to luminal A/non-Amplified tumours since they show similar characteristics as more aggressive luminal B subtype tumours[299].

It has been previously suggested that *CCND1*-amplification could act as a biomarker for predicting patients who may respond to CDK 4/6 inhibitors[6]. Given that, we explored the expression of *CDK4* and *CDK6* genes within *CCND1*-amplfied and non-Amplified tumours; our data from the gene expression analysis implied that *CCND1*-amplification would perform poorly as a predictive biomarker in this setting which are in agreement with the findings of PALOMA-1/TRIO-18 randomised phase II trial showing patients with *CCND1*-amplified tumours do not benefit from the CDK4/6 inhibitor Palbociclib[300].

Finally, differential gene expression analysis was performed between *CCND1*-amplfieid and non-amplified tumours in luminal A and B subtypes, separately. Notably, these two PAM50 subtypes share common genes within *CCND1*-amplfied tumours.

7.4 STUDY IV

Genomic alterations are main causes of loss of control over the cell cycle which is a hallmark of cancer[2]. In **Study IV**, we explored the genomic alterations including DNA mutations and chromosomal arm-level aneuploidy among and within tumours with low, intermediate and high cell cycle activity using the CCS signature[225]. The chromosomal arm-level aneuploidy was derived from aneuploidy score[294] which represents a total number of chromosomal changes at an arm-level in a given sample.

We demonstrated that cell cycle activity varies a lot among cancer types with KICH, PCPG, KIRP and PRAD tumours exhibiting lowest and TGCT, DLBC, HNSC and CESC tumours highest level of cell activity among 33 cancer types. Furthermore, we examined DNA mutations of 299 well defined onco- and tumour suppressor driver genes in relation to cell cycle activity among (inter) and within each cancer type (intra)[301].

TP53 and *PIK3CA* had the highest mutation rate within subgroups of high and intermediate inter-CCS while *KRAS* and *BRAF* mutations were prominent in the low and intermediate inter-CCS. The intra-CCS subgroups were also dominated by *TP53* and *PIK3CA* mutations in all subgroups of low, intermediate and high intra-CCS. *BRAF* mutations were among top 15 genes with highest mutation rate in high intra-CCS. These observations imply that

mutations of the aforementioned genes are not only associated with highly proliferative cancers, despite their direct association with cell cycle progression[21, 302].

The finding from chromosomal arm-level gains/losses analyses showed that the aneuploidy score significantly increases with increasing cell cycle activity among cancers (inter-CCS) but not in intra-CCS level. Higher level of amplification was observed in chromosomes 7p, 20q and 8q in the subgroups of inter/intra-CCS while the most frequent deletions were observed in 17q arm followed by 8p and 18q in the similar subgroups. Next, we assessed the PFI using KM analysis between inter- and intra-CCS groups. The inter-CCS had prognostic capacity whereby patients with low inter-CCS had longer PFI relative to intermediate or high inter-CCS groups ($P < 0.001$); while no difference in PFI was found between intra-CCS groups. Finally, inter-/intra-CCS were tested in multivariate Cox regression models. In multivariate analysis inter-CCS remained statistically significant after adjustments for tumour type, age, gender, pathological grade and radiotherapy demonstrating its capacity to provide independent prognostic information.

8 CONCLUSIONS

8.1 STUDY I

Our findings in this study indicate that RS and PAM50 signatures can provide more prognostic information compared to the IHC subtypes in all breast cancer patients. Notably, IHC subtypes did not add prognostic value to the PAM50 signature.

8.2 STUDY II

We showed for the first time that PAM50 signature provides prognostic information when applied to LN metastases and could potentially be used in the metastatic setting to aid treatment decisions.

8.3 STUDY III

We demonstrated that amplification of *CCND1* is correlated with poorer 15-year BCSS survival in patients with ER+/LN-/HER2-, luminal A and luminal B subtype tumours. Furthermore, luminal A tumours exhibiting amplification of *CCND1* have similar gene expression changes as tumours with a luminal B subtype. Our findings emphasize the potential of *CCND1* amplification to be used as a biomarker to identify patients who may benefit from aggressive treatment strategies.

8.4 STUDY IV

The results of this study indicate that DNA alterations are presented at all cell cycle activity levels, whilst some significant exceptions exist such as *KRAS* and *BRAF* mutations. We also demonstrated that cell cycle activity deviates extensively across and within cancers. Finally, our data show that a simple gene expression signature representing cell cycle activity can provide independent prognostic information in different cancer types.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to:

Karolinska Institutet for giving me the opportunity to work in an amazing scientific environment and the **Department of Oncology and Pathology** for creating a friendly and interactive workplace to research and collaborate with professional clinicians and biologists.

Supervisors

This amazing four years would not have been possible without the support and scientific assistance of my brilliant supervisors.

I am more than grateful for the privilege of having **Associate Prof. Nicholas P. Tobin** as my main supervisor. I cannot thank you enough for your patience, guidance and encouragements during this memorable journey. You are truly the best mentor someone could wish for; I am not only proud of the work we have accomplished together, but I have also learnt a lot from you about science, critical thinking and more. I appreciate you *always* being there helping and supporting me. Thank you, Nick, for making this journey as perfect as it can be.
It was my great pleasure working with **Professor Jonas Bergh** during my Ph.D. I am so humbled and honored for being part your team. I was always amazed at your vast knowledge and be willing to share your insights with others. I am so thankful for your constructive feedback and support. You are truly a great inspiration for me.
I feel so fortunate to have **Associate Prof. Linda S. Lindström** as my co-supervisor. I have learned a lot about biostatistics and statistical thinking from you. Thank you for your help and sharing your knowledge with me during these years. You have been the nicest supervisor I have ever had.
I would like to thank **Associate Prof. Johan Hartman** and **Associate Prof. Ola Larsson** for being my co-supervisors. Your valuable inputs helped me very much over past four years.

Collaborators

I am so thankful for the opportunity to collaborate with **Associate Prof. Theodoros Foukakis** and his clinical feedbacks.
I am so grateful for my collaboration with **Professor. Charles M. Perou** over the past four years. Thank you for your inputs which certainly added a great scientific value to my work. It was a great pleasure meeting you and your team in person.

Friends and colleagues

Amandine Anastácio: I have a lot to say but I am supposed to keep it short ☺ . I don't know how these years would be like if I didn't have a lovely friend like you. Thanks for a

great company, great food “Bacalhau com natas”, happy and sad moments, all together. You have been an amazing support at my down times. I feel so lucky to have you in my life. Keep smiling and spread your positive energy wherever you go. I should definitely thank **Associate Prof. Kenny Rodriguez-Wallberg** for having you here! Muito obrigado minha doce Amandine! **Ikram Ullah:** It was my pleasure to have a colleague and friend like you. I have learnt a lot during the time working together. Thanks for your positive and friendly attitude. I wish the best for you and your lovely family. **Claudette Falato:** Thanks for the dramas, salsa parties and gossips. I hope you are enjoying the new chapter of your life. **John Lövrot:** Thanks for being a great colleague and an amazing scientist and for your comments on my work. **Julie Lorent:** I was more than lucky to work with you and get your help when I’ve just started working with the group. You make science a great pleasure with your positive and helpful attitude. **Helen Eriksson:** I am so thankful for your administrative support and being there with a smile and friendly attitude whenever I bothered you despite your busy days. **Xia Han:** Thanks for happy days, laughs and interesting foods; best of luck with your Ph.D. **Emmanouil Sifakis:** Thanks for the scientific discussions, friendly talks, laughs and tasty Greek fikas. **Parisa Rabei Far, Magali Merrien, Ksenia Goroshchuk** and **Gonzalo Fernandez Lahore:** I am not sure if I should thank you or Barbro…! ☺ It was my lucky day attending the “Pathology course” and meeting you there. I am so glad having friends like you and I hope our friendship will become even better after this chapter. Btw, good luck with your projects! **Aafke Duinmeijer** and **Fleur Wiggenraad:** Despite the short time we have been working together, I think we made a great bond which I am sure will get stronger. Thanks for all the laughs and stroopwafels! #Dutchbros **Vasiliki Arapi** and **Ravi Saini:** Thank you guys for lunch/fika times. Thanks, Valia for your delicious cookies! ☺
Maryam: I cannot find words to describe how grateful I am for having you as my sister. You are the one who believed in me and pushed me forward along this journey. **Behrooz:** Nobody could understand how much you have helped me to believe in myself during sad and hard times. Thanks for being there and for all the great times we have had together over past 16+ years. Thank you again for offering your help and creating this beautiful cover for my thesis book. **Rozita:** I appreciate your frankness and the slap when I needed it; you have definitely changed my life for better. **Magnus:** I’d have never thought that we could come this far this good together. I feel so loved and blessed for having you in my life. Thank you for being there, being patient and dealing with my dramas. I couldn’t have asked for more. **Gunnar** och **Gudrun:** Tack för att ni har blivit min nya familj och för att ni är så snälla och de finaste föräldrarna på hela västkusten.

<u>My family back in Iran</u>

با اینکه خیلی از هم دور هستیم اما **مامان و بابا** ممنون از همه انرژی‌های مثبتی که از راه دور برام می‌فرستید. امیدوارم از این اتفاق خوشحال شده باشید. **ارحام،حسام و ندا (و البته رایمون!)** مرسی از همفکری‌ها و هم‌صحبتی‌ها و انژری‌های خوبی که بهم میفرستین و از اینکه من را تو ذهن رایمون نگه‌داشتید با همه فاصله که باهم داریم.

It is impossible to thank everyone that helped me during this journey, so to those whom I have not specifically named, I give thanks for your advice, support and contribution in this work.

REFERENCES

1. Hanahan D, Weinberg RA: The Hallmarks of Cancer [Internet]. Cell 100:57–70, 2000[cited 2017 Jan 24] Available from: //www.sciencedirect.com/science/article/pii/S0092867400816839

2. Hanahan D, Weinberg RA: Hallmarks of Cancer: The Next Generation [Internet]. Cell 144:646–674, 2011[cited 2017 Jan 24] Available from: //www.sciencedirect.com/science/article/pii/S0092867411001279

3. Cooper GM, Hausman RE: The cell: a molecular approach Seventh edition. Sunderland, Massachusetts, U.S.A, Sinauer Associates, Inc., Publishers, 2016

4. Cooper GM: Elements of Human Cancer. Jones & Bartlett Learning, 1992

5. Deshpande A, Sicinski P, Hinds PW: Cyclins and cdks in development and cancer: a perspective [Internet]. Oncogene 24:2909–2915, 2005[cited 2017 Jan 20] Available from: http://www.nature.com/onc/journal/v24/n17/full/1208618a.html

6. Musgrove EA, Caldon CE, Barraclough J, et al: Cyclin D as a therapeutic target in cancer [Internet]. Nature Reviews Cancer 11:558–572, 2011[cited 2018 Oct 21] Available from: http://www.nature.com/articles/nrc3090

7. Hartwell LH, Culotti J, Pringle JR, et al: Genetic Control of the Cell Division Cycle in Yeast: A model to account for the order of cell cycle events is deduced from the phenotypes of yeast mutants [Internet]. Science 183:46–51, 1974[cited 2018 Nov 14] Available from: http://science.sciencemag.org/content/183/4120/46

8. Lee MG, Nurse P: Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2 [Internet]. Nature 327:31–35, 1987[cited 2018 Nov 14] Available from: http://www.nature.com/articles/327031a0

9. Evans T, Rosenthal ET, Youngblom J, et al: Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division [Internet]. Cell 33:389–396, 1983[cited 2018 Nov 14] Available from: http://www.sciencedirect.com/science/article/pii/0092867483904208

10. Alberts B: Molecular biology of the cell Sixth edition. New York, NY, Garland Science, Taylor and Francis Group, 2015

11. Malumbres M, Barbacid M: Cell cycle, CDKs and cancer: a changing paradigm [Internet]. Nat Rev Cancer 9:153–166, 2009[cited 2017 Jan 20] Available from: http://www.nature.com/nrc/journal/v9/n3/full/nrc2602.html

12. Knudsen KE, Diehl JA, Haiman CA, et al: Cyclin D1: polymorphism, aberrant splicing and cancer risk [Internet]. Oncogene 25:1620–1628, 2006[cited 2017 Jan 31] Available from: http://www.nature.com/onc/journal/v25/n11/full/1209371a.html

13. Hwang HC, Clurman BE: Cyclin E in normal and neoplastic cell cycles [Internet]. Oncogene 24:2776–2786, 2005[cited 2017 Jan 31] Available from: http://www.nature.com/onc/journal/v24/n17/full/1208613a.html

14. Spruck CH, Won K-A, Reed SI: Deregulated cyclin E induces chromosome instability [Internet]. Nature 401:297–300, 1999[cited 2017 Jan 31] Available from: http://www.nature.com/nature/journal/v401/n6750/full/401297a0.html

15. Husdal A, Bukholm G, Bukholm IRK: The prognostic value and overexpression of cyclin A is correlated with gene amplification of both cyclin A and cyclin E in breast cancer patient. Cell Oncol 28:107–116, 2006

16. De Boer L, Oakes V, Beamish H, et al: Cyclin A/cdk2 coordinates centrosomal and nuclear mitotic events [Internet]. Oncogene 27:4261–4268, 2008[cited 2017 Jan 31] Available from: http://www.nature.com/onc/journal/v27/n31/full/onc200874a.html

17. Aaltonen K, Amini R-M, Heikkilä P, et al: High cyclin B1 expression is associated with poor survival in breast cancer [Internet]. Br J Cancer 100:1055–1060, 2009[cited 2017 Jan 31] Available from: http://www.nature.com/bjc/journal/v100/n7/full/6604874a.html

18. Yuan J, Yan R, Krämer A, et al: Cyclin B1 depletion inhibits proliferation and induces apoptosis in human tumor cells [Internet]. Oncogene 23:5843–5852, 2004[cited 2017 Jan 31] Available from: http://www.nature.com/onc/journal/v23/n34/full/1207757a.html

19. Sancar A, Lindsey-Boltz LA, Unsal-Kaçmaz K, et al: Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem 73:39–85, 2004

20. Tanaka TU, Desai A: Kinetochore-microtubule interactions: the means to the end. Curr Opin Cell Biol 20:53–63, 2008

21. Chen J: The Cell-Cycle Arrest and Apoptotic Functions of p53 in Tumor Initiation and Progression. Cold Spring Harb Perspect Med 6:a026104, 2016

22. Fridman JS, Lowe SW: Control of apoptosis by p53. Oncogene 22:9030–9040, 2003

23. Li Z, Wang C, Jiao X, et al: Cyclin D1 regulates cellular migration through the inhibition of thrombospondin 1 and ROCK signaling. Mol Cell Biol 26:4240–4256, 2006

24. Shan J, Zhao W, Gu W: Suppression of cancer cell growth by promoting cyclin D1 degradation. Mol Cell 36:469–476, 2009

25. Bendris N, Lemmers B, Blanchard JM: Cell cycle, cytoskeleton dynamics and beyond: the many functions of cyclins and CDK inhibitors. Cell Cycle 14:1786–1798, 2015

26. Nelsen CJ, Kuriyama R, Hirsch B, et al: Short term cyclin D1 overexpression induces centrosome amplification, mitotic spindle abnormalities, and aneuploidy. J Biol Chem 280:768–776, 2005

27. Butt AJ, McNeil CM, Musgrove EA, et al: Downstream targets of growth factor and oestrogen signalling and endocrine resistance: the potential roles of c-Myc, cyclin D1 and cyclin E. Endocr Relat Cancer 12 Suppl 1:S47-59, 2005

28. Prall OW, Rogan EM, Musgrove EA, et al: c-Myc or cyclin D1 mimics estrogen effects on cyclin E-Cdk2 activation and cell cycle reentry. Mol Cell Biol 18:4499–4508, 1998

29. Jirawatnotai S, Hu Y, Michowski W, et al: A function for cyclin D1 in DNA repair uncovered by protein interactome analyses in human cancers. Nature 474:230–234, 2011

30. Reutens AT, Fu M, Wang C, et al: Cyclin D1 binds the androgen receptor and regulates hormone-dependent signaling in a p300/CBP-associated factor (P/CAF)-dependent manner. Mol Endocrinol 15:797–811, 2001

31. Sabbah M, Courilleau D, Mester J, et al: Estrogen induction of the cyclin D1 promoter: involvement of a cAMP response-like element. Proc Natl Acad Sci USA 96:11217–11222, 1999

32. Ishizuka T, Tanabe C, Sakamoto H, et al: Gene amplification profiling of esophageal squamous cell carcinomas by DNA array CGH. Biochem Biophys Res Commun 296:152–155, 2002

33. Nagasawa S, Onda M, Sasajima K, et al: Cyclin D1 overexpression as a prognostic factor in patients with esophageal carcinoma. J Surg Oncol 78:208–214, 2001

34. Mishina T, Dosaka-Akita H, Kinoshita I, et al: Cyclin D1 expression in non-small-cell lung cancers: its association with altered p53 expression, cell proliferation and clinical outcome. Br J Cancer 80:1289–1295, 1999

35. Peurala E, Koivunen P, Haapasaari K-M, et al: The prognostic significance and value of cyclin D1, CDK4 and p16 in human breast cancer. Breast Cancer Res 15:R5, 2013

36. Lundgren K, Brown M, Pineda S, et al: Effects of cyclin D1 gene amplification and protein expression on time to recurrence in postmenopausal breast cancer patients treated with anastrozole or tamoxifen: a TransATAC study [Internet]. Breast Cancer Research 14:R57, 2012[cited 2018 Mar 14] Available from: https://doi.org/10.1186/bcr3161

37. Ortiz AB, Garcia D, Vicente Y, et al: Prognostic significance of cyclin D1 protein expression and gene amplification in invasive breast carcinoma. PLoS ONE 12:e0188068, 2017

38. Stendahl M, Kronblad A, Rydén L, et al: Cyclin D1 overexpression is a negative predictive factor for tamoxifen response in postmenopausal breast cancer patients. Br J Cancer 90:1942–1948, 2004

39. World Health Organization (WHO) - Breast Cancer 2018 [Internet]. World Health Organization (WHO) [cited 2018 Nov 29] Available from: http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/

40. Bray F, Ferlay J, Soerjomataram I, et al: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [Internet]. CA: A Cancer Journal for Clinicians 68:394–424, 2018[cited 2018 Nov 29] Available from: http://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21492

41. Engholm G, Ferlay J, Christensen N, et al: NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol 49:725–736, 2010

42. Swedish Cancer Registry [Internet][cited 2018 Dec 3] Available from: http://www.socialstyrelsen.se/register/halsodataregister/cancerregistret/inenglish

43. Statistics on Cancer Incidence in Sweden 2016 [Internet]Available from: https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/20787/2017-12-31.pdf

44. Seitz HK, Stickel F: Molecular mechanisms of alcohol-mediated carcinogenesis [Internet]. Nat Rev Cancer 7:599–612, 2007[cited 2017 Jan 31] Available from: http://www.nature.com/nrc/journal/v7/n8/full/nrc2191.html

45. Schuller HM: Mechanisms of smoking-related lung and pancreatic adenocarcinoma development [Internet]. Nat Rev Cancer 2:455–463, 2002[cited 2017 Jan 31] Available from: http://www.nature.com/nrc/journal/v2/n6/abs/nrc824.html

46. Kant P, Hull MA: Excess body weight and obesity|[mdash]|the link with gastrointestinal and hepatobiliary cancer [Internet]. Nature Reviews Gastroenterology and Hepatology 8:224–238, 2011[cited 2017 Jan 31] Available from: http://www.nature.com/nrgastro/journal/v8/n4/pdf/nrgastro.2011.23.pdf%3FWT.ec_id%3DNRGASTRO-201104

47. Stein CJ, Colditz GA: Modifiable risk factors for cancer [Internet]. Br J Cancer 90:299–303, 2004[cited 2017 Jan 20] Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2410150/

48. Barnett GC, Shah M, Redman K, et al: Risk Factors for the Incidence of Breast Cancer: Do They Affect Survival From the Disease? [Internet]. JCO 26:3310–3316, 2008[cited 2017 Jan 20] Available from: http://ascopubs.org/doi/abs/10.1200/jco.2006.10.3168

49. Dawson S-J, Rueda OM, Aparicio S, et al: A new genome-driven integrated classification of breast cancer and its implications [Internet]. EMBO J 32:617–628, 2013[cited 2017 Jan 20] Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3590990/

50. Leonard GD, Swain SM: Ductal Carcinoma In Situ , Complexities and Challenges [Internet]. JNCI J Natl Cancer Inst 96:906–920, 2004[cited 2017 Jan 20] Available from: https://academic.oup.com/jnci/article/96/12/906/2520795/Ductal-Carcinoma-In-Situ-Complexities-and

51. Espina V, Liotta LA: What is the malignant nature of human ductal carcinoma in situ? [Internet]. Nat Rev Cancer 11:68–75, 2011[cited 2017 Jan 20] Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3756606/

52. Weigelt B, Peterse JL, van't Veer LJ: Breast cancer metastasis: markers and models [Internet]. Nat Rev Cancer 5:591–602, 2005[cited 2017 Jan 20] Available from: http://www.nature.com/nrc/journal/v5/n8/full/nrc1670.html

53. Rakha EA, Patel A, Powe DG, et al: Clinical and biological significance of E-cadherin protein expression in invasive lobular carcinoma of the breast. Am J Surg Pathol 34:1472–1479, 2010

54. Schnitt SJ: Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy [Internet]. Mod Pathol 23:S60–S64, 2010[cited 2017 Jan 20] Available from: http://www.nature.com/modpathol/journal/v23/n2s/full/modpathol201033a.html

55. Dellaire G, Berman JN, Arceci R (eds): Cancer genomics: from bench to personalized medicine. Amsterdam, Elsevier, AP, Academic Press is an imprint of Elsevier, 2014

56. Edge SB, Compton CC: The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 17:1471–1474, 2010

57. Amin MB, Greene FL, Edge SB, et al: The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin 67:93–99, 2017

58. Cianfrocca M, Goldstein LJ: Prognostic and Predictive Factors in Early-Stage Breast Cancer [Internet]. The Oncologist 9:606–616, 2004[cited 2017 Jan 20] Available from: http://theoncologist.alphamedpress.org/content/9/6/606

59. Allred DC, Harvey JM, Berardo M, et al: Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Mod Pathol 11:155–168, 1998

60. Davies C, Pan H, Godwin J, 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 [Internet]. The Lancet 381:805–816, 2013[cited 2018 Nov 18] Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)61963-1/abstract

61. Dowsett M, Cuzick J, Wale C, et al: Retrospective analysis of time to recurrence in the ATAC trial according to hormone receptor status: an hypothesis-generating study. J Clin Oncol 23:7512–7517, 2005

62. Iwamoto T, Booser D, Valero V, et al: Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. J Clin Oncol 30:729–734, 2012

63. Yi M, Huo L, Koenig KB, et al: Which threshold for ER positivity? a retrospective study based on 9639 patients. Ann Oncol 25:1004–1011, 2014

64. Mohammed H, Russell IA, Stark R, et al: Progesterone receptor modulates ERα action in breast cancer. Nature 523:313–317, 2015

65. Purdie CA, Quinlan P, Jordan LB, et al: Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. Br J Cancer 110:565–572, 2014

66. Bardou V-J, Arpino G, Elledge RM, et al: Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. J Clin Oncol 21:1973–1979, 2003

67. Tan M, Yu D: Molecular Mechanisms of ErbB2-Mediated Breast Cancer Chemoresistance [Internet]. Landes Bioscience, 2013[cited 2018 Nov 30] Available from: http://www.ncbi.nlm.nih.gov/books/NBK6194/

68. Dean L: Trastuzumab (Herceptin) Therapy and ERBB2 (HER2) Genotype [Internet], in Pratt V, McLeod H, Rubinstein W, et al (eds): Medical Genetics Summaries. Bethesda (MD), National Center for Biotechnology Information (US), 2012[cited 2018 Nov 30] Available from: http://www.ncbi.nlm.nih.gov/books/NBK310376/

69. Beresford MJ, Wilson GD, Makris A: Measuring proliferation in breast cancer: practicalities and applications. Breast Cancer Res 8:216, 2006

70. Scholzen T, Gerdes J: The Ki-67 protein: from the known and the unknown. J Cell Physiol 182:311–322, 2000

71. Azambuja E de, Cardoso F, Jr G de C, et al: Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients [Internet]. British Journal of Cancer 96:1504–1513, 2007[cited 2018 Apr 26] Available from: https://www.nature.com/articles/6603756

72. Inwald EC, Klinkhammer-Schalke M, Hofstädter F, et al: Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry [Internet]. Breast Cancer Res Treat 139:539–552, 2013[cited 2017 Jan 20] Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3669503/

73. Cuylen S, Blaukopf C, Politi AZ, et al: Ki-67 acts as a biological surfactant to disperse mitotic chromosomes [Internet]. Nature 535:308–312, 2016[cited 2018 Nov 30] Available from: http://www.nature.com/articles/nature18610

74. Criscitiello C, Disalvatore D, De Laurentiis M, et al: High Ki-67 score is indicative of a greater benefit from adjuvant chemotherapy when added to endocrine therapy in Luminal B HER2 negative and node-positive breast cancer [Internet]. The Breast 23:69–75, 2014[cited 2018 Nov 30] Available from: https://linkinghub.elsevier.com/retrieve/pii/S0960977613002968

75. Dowsett M, Nielsen TO, A'Hern R, et al: Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst 103:1656–1664, 2011

76. Coates AS, Winer EP, Goldhirsch A, et al: Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol 26:1533–1546, 2015

77. Rouzier R, Perou CM, Symmans WF, et al: Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 11:5678–5685, 2005

78. Prat A, Bianchini G, Thomas M, et al: Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. Clin Cancer Res 20:511–521, 2014

79. Cheang MCU, Voduc KD, Tu D, et al: Responsiveness of intrinsic subtypes to adjuvant anthracycline substitution in the NCIC.CTG MA.5 randomized trial. Clin Cancer Res 18:2402–2412, 2012

80. Hayes DF, Thor AD, Dressler LG, et al: HER2 and Response to Paclitaxel in Node-Positive Breast Cancer [Internet]. New England Journal of Medicine 357:1496–1506, 2007[cited 2017 Jan 31] Available from: http://dx.doi.org/10.1056/NEJMoa071167

81. Goldhirsch A, Winer EP, Coates AS, et al: Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013 [Internet]. Ann Oncol 24:2206–2223, 2013[cited 2017 Jan 20] Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755334/

82. Badve S, Dabbs DJ, Schnitt SJ, et al: Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists [Internet]. Mod Pathol 24:157–167, 2011[cited 2017 Jan 30] Available from: http://www.nature.com/modpathol/journal/v24/n2/full/modpathol2010200a.html

83. Dai X, Xiang L, Li T, et al: Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes [Internet]. J Cancer 7:1281–1294, 2016[cited 2017 Jan 20] Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4934037/

84. Boyle P: Triple-negative breast cancer: epidemiological considerations and recommendations [Internet]. Ann Oncol 23:vi7–vi12, 2012[cited 2017 Jan 20] Available from: https://academic.oup.com/annonc/article/23/suppl_6/vi7/173019/Triple-negative-breast-cancer-epidemiological

85. Bianchini G, Balko JM, Mayer IA, et al: Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease [Internet]. Nat Rev Clin Oncol 13:674–690, 2016[cited 2017 Jan 20] Available from: http://www.nature.com/nrclinonc/journal/v13/n11/abs/nrclinonc.2016.66.html

86. Huo D, Ikpatt F, Khramtsov A, et al: Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. J Clin Oncol 27:4515–4521, 2009

87. Perou CM, Sørlie T, Eisen MB, et al: Molecular portraits of human breast tumours [Internet]. Nature 406:747–752, 2000[cited 2017 Jan 20] Available from: http://www.nature.com/nature/journal/v406/n6797/full/406747a0.html

88. Sørlie T, Perou CM, Tibshirani R, et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications [Internet]. PNAS 98:10869–10874, 2001[cited 2017 Jan 31] Available from: http://www.pnas.org/content/98/19/10869

89. Hu Z, Fan C, Oh DS, et al: The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7:96, 2006

90. Parker JS, Mullins M, Cheang MCU, et al: Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes [Internet]. JCO 27:1160–1167, 2009[cited 2015 Aug 12] Available from: http://jco.ascopubs.org/content/27/8/1160

91. Ohnstad HO, Borgen E, Falk RS, et al: Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up. Breast Cancer Res 19:120, 2017

92. Liu MC, Pitcher BN, Mardis ER, et al: PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). NPJ Breast Cancer 2, 2016

93. Tobin NP, Lundberg A, Lindström LS, et al: PAM50 Provides Prognostic Information When Applied to the Lymph Node Metastases of Advanced Breast Cancer Patients. Clin Cancer Res 23:7225–7231, 2017

94. Caan BJ, Sweeney C, Habel LA, et al: Intrinsic subtypes from the PAM50 gene expression assay in a population-based breast cancer survivor cohort: prognostication of short- and long-term outcomes. Cancer Epidemiol Biomarkers Prev 23:725–734, 2014

95. Arriagada R, Le MG, Dunant A, et al: Twenty-five years of follow-up in patients with operable breast carcinoma: correlation between clinicopathologic factors and the risk of death in each 5-year period. Cancer 106:743–750, 2006

96. Weiss RB, Woolf SH, Demakos E, et al: Natural history of more than 20 years of node-positive primary breast carcinoma treated with cyclophosphamide, methotrexate, and fluorouracil-based adjuvant chemotherapy: a study by the Cancer and Leukemia Group B. J Clin Oncol 21:1825–1835, 2003

97. Fisher B, Bauer M, Wickerham DL, et al: Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update. Cancer 52:1551–1557, 1983

98. Tang C, Wang P, Li X, et al: Lymph node status have a prognostic impact in breast cancer patients with distant metastasis. PLoS ONE 12:e0182953, 2017

99. Gobardhan PD, Elias SG, Madsen EVE, et al: Prognostic value of lymph node micrometastases in breast cancer: a multicenter cohort study. Ann Surg Oncol 18:1657–1664, 2011

100. de Mascarel I, Bonichon F, Coindre JM, et al: Prognostic significance of breast cancer axillary lymph node micrometastases assessed by two special techniques: reevaluation with longer follow-up. Br J Cancer 66:523–527, 1992

101. Liao G-S, Chou Y-C, Hsu H-M, et al: The prognostic value of lymph node status among breast cancer subtypes. Am J Surg 209:717–724, 2015

102. Ullah I, Karthik G-M, Alkodsi A, et al: Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes [Internet]. J Clin Invest 128:1355–1370, 2018[cited 2018 Nov 14] Available from: https://www.jci.org/articles/view/96149

103. Siegel MB, He X, Hoadley KA, et al: Integrated RNA and DNA sequencing reveals early drivers of metastatic breast cancer [Internet]. J Clin Invest 128:1371–1383, 2018[cited 2018 Dec 10] Available from: https://www.jci.org/articles/view/96153#B38

104. Elkin EB, Hudis C, Begg CB, et al: The effect of changes in tumor size on breast carcinoma survival in the U.S.: 1975-1999. Cancer 104:1149–1157, 2005

105. Carter CL, Allen C, Henson DE: Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer 63:181–187, 1989

106. Rosen PP, Groshen S, Kinne DW: Prognosis in T2N0M0 stage I breast carcinoma: a 20-year follow-up study. J Clin Oncol 9:1650–1661, 1991

107. Fung F, Cornacchi SD, Vanniyasingam T, et al: Predictors of 5-year local, regional, and distant recurrent events in a population-based cohort of breast cancer patients [Internet]. The American Journal of Surgery 213:418–425, 2017[cited 2018 Nov 14] Available from: http://www.sciencedirect.com/science/article/pii/S0002961016302902

108. Amat S, Penault-Llorca F, Cure H, et al: Scarff-Bloom-Richardson (SBR) grading: a pleiotropic marker of chemosensitivity in invasive ductal breast carcinomas treated by neoadjuvant chemotherapy. Int J Oncol 20:791–796, 2002

109. Chollet P, Amat S, Belembaogo E, et al: Is Nottingham prognostic index useful after induction chemotherapy in operable breast cancer? Br J Cancer 89:1185–1191, 2003

110. Rakha EA, Martin S, Lee AHS, et al: The prognostic significance of lymphovascular invasion in invasive breast carcinoma [Internet]. Cancer 118:3670–3680, 2012[cited 2018 Nov 14] Available from: http://onlinelibrary.wiley.com/doi/abs/10.1002/cncr.26711

111. Song YJ, Shin SH, Cho JS, et al: The role of lymphovascular invasion as a prognostic factor in patients with lymph node-positive operable invasive breast cancer. J Breast Cancer 14:198–203, 2011

112. Ejlertsen B, Jensen M-B, Rank F, et al: Population-based study of peritumoral lymphovascular invasion and outcome among patients with operable breast cancer. J Natl Cancer Inst 101:729–735, 2009

113. Liu YL, Saraf A, Lee SM, et al: Lymphovascular Invasion is an Independent Predictor of Survival in Breast Cancer after Neoadjuvant Chemotherapy [Internet]. Breast Cancer Res Treat 157:555–564, 2016[cited 2018 Nov 14] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5337480/

114. Bergman L, Kluck HM, van Leeuwen FE, et al: The influence of age on treatment choice and survival of elderly breast cancer patients in south-eastern Netherlands: a population-based study. Eur J Cancer 28A:1475–1480, 1992

115. Høst H, Lund E: Age as a prognostic factor in breast cancer [Internet]. Cancer 57:2217–2221, 1986[cited 2018 Nov 30] Available from: http://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%2819860601%2957%3A11%3C2217%3A%3AAID-CNCR2820571124%3E3.0.CO%3B2-T

116. Chung M, Chang HR, Bland KI, et al: Younger women with breast carcinoma have a poorer prognosis than older women. Cancer 77:97–103, 1996

117. Maskarinec G, Sen C, Koga K, et al: Ethnic Differences in Breast Cancer Survival: Status and Determinants [Internet]. Womens Health (Lond Engl) 7:677–687, 2011[cited 2018 Nov 14] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3256927/

118. Pierce L, Fowble B, Solin LJ, et al: Conservative surgery and radiation therapy in black women with early stage breast cancer. Patterns of failure and analysis of outcome. Cancer 69:2831–2841, 1992

119. dos Santos Silva I, Mangtani P, De Stavola BL, et al: Survival from breast cancer among South Asian and non-South Asian women resident in South East England. Br J Cancer 89:508–512, 2003

120. Chang J, Powles TJ, Allred DC, et al: Prediction of clinical outcome from primary tamoxifen by expression of biologic markers in breast cancer patients. Clin Cancer Res 6:616–621, 2000

121. Miller WR, White S, Dixon JM, et al: Proliferation, steroid receptors and clinical/pathological response in breast cancer treated with letrozole. Br J Cancer 94:1051–1056, 2006

122. Viale G, Giobbie-Hurder A, Regan MM, et al: Prognostic and predictive value of centrally reviewed Ki-67 labeling index in postmenopausal women with endocrine-

responsive breast cancer: results from Breast International Group Trial 1-98 comparing adjuvant tamoxifen with letrozole. J Clin Oncol 26:5569–5575, 2008

123. Fisher B, Redmond C, Fisher ER, et al: Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. [Internet]. JCO 6:1076–1087, 1988[cited 2018 Nov 14] Available from: http://ascopubs.org/doi/abs/10.1200/JCO.1988.6.7.1076

124. Hilsenbeck SG, Ravdin PM, de Moor CA, et al: Time-dependence of hazard ratios for prognostic factors in primary breast cancer. Breast Cancer Res Treat 52:227–237, 1998

125. Katzorke N, Rack BK, Haeberle L, et al: Prognostic value of HER2 on breast cancer survival. [Internet]. JCO 31:640–640, 2013[cited 2018 Nov 14] Available from: http://ascopubs.org/doi/abs/10.1200/jco.2013.31.15_suppl.640

126. O'Shaughnessy J: Extending survival with chemotherapy in metastatic breast cancer. Oncologist 10 Suppl 3:20–29, 2005

127. Cardoso F, Costa A, Norton L, et al: 1st International consensus guidelines for advanced breast cancer (ABC 1) [Internet]. The Breast 21:242–252, 2012[cited 2018 Nov 15] Available from: http://www.sciencedirect.com/science/article/pii/S0960977612000628

128. Kennecke H, Yerushalmi R, Woods R, et al: Metastatic Behavior of Breast Cancer Subtypes [Internet]. JCO 28:3271–3277, 2010[cited 2018 Nov 15] Available from: http://ascopubs.org/doi/full/10.1200/JCO.2009.25.9820

129. Lobbezoo DJA, van Kampen RJW, Voogd AC, et al: Prognosis of metastatic breast cancer subtypes: the hormone receptor/HER2-positive subtype is associated with the most favorable outcome. Breast Cancer Res Treat 141:507–514, 2013

130. Dawood S, Broglio K, Buzdar AU, et al: Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. J Clin Oncol 28:92–98, 2010

131. Amir E, Ooi WS, Simmons C, et al: Discordance between Receptor Status in Primary and Metastatic Breast Cancer: an Exploratory Study of Bone and Bone Marrow Biopsies [Internet]. Clinical Oncology 20:763–768, 2008[cited 2018 Nov 15] Available from: http://www.sciencedirect.com/science/article/pii/S0936655508003609

132. Simmons C, Miller N, Geddie W, et al: Does confirmatory tumor biopsy alter the management of breast cancer patients with distant metastases? [Internet]. Ann Oncol 20:1499–1504, 2009[cited 2018 Nov 15] Available from: http://academic.oup.com/annonc/article/20/9/1499/217745

133. Thompson AM, Jordan LB, Quinlan P, et al: Prospective comparison of switches in biomarker status between primary and recurrent breast cancer: the Breast Recurrence In

Tissues Study (BRITS) [Internet]. Breast Cancer Research 12:R92, 2010[cited 2018 Nov 15] Available from: https://doi.org/10.1186/bcr2771

134. Aurilio G, Disalvatore D, Pruneri G, et al: A meta-analysis of oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 discordance between primary breast cancer and metastases [Internet]. European Journal of Cancer 50:277–289, 2014[cited 2018 Nov 15] Available from: http://www.sciencedirect.com/science/article/pii/S0959804913009040

135. Lindström LS, Karlsson E, Wilking UM, et al: 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 30:2601–2608, 2012

136. Largillier R, Ferrero J-M, Doyen J, et al: Prognostic factors in 1,038 women with metastatic breast cancer. Ann Oncol 19:2012–2019, 2008

137. Solomayer EF, Diel IJ, Meyberg GC, et al: Metastatic breast cancer: clinical course, prognosis and therapy related to the first site of metastasis. Breast Cancer Res Treat 59:271–278, 2000

138. Smalley RV, Scogna DM, Malmud LS: Advanced breast cancer with bone-only metastases: a chemotherapeutically responsive pattern of metastases. Am J Clin Oncol 5:161–166, 1982

139. Chen M-T, Sun H-F, Zhao Y, et al: Comparison of patterns and prognosis among distant metastatic breast cancer patients by age groups: a SEER population-based analysis. Sci Rep 7:9254, 2017

140. Yamamoto N, Watanabe T, Katsumata N, et al: Construction and validation of a practical prognostic index for patients with metastatic breast cancer. J Clin Oncol 16:2401–2408, 1998

141. Crivellari D, Aapro M, Leonard R, et al: Breast cancer in the elderly. J Clin Oncol 25:1882–1890, 2007

142. Zhu W, Perez EA, Hong R, et al: Age-Related Disparity in Immediate Prognosis of Patients with Triple-Negative Breast Cancer: A Population-Based Study from SEER Cancer Registries. PLoS ONE 10:e0128345, 2015

143. Sabiani L, Houvenaeghel G, Heinemann M, et al: Breast cancer in young women: Pathologic features and molecular phenotype. Breast 29:109–116, 2016

144. Andre F, Slimane K, Bachelot T, et al: Breast cancer with synchronous metastases: trends in survival during a 14-year period. J Clin Oncol 22:3302–3308, 2004

145. Dawood S, Haaland B, Albaracin C, et al: Is the Proportion of Patients Diagnosed with Synchronous Stage IV Breast Cancer Who Survive More than Two Years Increasing over

Time? [Internet]. OCL 89:79–87, 2015[cited 2018 Nov 16] Available from: http://www.karger.com/Article/FullText/371746

146. Gennari A, Conte P, Rosso R, et al: Survival of metastatic breast carcinoma patients over a 20-year period: a retrospective analysis based on individual patient data from six consecutive studies. Cancer 104:1742–1750, 2005

147. van den Hurk CJG, Eckel R, van de Poll-Franse LV, et al: Unfavourable pattern of metastases in M0 breast cancer patients during 1978-2008: a population-based analysis of the Munich Cancer Registry. Breast Cancer Res Treat 128:795–805, 2011

148. Foukakis T, Fornander T, Lekberg T, et al: Age-specific trends of survival in metastatic breast cancer: 26 years longitudinal data from a population-based cancer registry in Stockholm, Sweden. Breast Cancer Res Treat 130:553–560, 2011

149. Shen T, Gao C, Zhang K, et al: Prognostic outcomes in advanced breast cancer: the metastasis-free interval is important [Internet]. Human Pathology 70:70–76, 2017[cited 2018 Nov 16] Available from: http://www.sciencedirect.com/science/article/pii/S0046817717303611

150. Lobbezoo DJA, van Kampen RJW, Voogd AC, et al: Prognosis of metastatic breast cancer: are there differences between patients with de novo and recurrent metastatic breast cancer? Br J Cancer 112:1445–1451, 2015

151. Chang E, Mougalian SS, Adelson KB, et al: Association between prolonged metastatic free interval and recurrent metastatic breast cancer survival: findings from the SEER database. Breast Cancer Res Treat , 2018

152. Esteva FJ, Guo H, Zhang S, et al: PTEN, PIK3CA, p-AKT, and p-p70S6K status: association with trastuzumab response and survival in patients with HER2-positive metastatic breast cancer. Am J Pathol 177:1647–1656, 2010

153. Berns K, Horlings HM, Hennessy BT, et al: A Functional Genetic Approach Identifies the PI3K Pathway as a Major Determinant of Trastuzumab Resistance in Breast Cancer [Internet]. Cancer Cell 12:395–402, 2007[cited 2018 Nov 16] Available from: http://www.sciencedirect.com/science/article/pii/S1535610807002620

154. de Bono J, Ramanathan RK, Mina L, et al: Phase I, Dose-Escalation, Two-Part Trial of the PARP Inhibitor Talazoparib in Patients with Advanced Germline BRCA1/2 Mutations and Selected Sporadic Cancers. Cancer Discov 7:620–629, 2017

155. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. N Engl J Med 377:1700, 2017

156. Goodwin PJ, Phillips K-A, West DW, et al: Breast Cancer Prognosis in BRCA1 and BRCA2 Mutation Carriers: An International Prospective Breast Cancer Family Registry Population-Based Cohort Study [Internet]. JCO 30:19–26, 2012[cited 2018 Nov 16] Available from: http://ascopubs.org/doi/full/10.1200/JCO.2010.33.0068

157. Toy W, Shen Y, Won H, et al: ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. Nat Genet 45:1439–1445, 2013

158. Robinson DR, Wu Y-M, Vats P, et al: Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat Genet 45:1446–1451, 2013

159. Jeselsohn R, Buchwalter G, De Angelis C, et al: ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. Nat Rev Clin Oncol 12:573–583, 2015

160. Schiavon G, Hrebien S, Garcia-Murillas I, et al: Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. Sci Transl Med 7:313ra182, 2015

161. Fribbens C, O'Leary B, Kilburn L, et al: Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer. J Clin Oncol 34:2961–2968, 2016

162. Chandarlapaty S, Chen D, He W, et al: Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial. JAMA Oncol 2:1310–1315, 2016

163. Bose R, Kavuri SM, Searleman AC, et al: Activating HER2 mutations in HER2 gene amplification negative breast cancer [Internet]. Cancer Discov 3:224–237, 2013[cited 2018 Nov 16] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3570596/

164. Ben-Baruch NE, Bose R, Kavuri SM, et al: HER2-Mutated Breast Cancer Responds to Treatment With Single-Agent Neratinib, a Second-Generation HER2/EGFR Tyrosine Kinase Inhibitor. J Natl Compr Canc Netw 13:1061–1064, 2015

165. Amir E, Miller N, Geddie W, et al: Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. J Clin Oncol 30:587–592, 2012

166. Tobin NP, Harrell JC, Lövrot J, et al: Molecular subtype and tumor characteristics of breast cancer metastases as assessed by gene expression significantly influence patient post-relapse survival. Ann Oncol 26:81–88, 2015

167. Wolmark N, Wang J, Mamounas E, et al: Preoperative Chemotherapy in Patients With Operable Breast Cancer: Nine-Year Results From National Surgical Adjuvant Breast and Bowel Project B-18 [Internet]. J Natl Cancer Inst Monogr 2001:96–102, 2001[cited 2018 Dec 18] Available from: http://academic.oup.com/jncimono/article/2001/30/96/936263

168. Hage JA van der, Velde CJH van de, Julien J-P, et al: Preoperative Chemotherapy in Primary Operable Breast Cancer: Results From the European Organization for Research and Treatment of Cancer Trial 10902 [Internet]. Journal of Clinical Oncology , 2016[cited 2018 Dec 18] Available from: http://ascopubs.org/doi/10.1200/JCO.2001.19.22.4224

169. Golshan M, Cirrincione CT, Carey LA, et al: Impact of neoadjuvant therapy on breast conservation rates in triple-negative and HER2-positive breast cancer: Combined results of

CALGB 40603 and 40601 (Alliance). [Internet]. JCO 33:1007–1007, 2015[cited 2018 Nov 18] Available from: http://ascopubs.org/doi/abs/10.1200/jco.2015.33.15_suppl.1007

170. King TA, Morrow M: Surgical issues in patients with breast cancer receiving neoadjuvant chemotherapy. Nat Rev Clin Oncol 12:335–343, 2015

171. Kim J, Oktay K, Gracia C, et al: Which Patients Pursue Fertility Preservation Treatments? A Multi-Center Analysis of the Predictors of Fertility Preservation in Women with Breast Cancer [Internet]. Fertil Steril 97:671–676, 2012[cited 2018 Nov 18] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4191895/

172. Hage JH van der, Velde CC van de, Mieog SJ: Preoperative chemotherapy for women with operable breast cancer [Internet]. Cochrane Database of Systematic Reviews , 2007[cited 2018 Nov 18] Available from: http://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD005002.pub2/abstract

173. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Darby S, McGale P, et al: Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. Lancet 378:1707–1716, 2011

174. EBCTCG (Early Breast Cancer Trialists' Collaborative Group): Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials [Internet]. Lancet 383:2127–2135, 2014[cited 2018 Nov 18] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5015598/

175. Mansour EG, Gray R, Shatila AH, et al: Survival advantage of adjuvant chemotherapy in high-risk node-negative breast cancer: ten-year analysis--an intergroup study. [Internet]. JCO 16:3486–3492, 1998[cited 2018 Nov 18] Available from: http://ascopubs.org/doi/abs/10.1200/JCO.1998.16.11.3486

176. Early Breast Cancer Trialists' Collaborative Group (EBCTCG): Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials [Internet]. Lancet 379:432–444, 2012[cited 2018 Nov 18] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3273723/

177. Early Breast Cancer Trialists' Collaborative Group: Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women. N Engl J Med 319:1681–1692, 1988

178. Bonadonna G, Brusamolino E, Valagussa P, et al: Combination Chemotherapy as an Adjuvant Treatment in Operable Breast Cancer [Internet]. New England Journal of Medicine 294:405–410, 1976[cited 2018 Nov 18] Available from: https://doi.org/10.1056/NEJM197602192940801

179. Fisher B, Redmond C, Wickerham DL, et al: Doxorubicin-containing regimens for the treatment of stage II breast cancer: The National Surgical Adjuvant Breast and Bowel Project experience. J Clin Oncol 7:572–582, 1989

180. Fisher B, Brown AM, Dimitrov NV, et al: Two months of doxorubicin-cyclophosphamide with and without interval reinduction therapy compared with 6 months of cyclophosphamide, methotrexate, and fluorouracil in positive-node breast cancer patients with tamoxifen-nonresponsive tumors: results from the National Surgical Adjuvant Breast and Bowel Project B-15. J Clin Oncol 8:1483–1496, 1990

181. Fisher B, Anderson S, Tan-Chiu E, et al: Tamoxifen and chemotherapy for axillary node-negative, estrogen receptor-negative breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-23. J Clin Oncol 19:931–942, 2001

182. Namer M, Fargeot P, Roché H, et al: Improved disease-free survival with epirubicin-based chemoendocrine adjuvant therapy compared with tamoxifen alone in one to three node-positive, estrogen-receptor-positive, postmenopausal breast cancer patients: results of French Adjuvant Study Group 02 and 07 trials [Internet]. Ann Oncol 17:65–73, 2006[cited 2018 Nov 18] Available from: http://academic.oup.com/annonc/article/17/1/65/160873

183. Benefit of a High-Dose Epirubicin Regimen in Adjuvant Chemotherapy for Node-Positive Breast Cancer Patients With Poor Prognostic Factors: 5-Year Follow-Up Results of French Adjuvant Study Group 05 Randomized Trial [Internet]. JCO 19:602–611, 2001[cited 2018 Nov 18] Available from: http://ascopubs.org/doi/full/10.1200/JCO.2001.19.3.602

184. Martin M, Pienkowski T, Mackey J, et al: Adjuvant docetaxel for node-positive breast cancer. N Engl J Med 352:2302–2313, 2005

185. Martín M, Seguí MA, Antón A, et al: Adjuvant docetaxel for high-risk, node-negative breast cancer. N Engl J Med 363:2200–2210, 2010

186. Hackshaw A, Roughton M, Forsyth S, et al: Long-term benefits of 5 years of tamoxifen: 10-year follow-up of a large randomized trial in women at least 50 years of age with early breast cancer. J Clin Oncol 29:1657–1663, 2011

187. No Difference in Overall Survival with Shorter Extended AI Therapy [Internet]. OncLive [cited 2018 Nov 30] Available from: https://www.onclive.com/conference-coverage/sabcs-2017/no-difference-in-overall-survival-with-shorter-extended-ai-therapy

188. Perez EA, Suman VJ, Davidson NE, et al: Sequential versus concurrent trastuzumab in adjuvant chemotherapy for breast cancer. J Clin Oncol 29:4491–4497, 2011

189. Niraula S, Gyawali B: Optimal duration of adjuvant trastuzumab in treatment of early breast cancer: a meta-analysis of randomized controlled trials. Breast Cancer Res Treat , 2018

190. Gyawali B, Niraula S: Duration of adjuvant trastuzumab in HER2 positive breast cancer: Overall and disease free survival results from meta-analyses of randomized controlled trials [Internet]. Cancer Treatment Reviews 60:18–23, 2017[cited 2018 Nov 18] Available from: http://www.sciencedirect.com/science/article/pii/S030573721730124X

191. Earl HM, Hiller L, Vallier A-L, et al: PERSEPHONE: 6 versus 12 months (m) of adjuvant trastuzumab in patients (pts) with HER2 positive (+) early breast cancer (EBC): Randomised phase 3 non-inferiority trial with definitive 4-year (yr) disease-free survival (DFS) results. [Internet]. JCO 36:506–506, 2018[cited 2018 Nov 18] Available from: http://ascopubs.org/doi/abs/10.1200/JCO.2018.36.15_suppl.506

192. Amiri-Kordestani L, Wedam S, Zhang L, et al: First FDA approval of neoadjuvant therapy for breast cancer: pertuzumab for the treatment of patients with HER2-positive breast cancer. Clin Cancer Res 20:5359–5364, 2014

193. Jhaveri K, Esteva FJ: Pertuzumab in the treatment of HER2+ breast cancer. J Natl Compr Canc Netw 12:591–598, 2014

194. Swain SM, Baselga J, Kim S-B, et al: Pertuzumab, Trastuzumab, and Docetaxel in HER2-Positive Metastatic Breast Cancer [Internet]. New England Journal of Medicine 372:724–734, 2015[cited 2018 Nov 18] Available from: https://doi.org/10.1056/NEJMoa1413513

195. Schena M, Shalon D, Davis RW, et al: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470, 1995

196. Lima Passos V, Tan FES, Winkens B, et al: Optimal designs for one- and two-color microarrays using mixed models: a comparative evaluation of their efficiencies. J Comput Biol 16:67–83, 2009

197. Bloom JS, Khan Z, Kruglyak L, et al: Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays [Internet]. BMC Genomics 10:221, 2009[cited 2018 Nov 30] Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-10-221

198. Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, et al: The affymetrix GeneChip platform: an overview. Meth Enzymol 410:3–28, 2006

199. Cope LM, Irizarry RA, Jaffee HA, et al: A benchmark for Affymetrix GeneChip expression measures. Bioinformatics 20:323–331, 2004

200. Bolstad BM, Irizarry RA, Astrand M, et al: A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193, 2003

201. Shakya K, Ruskin HJ, Kerr G, et al: Comparison of microarray preprocessing methods. Adv Exp Med Biol 680:139–147, 2010

202. Microarray Analysis - SE [Internet]. Advance your research with Affymetrix microarray analysis products [cited 2018 Nov 30] Available from: https://www.thermofisher.com/uk/en/home/life-science/microarray-analysis.html

203. mAdb Training/Affymetrix, Statistical Algorithms Description Document [Internet][cited 2018 Nov 30] Available from: https://madb.nci.nih.gov/mAdb-public/Training/index.shtml

204. Irizarry RA, Hobbs B, Collin F, et al: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264, 2003

205. McGee M, Chen Z: Parameter estimation for the exponential-normal convolution model for background correction of affymetrix GeneChip data. Stat Appl Genet Mol Biol 5:Article24, 2006

206. Seo J, Hoffman EP: Probe set algorithms: is there a rational best bet? BMC Bioinformatics 7:395, 2006

207. Harr B, Schlötterer C: Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons. Nucleic Acids Res 34:e8, 2006

208. Yang YH, Buckley MJ, Speed TP: Analysis of cDNA microarray images. Brief Bioinformatics 2:341–349, 2001

209. Tran PH, Peiffer DA, Shin Y, et al: Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals. Nucleic Acids Res 30:e54, 2002

210. Scharpf RB, Iacobuzio-Donahue CA, Sneddon JB, et al: When should one subtract background fluorescence in 2-color microarrays? Biostatistics 8:695–707, 2007

211. Yang YH, Dudoit S, Luu P, et al: Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30:e15, 2002

212. Kouadjo KE, Nishida Y, Cadrin-Girard JF, et al: Housekeeping and tissue-specific genes in mouse tissues. BMC Genomics 8:127, 2007

213. Gautier L, Møller M, Friis-Hansen L, et al: Alternative mapping of probes to genes for Affymetrix chips. BMC Bioinformatics 5:111, 2004

214. Dai M, Wang P, Boyd AD, et al: Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33:e175, 2005

215. Loganantharaj R, Cheepala S, Clifford J: Metric for measuring the effectiveness of clustering of DNA microarray expression. BMC Bioinformatics 7 Suppl 2:S5, 2006

216. D'haeseleer P: How does gene expression clustering work? Nat Biotechnol 23:1499–1501, 2005

217. Raychaudhuri S, Stuart JM, Altman RB: Principal components analysis to summarize microarray experiments: application to sporulation time series. Pac Symp Biocomput 455–466, 2000

218. Ringnér M: What is principal component analysis? Nat Biotechnol 26:303–304, 2008

219. Zhu Y, Shen X, Pan W: Network-based support vector machine for classification of microarray samples. BMC Bioinformatics 10 Suppl 1:S21, 2009

220. Liao JG, Chin K-V: Logistic regression for disease classification using microarray data: model selection in a large p and small n case. Bioinformatics 23:1945–1951, 2007

221. O'Neill MC, Song L: Neural network analysis of lymphoma microarray data: prognosis and diagnosis near-perfect. BMC Bioinformatics 4:13, 2003

222. Alizadeh AA, Eisen MB, Davis RE, et al: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511, 2000

223. Cardoso F, van't Veer LJ, Bogaerts J, et al: 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. N Engl J Med 375:717–729, 2016

224. Sotiriou C, Wirapati P, Loi S, et al: Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 98:262–272, 2006

225. Lundberg A, Lindström LS, Harrell JC, et al: Gene Expression Signatures and Immunohistochemical Subtypes Add Prognostic Value to Each Other in Breast Cancer Cohorts. Clin Cancer Res 23:7512–7520, 2017

226. Chibon F: Cancer gene expression signatures - the rise and fall? Eur J Cancer 49:2000–2009, 2013

227. van 't Veer LJ, Dai H, van de Vijver MJ, et al: Gene expression profiling predicts clinical outcome of breast cancer [Internet]. Nature 415:530–536, 2002[cited 2015 Aug 12] Available from: http://www.nature.com/nature/journal/v415/n6871/full/415530a.html

228. Mook S, Schmidt MK, Viale G, et al: The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. Breast Cancer Res Treat 116:295–302, 2009

229. Saghatchian M, Mook S, Pruneri G, et al: Additional prognostic value of the 70-gene signature (MammaPrint(®)) among breast cancer patients with 4-9 positive lymph nodes. Breast 22:682–690, 2013

230. Drukker CA, van Tinteren H, Schmidt MK, et al: Long-term impact of the 70-gene signature on breast cancer outcome. Breast Cancer Res Treat 143:587–592, 2014

231. Drukker CA, Bueno-de-Mesquita JM, Retèl VP, et al: A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. Int J Cancer 133:929–936, 2013

232. Elston CW, Ellis IO: Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 41:154–161, 2002

233. Loi S, Haibe-Kains B, Desmedt C, et al: Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 25:1239–1246, 2007

234. Liedtke C, Hatzis C, Symmans WF, et al: Genomic grade index is associated with response to chemotherapy in patients with breast cancer. J Clin Oncol 27:3185–3191, 2009

235. Toussaint J, Sieuwerts AM, Haibe-Kains B, et al: Improvement of the clinical applicability of the Genomic Grade Index through a qRT-PCR test performed on frozen and formalin-fixed paraffin-embedded tissues. BMC Genomics 10:424, 2009

236. Early Breast Cancer Trialists' Collaborative Group (EBCTCG): Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365:1687–1717, 2005

237. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Davies C, Godwin J, et al: Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet 378:771–784, 2011

238. Sparano JA, Paik S: Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol 26:721–728, 2008

239. Paik S, Shak S, Tang G, et al: A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer [Internet]. New England Journal of Medicine 351:2817–2826, 2004[cited 2017 Jan 31] Available from: http://dx.doi.org/10.1056/NEJMoa041588

240. Albain KS, Barlow WE, Shak S, et al: Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial [Internet]. The Lancet Oncology 11:55–65, 2010[cited 2017 Jan 25] Available from: /journals/lanonc/article/PIIS1470-2045(09)70314-6/abstract

241. Sparano JA, Gray RJ, Makower DF, et al: Prospective Validation of a 21-Gene Expression Assay in Breast Cancer [Internet]. New England Journal of Medicine 373:2005–2014, 2015[cited 2017 Jan 24] Available from: http://dx.doi.org/10.1056/NEJMoa1510764

242. Sparano JA, Gray RJ, Makower DF, et al: Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med 379:111–121, 2018

243. Nielsen TO, Parker JS, Leung S, et al: A Comparison of PAM50 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor–Positive Breast Cancer [Internet]. Clin Cancer Res 16:5222–5232, 2010[cited 2017 Jan 31] Available from: http://clincancerres.aacrjournals.org/content/16/21/5222

244. Dowsett M, Sestak I, Lopez-Knowles E, et al: Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. J Clin Oncol 31:2783–2790, 2013

245. Filipits M, Nielsen TO, Rudas M, et al: The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. Clin Cancer Res 20:1298–1305, 2014

246. Sestak I, Cuzick J, Dowsett M, et al: Prediction of Late Distant Recurrence After 5 Years of Endocrine Treatment: A Combined Analysis of Patients From the Austrian Breast and Colorectal Cancer Study Group 8 and Arimidex, Tamoxifen Alone or in Combination Randomized Trials Using the PAM50 Risk of Recurrence Score [Internet]. JCO 33:916–922, 2015[cited 2017 Jan 31] Available from: http://ascopubs.org/doi/abs/10.1200/JCO.2014.55.6894

247. Wallden B, Storhoff J, Nielsen T, et al: Development and verification of the PAM50-based Prosigna breast cancer gene signature assay [Internet]. BMC Medical Genomics 8, 2015[cited 2018 Sep 25] Available from: http://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-015-0129-6

248. Gnant M, Sestak I, Filipits M, et al: Identifying clinically relevant prognostic subgroups of postmenopausal women with node-positive hormone receptor-positive early-stage breast cancer treated with endocrine therapy: a combined analysis of ABCSG-8 and ATAC using the PAM50 risk of recurrence score and intrinsic subtype. Ann Oncol 26:1685–1691, 2015

249. Harris LN, Ismaila N, McShane LM, et al: Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 34:1134–1150, 2016

250. Prosigna® Breast Cancer Prognostic Gene Signature Assay [package insert]. Seattle, WA: NanoString Technologies, Inc; 2015

251. Dancik GM, Theodorescu D: Robust prognostic gene expression signatures in bladder cancer and lung adenocarcinoma depend on cell cycle related genes. PLoS ONE 9:e85249, 2014

252. Mosley JD, Keri RA: Cell cycle correlated genes dictate the prognostic power of breast cancer gene lists. BMC Med Genomics 1:11, 2008

253. Sotiriou C, Pusztai L: Gene-expression signatures in breast cancer. N Engl J Med 360:790–800, 2009

254. Gray KA, Yates B, Seal RL, et al: Genenames.org: the HGNC resources in 2015. Nucleic Acids Res 43:D1079-1085, 2015

255. Kanehisa M, Sato Y, Kawashima M, et al: KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457-462, 2016

256. Santos A, Wernersson R, Jensen LJ: Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes [Internet]. Nucleic Acids Research 43:D1140–D1144, 2015[cited 2016 Nov 21] Available from: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1092

257. Ma X-J, Salunga R, Dahiya S, et al: A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res 14:2601–2608, 2008

258. Zhang Y, Schnabel CA, Schroeder BE, et al: Breast cancer index identifies early-stage estrogen receptor-positive breast cancer patients at risk for early- and late-distant recurrence. Clin Cancer Res 19:4196–4205, 2013

259. Sgroi DC, Sestak I, Cuzick J, et al: Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. Lancet Oncol 14:1067–1076, 2013

260. Sanft T, Aktas B, Schroeder B, et al: Prospective assessment of the decision-making impact of the Breast Cancer Index in recommending extended adjuvant endocrine therapy for patients with early-stage ER-positive breast cancer. Breast Cancer Res Treat 154:533–541, 2015

261. Filipits M, Rudas M, Jakesz R, et al: A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res 17:6012–6020, 2011

262. Dubsky P, Brase JC, Jakesz R, et al: The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. Br J Cancer 109:2959–2964, 2013

263. Martin M, Brase JC, Ruiz A, et al: Prognostic ability of EndoPredict compared to research-based versions of the PAM50 risk of recurrence (ROR) scores in node-positive, estrogen receptor-positive, and HER2-negative breast cancer. A GEICAM/9906 sub-study. Breast Cancer Res Treat 156:81–89, 2016

264. Buus R, Sestak I, Kronenwett R, et al: Comparison of EndoPredict and EPclin With Oncotype DX Recurrence Score for Prediction of Risk of Distant Recurrence After Endocrine Therapy. J Natl Cancer Inst 108, 2016

265. Tobin NP, Lindström LS, Carlson JW, et al: Multi-level gene expression signatures, but not binary, outperform Ki67 for the long term prognostication of breast cancer patients. Mol Oncol 8:741–752, 2014

266. Reyal F, Bollet MA, Caly M, et al: Respective prognostic value of genomic grade and histological proliferation markers in early stage (pN0) breast carcinoma. PLoS ONE 7:e35184, 2012

267. Niikura N, Iwamoto T, Masuda S, et al: Immunohistochemical Ki67 labeling index has similar proliferation predictive power to various gene signatures in breast cancer. Cancer Sci 103:1508–1512, 2012

268. Cheang MCU, Chia SK, Voduc D, et al: Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst 101:736–750, 2009

269. Varga Z, Sinn P, Fritzsche F, et al: Comparison of EndoPredict and Oncotype DX test results in hormone receptor positive invasive breast cancer. PLoS ONE 8:e58483, 2013

270. Cancer Genome Atlas Research Network: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455:1061–1068, 2008

271. Cancer Genome Atlas Research Network: Integrated genomic analyses of ovarian carcinoma. Nature 474:609–615, 2011

272. Cancer Genome Atlas Network: Comprehensive molecular portraits of human breast tumours. Nature 490:61–70, 2012

273. Grieb BC, Chen X, Eischen CM: MTBP is overexpressed in triple-negative breast cancer and contributes to its growth and survival. Mol Cancer Res 12:1216–1224, 2014

274. Cancer Genome Atlas Research Network: Comprehensive genomic characterization of squamous cell lung cancers. Nature 489:519–525, 2012

275. Ceccarelli M, Barthel FP, Malta TM, et al: Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. Cell 164:550–563, 2016

276. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al: The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45:1113–1120, 2013

277. Hoadley KA, Yau C, Hinoue T, et al: Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer [Internet]. Cell 173:291-304.e6, 2018[cited 2018 Oct 15] Available from: http://www.sciencedirect.com/science/article/pii/S0092867418303027

278. Ding L, Bailey MH, Porta-Pardo E, et al: Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. Cell 173:305-320.e10, 2018

279. Sanchez-Vega F, Mina M, Armenia J, et al: Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell 173:321-337.e10, 2018

280. Barrett T, Wilhite SE, Ledoux P, et al: NCBI GEO: archive for functional genomics data sets—update [Internet]. Nucleic Acids Res 41:D991–D995, 2013[cited 2018 Dec 7] Available from: http://academic.oup.com/nar/article/41/D1/D991/1067995

281. Lappalainen I, Almeida-King J, Kumanduri V, et al: The European Genome-phenome Archive of human data consented for biomedical research [Internet]. Nature Genetics 47:692–695, 2015[cited 2018 Dec 3] Available from: http://www.nature.com/articles/ng.3312

282. PanCanAtlas Publications | NCI Genomic Data Commons [Internet][cited 2018 Oct 15] Available from: https://gdc.cancer.gov/about-data/publications/pancanatlas

283. Hatschek T, Carlsson L, Einbeigi Z, et al: Individually tailored treatment with epirubicin and paclitaxel with or without capecitabine as first-line chemotherapy in metastatic breast cancer: a randomized multicenter trial. Breast Cancer Res Treat 131:939–947, 2012

284. Curtis C, Shah SP, Chin S-F, et al: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346–352, 2012

285. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

286. Bengtsson H, Irizarry R, Carvalho B, et al: Estimation and assessment of raw copy numbers at the single locus level. Bioinformatics 24:759–767, 2008

287. Li B, Dewey CN: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome [Internet]. BMC Bioinformatics 12:323, 2011[cited 2018 Dec 3] Available from: https://doi.org/10.1186/1471-2105-12-323

288. Liu J, Lichtenberg T, Hoadley KA, et al: An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics [Internet]. Cell 173:400-416.e11, 2018[cited 2018 Nov 6] Available from: http://www.sciencedirect.com/science/article/pii/S0092867418302290

289. GenomeStudio Software [Internet][cited 2018 Dec 8] Available from: https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html

290. Olshen AB, Venkatraman ES, Lucito R, et al: Circular binary segmentation for the analysis of array-based DNA copy number data [Internet]. Biostatistics 5:557–572,

2004[cited 2018 Dec 3] Available from: http://academic.oup.com/biostatistics/article/5/4/557/275197

291. Venkatraman E, Olhsen S, Olshen A: DNAcopy: DNA copy number data analysis. R package (version 1.54.0)

292. Beroukhim R, Mermel CH, Porter D, et al: The landscape of somatic copy-number alteration across human cancers. Nature 463:899–905, 2010

293. Mayakonda A, Lin D-C, Assenov Y, et al: Maftools: efficient and comprehensive analysis of somatic variants in cancer [Internet]. Genome Res 28:1747–1756, 2018[cited 2018 Dec 3] Available from: http://genome.cshlp.org/content/28/11/1747

294. Taylor AM, Shih J, Ha G, et al: Genomic and Functional Approaches to Understanding Cancer Aneuploidy [Internet]. Cancer Cell 33:676-689.e3, 2018[cited 2018 Nov 4] Available from: http://www.sciencedirect.com/science/article/pii/S1535610818301119

295. Barnes DM, Gillett CE: Cyclin D1 in breast cancer. Breast Cancer Res Treat 52:1–15, 1998

296. Guo L-L, Gao P, Wu Y-G, et al: Alteration of Cyclin D1 in Chinese Patients with Breast Carcinoma and its Correlation with Ki-67, pRb, and p53 [Internet]. Archives of Medical Research 38:846–852, 2007[cited 2018 Dec 4] Available from: http://www.sciencedirect.com/science/article/pii/S0188440907002044

297. Rudas M, Lehnert M, Huynh A, et al: Cyclin D1 expression in breast cancer patients receiving adjuvant tamoxifen-based therapy. Clin Cancer Res 14:1767–1774, 2008

298. Elsheikh S, Green AR, Aleskandarany MA, et al: CCND1 amplification and cyclin D1 expression in breast cancer and their relation with proteomic subgroups and patient outcome. Breast Cancer Res Treat 109:325–335, 2008

299. Prat A, Cheang MCU, Martín M, et al: Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. J Clin Oncol 31:203–209, 2013

300. Finn RS, Crown JP, Lang I, et al: The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. Lancet Oncol 16:25–35, 2015

301. Bailey MH, Tokheim C, Porta-Pardo E, et al: Comprehensive Characterization of Cancer Driver Genes and Mutations [Internet]. Cell 173:371-385.e18, 2018[cited 2018 Nov 4] Available from: https://www.cell.com/cell/abstract/S0092-8674(18)30237-X

302. Chang F, Lee JT, Navolanic PM, et al: Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. Leukemia 17:590–603, 2003