Aus der Klinik und Poliklinik für Frauenheilkunde und Geburthilfe,

Klinikum der Ludwig-Maximilians-Universität München

Vorstand: Prof. Dr. Sven Mahner

Evaluation of molecular prognostic markers for premenopausal patients with HR+/HER2− early breast cancer

Dissertation

zum Erwerb des Doktorgrades der Medizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Hua Ni

aus

Lhasa, Tibet, China

Jahr

2022

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. Nadia Harbeck
Mitberichterstatter:	Prof. Dr. Eva-Maria Grischke Priv. Doz. Dr. Stefanie Corradini
Mitbetreuung durch den promovierten Mitarbeiter:	PD Dr. Rachel Würstlein
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	27.10.2022

Table of content

Table of content **3**

Zusammenfassung (Deutsch): **5**

Abstract (English): **5**

List of figures **7**

List of tables **7**

List of abbreviations **8**

1. Introduction **9**

1.1 A brief review of the development of diagnosis and treatment of breast cancer 9

1.2 Refining the treatment for patients with HR+/HER2– EBC 10

1.2.1 De-escalation of systemic treatment 10

1.2.2 Multigene assays as risk predictors 12

1.2.3 Escalation of systemic treatment 16

1.3 The necessity of establishing molecular tests for premenopausal patients 17

2. Material and Methods **20**

2.1 Patient 20

2.1.1 Study design 20

2.1.2 Ethical approval 20

2.1.3 Sample collection 20

2.2 Reagents 23

2.2.1 RNA extraction 23

2.2.2 Nanostring analysis 24

2.3 Equipment 24

2.3.1 RNA extraction 24

2.3.2 Nanostring 24

2.4 Tools/Software 25

2.5 Methods 25

2.5.1 RNA extraction 25

2.5.2 Nanostring profiling 26

2.5.3 Data analysis 27

3. Results **30**

3.1 Patient characteristics 30

3.1.1 Follow-up information 30

3.1.2 Patient information 30

3.1.3 Survival relevance of clinical parameters 32

3.2 PAM50 analysis 32

3.2.1 Subtype distribution 32

3.2.2 PAM50 subtype scores and risk of distant metastasis 32

3.3 BC360® signature analysis 34

3.3.1 Signature score overview for each subtype 34

3.3.2 BC360® signature expression 36
3.3.3 The four significant signatures were also associated with survival 37
3.3.4 Subgroup analysis of BC360® signatures 38
3.4 GSEA analysis 40
3.5 Analysis of differentially expressed genes 41
3.5.1 Differential expression of single genes 41
3.5.2 Survival relevancy of the DEGs 43
3.5.3 Functional analysis of the 19 survival-related DEGs 45
3.5.4 Differentially expressed genes in luminal A and luminal B subgroups 48
3.6 Multivariate survival analysis 49

4. Discussion 52

4.1 Clinical risk factors for developing distant metastasis in HR+/HER2− early breast cancer 52
4.2 Molecular subtype and risk of developing distant metastasis 54
4.3 BC360® signatures and the risk of developing distant metastasis 55
4.4 GSEA pathways and the risk of developing distant metastasis 56
4.5 Single gene markers and the risk of distant metastasis 57
4.5.1 The prognostic value of LRP2 in breast cancer 58
4.5.2 The prognostic value of SCUBE2 in breast cancer 59
4.5.3 Other single prognostic gene markers in breast cancer 59
4.6 Summary of the prognostic factors 62

References 63

Apendix A: Key codes used in R analysis 75

Acknowledgements 77

Affidavit 79

List of publications 80

Your contribution to the publication 81

Zusammenfassung (Deutsch):

Prämenopausale Patientinnen, bei denen ein Hormonrezeptor-positiver (HR+)/humaner epidermaler Wachstumsfaktorrezeptor-2-negativer (HER2-) Brustkrebs im Frühstadium (EBC) diagnostiziert wurde, haben tendenziell eine schlechtere Prognose als postmenopausale Patientinnen. Luminale Mammakarzinome bei prämenopausalen Patientinnen können einzigartige molekulare Muster aufweisen, und Experten haben präzisere polygene Werkzeuge gefordert, um individualisierte Behandlungen für diese Patientinnen zu entwickeln. Unser Projekt konzentriert sich auf prämenopausale HR+/HER2- EBC-Patientinnen und zielt darauf ab, molekulare Prognosefaktoren zu untersuchen und die altersspezifische Entwicklung und Interpretation von Genexpressionstests zu unterstützen.

In einer Fall-Kontrollstudie haben wir geeignete Patientinnen des LMU Brustzentrums mit mindestens 10 Jahren Nachbeobachtungszeit eingeschlossen, und nach dem Auftreten von Fernmetastasen in der Nachbeobachtungszeit gruppiert. Es konnten 97 Patienten in unsere Patientenkohorte aufgenommen werden. Mittels Nanostring nCounter®-Technologie wurden die Genexpressionsprofile der Primärtumorproben analysiert. Im Anschluss wurde eine bioinformatische Analyse durchgeführt, um nach signifikant dysregulierten Signaturen/Genen zu suchen und den prognostischen Wert von der gefundenen Markern in unserer Kohorte sowie in Online-Datenbanken untersucht.

Fünf und achtzig der 97 Tumorproben bestanden die RNA-Qualitätskontrolle und die PAM50-Subtypanalyse zeigte, dass fast alle (81/85) Tumoren zum luminalen Subtyp gehörten. Interessanterweise zeigten die metastasierten Tumoren eine engere Korrelation mit der HER2-Enriched-Biologie. In Bezug auf die Signalübertragung deutet unsere Studie darauf hin, dass ROR Score (Rückfallrisiko), PGR (Progesteronrezeptor) und mTORC1 (Säugetierziel des Rapamycinkomplexes 1) Signalübertragung bei prämenopausalen Patientinnen das Potenzial haben, als prognostische Faktoren zu dienen. Neben der Ermittlung etablierter krebsbezogener Signaturen haben wir auch 22 Einzelgene gescreent, die bei Patienten mit Fernmetastasen signifikant unterschiedlich exprimiert wurden und von denen 19 signifikant mit dem Fernmetastasen-freien Überleben (DMFS) assoziiert waren. Besonders hervorzuheben ist, dass der prognostische Wert von 15 der 19 Gene in externen klinischen Datensätzen validiert werden konnte. Gemäß der multivariaten Analyse, die alle DMFS-bezogenen klinischen Faktoren/Signaturen/DEGs (differenziell exprimierte Gene) umfasste, sind LRP2 (Low Density Lipoprotein Receptor-Related Protein 2), PTGER3 (Prostaglandin-E-Rezeptor 3) zusammen mit dem Lymphknotenstatus unabhängige Prognosefaktoren.

Unser Projekt untersuchte prognostische molekulare Faktoren bei prämenopausalen HR+/HER2–EBC-Patientinnen, die im Vergleich zu postmenopausalen Patientinnen z.T. ähnlich und z.T. einzigartig waren. Diese Ergebnisse können bei der personalisierten Behandlung von prämenopausalen Patientinnen weiterhelfen.

Abstract (English):

Premenopausal patients diagnosed with hormone receptor-positive (HR+) / human epidermal growth factor receptor 2-negative (HER2–) early breast cancer (EBC) often have a poorer prognosis compared to their postmenopausal counterparts. Luminal cancer in premenopausal patients can display unique molecular patterns and experts call for more precise multigene tools to support individualized treatment concepts for these patients. Our project focused on premenopausal

patients with HR+/ HER2– EBC and aimed to investigate the molecular prognostic factors and facilitate the age-tailored development and interpretation of gene expression tests.

In a case-control study, eligible patients treated in the LMU breast center with at least a ten-year follow-up were grouped according to whether they had developed distant metastases or not. Ninety-seven patients could be included in our patient cohort. Nanostring nCounter® technology was used to decipher the gene expression profile of the tumor samples acquired in their primary tumor specimens. Afterwards, bioinformatics analysis was carried out to screen out the significantly dysregulated signatures/genes and examine the prognostic value of the markers in our cohort and online databases.

Eighty-five of the 97 tumor samples passed the RNA quality control, and PAM50 subtyping analysis suggested that nearly all (81/85) of the tumors belonged to the luminal subtypes. Interestingly, tumors that had developed metastases showed closer correlation to HER2-Enriched biology. As for signatures, our study indicated the potential of applying ROR (risk of recurrence), PGR (progesterone receptor), and the enrichment of mTORC1 (mammalian target of rapamycin complex 1) signaling as prognostic factors in premenopausal patients. Except for excavating the established cancer-related signatures, we selected out 22 single genes that were significantly differentially expressed in patients who developed distant metastasis, and 19 of the 22 genes were significantly associated with distant metastasis-free survival (DMFS). When we used online databases to confirm the prognostic power of the 19 genes, 15 of them are prognostic. According to the multivariate analysis that included all DMFS-related clinical factors/signatures/DEGs (differentially expressed genes), LRP2 (low density lipoprotein receptor-related protein 2), PTGER3 (prostaglandin E receptor 3) along with node status are the most independent prognostic factors.

Our project studied prognostic molecular factors for premenopausal patients with HR+/ HER2– EBC that possess both similarities and uniqueness compared to their relevance in postmenopausal patients. These findings could further contribute to the personalized management of premenopausal patients.

List of figures

Figure 1-1 Algorithm for treating luminal early breast cancer
Figure 1-2 The mechanism of estrogen receptor (ER) in breast cancer
Figure 2-1 Study design
Figure 2-2 Work flow for sample collection
Figure 2-3 Overview of the RNA extraction procedure
Figure 2-4 Methodology of Nanostring nCounter® technology
Figure 3-1 Survival swimmer plot (each line represents one patient)
Figure 3-2 Differential analysis of the subtype scores
Figure 3-3 The overview of signature analysis for each subtype
Figure 3-4 The differentially expressed BC360® signatures
Figure 3-5 Survival analysis of four significant signatures
Figure 3-6 Subgroup analysis of BC360® signatures
Figure 3-7 GSEA analysis of hallmark gene sets in all patients/subgroups
Figure 3-8 The differentially expressed genes between two groups of patients
Figure 3-9 Survival analysis of the differentially expressed genes
Figure 3-10 Functional analysis of nineteen survival-related DEGs
Figure 3-11 The differentially expressed genes in subgroups

List of tables

Table 1-1 Clinical validation of gene expression tests in premenopausal patients with HR+/HER2−, early breast cancer (EBC)
Table 2-1 Tumor content in FFPE samples
Table 2-2 Tested GSEA sets
Table 3-1 Patient clinical characteristics
Table 3-2 Univariate survival analysis of clinical parameters
Table 3-3 Subtypes' distribution
Table 3-4 Overview of significant gene set enrichments
Table 3-5 The involved BC360® pathways of the19 survival-related DEGs
Table 3-6 The distribution of 19 DEGs in BC360® pathways
Table 3-7 Prognostic factors included in the multivariate analysis (the starting block)
Table 3-8 Multivariate survival analysis of 28 prognostic factors (method: forward stepwise)
Table 3-9 Correlation analysis of LRP2 and other prognostic factors
Table 3-10 Multivariate survival analysis of pN, pT, and differentially expressed signatures/genes (method: forward stepwise)

List of abbreviations

DEGs	Differentially expressed genes
DMFS	Distant metastasis-free survival
EBC	Early breast cancer
ER	Estrogen receptor
FFPE	Formalin-fixed paraffin-embedded
FISH	Fluorescence in situ hybridization
GSEA	Gene set enrichment analysis
GO	Gene ontology
HER2	Human epidermal growth factor receptor
HR	Hormone receptor
IHC	Immunohistochemistry
KEGG	Kyoto encyclopedia of genes and genomes
LRP2	Low density lipoprotein receptor-related protein 2
M0	No metastasis group
M1	Metastasis group
mTORC1	Mammalian target of rapamycin complex 1
PAM50	Prediction analysis of microarray
PGR	Progesterone receptor
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic
PR	Progesterone receptor
PTGER3	Prostaglandin E receptor 3
ROR	Risk of recurrence
STRING	Search tool for the retrieval of interacting genes/proteins

1. Introduction

Breast cancer is the most common cancer and the leading cause of cancer-related death in women [1]. Though diagnosis and treatment of breast cancer have evolved to a comprehensive and evidence-guided routine, individualized or personalized treatment is not yet a reality and a considerable proportion of patients are experiencing treatment failure, or over-/under-treatment. Exploring personalized prognostic factors/treatment targets is crucial for optimizing breast cancer management.

1.1 A brief review of the development of diagnosis and treatment of breast cancer

Cancer has been on this planet long before our civilization. The earliest evidence of cancer in mammals was discovered in the fossils of dinosaurs and the earliest cancer in humans was found in a fossilized foot from pre-historic times (around 1.7 million years ago). The earliest written record of cancer dates to around 1500BC in Egypt. According to the document, the tumors in chest (breast cancer) were treated by wound suturing and cauterization with a tool called the “fire drill”, and were considered nonetheless incurable. The term “cancer” originates from around 400 BC, when the Greek physician, Hippocrates, who was considered as “the father of medicine”, used the term “carcinoma” (which means crab in English) to refer to an abnormal body swelling. Later, the term was translated into its Latin form: “cancer”. According to Hippocrates’s theory, the diagnosis of cancer was based on visual observation and the treatment was a combination of medicine, surgery, and cauterization [2].

Though the notice of breast cancer was early, it was until the 18th century that the research and treatment of breast cancer embraced the breakthrough which was fueled by the rapid development of modern science and shaped the modern management of breast cancer. The invention of microscopes brought people’s sight to the cell level. The discovery of anesthesia, asepsis, and antibiotics, and the improved medical support propelled the upgrade of surgery, which was later refined into current mastectomy and lumpectomy. The discovery of X-rays and radioactive elements (namely uranium, radium, and polonium) gave rise to modern diagnostic and therapeutic radiology and nuclear medicine [2]. The anti-recurrence effect of oophorectomy, adrenalectomy and hypophysectomy lay the foundation for the wide application of endocrine treatment [3]. The anticancer potential of mustard gas that discovered during world war marked the beginning of the chemotherapy era [2]. By the end of 20 century, the basic frame of modern management of breast cancer was formed as a comprehensive diagnostic routine based on physical checks, pathology, and imaging together with a comprehensive treatment procedure composed of operation, radiotherapy, and systemic treatment [4].

Incubated during the 19th century, and rocketed since the middle 20th century, molecular research started a revolution in the field of biology and gave people the possibility to decipher the nature of life in the molecular world where direct visual observation is impossible [5, 6]. Genetic research in the breast cancer led to discovery of several oncogenes and tumor suppressor genes, thereby ending the era of blaming lifestyle as the main cause of cancer. A classic example is that BRCA1 (BRCA1 DNA repair associated) and/or BRCA2 (BRCA2 DNA repair associated) mutations were established as risk factors for breast cancer development and are now frequently tested in patients with metastatic breast cancer [6, 7]. Besides, the association between overexpression of human epidermal receptor 2 (HER2) gene and breast cancer development is also a remarkable

achievement of molecular research [6]. This finding led to identification of a distinctive breast cancer subgroup and assisted in generating the first monoclonal antibody against breast cancer [4].

High throughput testing methods again boosted the knowledge of breast cancer. Publishing the human genome sequence in the beginning of 21st century marked the dawn of precision medicine [8]. By analyzing the gene expression patterns in 65 tumor samples from 42 individuals with locally advanced breast cancer with a microarray representing 8102 human genes, Perou et al. discovered that breast cancer is not homogenous, but composed of four distinct molecular pattens: luminal-like, basal-like, Erb-B2+, normal-like [9]. Discovery of the heterogeneity of breast cancer reformed the clinical routine of breast cancer treatment: the subtype of breast cancer should be evaluated before the starting systemic treatment [4, 10]. Due to the high cost of multigene tests, surrogate subtyping of breast cancer which relies on immunohistological tests of key molecules is in wider clinical use and helps to stratify patients into five groups for treatment consideration: triple-negative, HER2-enriched (HER2-E), luminal B-like (HER2+), luminal B-like (HER2–), luminal A-like [10].

Diagnosis and treatment of breast cancer has a long history, yet today it has been experiencing the fastest development in recent centuries. Currently, breast cancer diagnosis relies on a combination of traditional physical examination, imaging, pathology, and gene tests; breast cancer treatment, meanwhile, is composed of local treatment including surgery and radiotherapy and systemic treatment, encompassing chemotherapy, endocrine therapy, and targeted therapy [10, 11]. Both, diagnosis and treatment of breast cancer await further improvement regarding efficiency, convenience, and affordability and aim to offer the best possible quality of life for each patient.

1.2 Refining the treatment for patients with HR+/HER2– EBC

Thanks to the evolvement of modern medicine, early breast cancer is now a curable disease. But, regarding the treatment of each subtype of breast cancer, quite a few conundrums remain unsolved. HR+/HER2– EBC, the subtype studied in this dissertation, is the most common breast cancer subtype (about 70%) [10, 12]. It has a relatively favorable prognosis compared to HER2+ or triple-negative subtype; nevertheless, given its frequency, more patients die from this subtype [12].

The mainstay of systemic therapy for HR+/ HER2– EBC is endocrine therapy and chemotherapy. But the current treatment is far from perfect: some patients respond very well to endocrine therapy and could safely omit chemotherapy [13, 14]. Around half of the patients are resistant to endocrine therapy [12] and nearly one third of the patients will experience distant metastases [15]. Therefore, more accurate stratification is necessary to select patients who do not need chemotherapy and patients who need more than endocrine therapy or chemotherapy: de-escalation and escalation [16].

1.2.1 De-escalation of systemic treatment

De-escalation is a cautious strategy that considers both treatment efficiency and quality of life [16]. Along with the development of modern medicine, de-escalation is gaining growing attention, and one question is now frequently asked before an aggressive regimen is prescribed: is this treatment absolutely necessary? The de-escalation concept has reformed and is still reforming

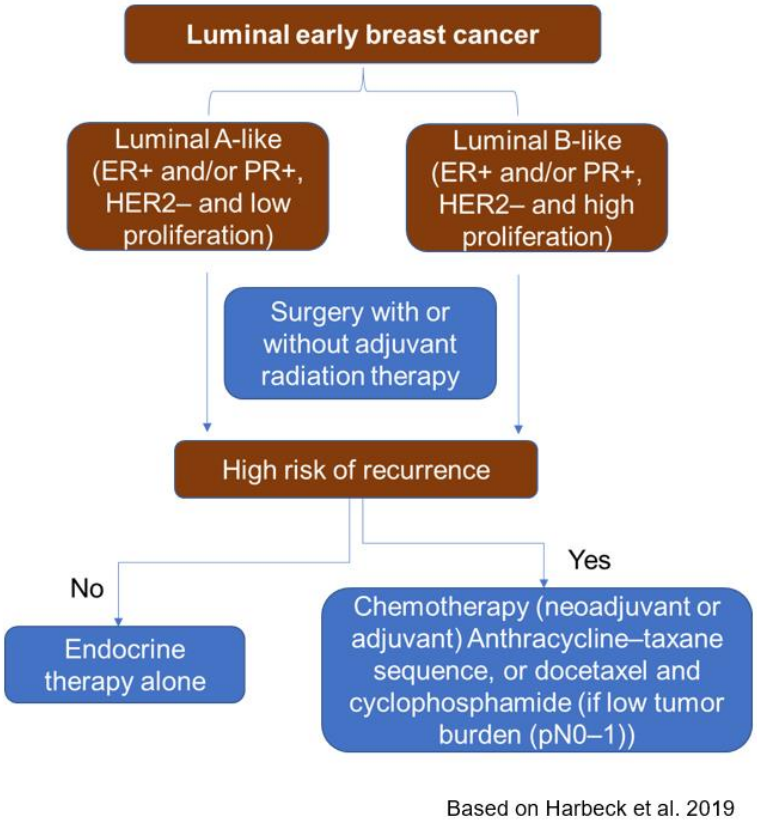


Figure 1-1 Algorithm for treating luminal early breast cancer [10]

many aspects of breast cancer management including screening, diagnosis, surgery, radiotherapy and systemic therapy [10].

Chemotherapy is a key component of systemic therapy and often prescribed along with endocrine therapy for HR+/ HER2– EBC. It targets all fast-growing cells including cancer cells, but thereby also impairs growth of normal cells like hair follicle cells, skin cells, and gastro-intestinal cells. Therefore, besides fighting cancer, chemotherapy frequently leads to unpleasant side effects like nausea, vomiting, fatigue, decreased appetite, changes in taste, hair loss, dry mouth, constipation, and myelosuppression [4, 17]. Most patients will experience some side effects of chemotherapy, but only a small percentage of them will actually substantially benefit from the cytotoxic therapy. According to an early meta-analysis, the absolute improvement for ten-year survival was about 10% for patients diagnosed before age 50, and under 3% for those diagnosed between age 50 and 69 [18]. Moreover, for node-negative patients under age 50, the ten-year survival rate was 77.6% if they had received chemotherapy and 71.9% if not [4, 18]. This means that out of 100 patients who received chemotherapy, only 5 patients actually benefited.

Considering the frequent and unpleasant reactions to cytotoxic therapies, avoiding chemotherapy in low-risk patients has frequently been advocated [4, 16, 19] (Figure 1-1). Besides clinical risk factors, intrinsic subtyping is also important for choosing a systemic treatment regimen [10]: the molecular patterns of breast cancer are heterogenous and possess prognostic/predictive significance for patients [20]. The intrinsic subtypes based on PAM50 analysis [9] of HR+/ HER2– breast cancers are mostly luminal A (around 60%) and luminal B (around 30%). The differences between the two luminal subtypes consist of the following: Luminal B tumors have higher expression of

proliferation/cell-cycle related genes or proteins, and lower expression of several luminal-related genes such as PGR (progesterone); luminal A tumors have a lower number of mutations across the genome and chromosomal copy-number changes, and less TP53 and more PIK3CA (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic) /MAP3K1 (mitogen-activated protein kinase kinase kinase 1) mutations. It is generally accepted that luminal A tumors have better prognosis and benefit less from chemotherapy than luminal B tumors [10, 21].

Intrinsic subtypes based on hierarchical clustering are not stable, have no accurate standard for diagnosis, and therefore hard to implement [22]. Therefore, surrogate intrinsic subtypes were proposed to stratify HR+/HER2– breast cancer into luminal A-like (ER (estrogen receptor) + and/or PR (progesterone receptor) +, HER2–, low proliferation) and luminal B-like (ER+ and/or PR+, HER2–, high proliferation) [23]. For evaluating proliferation, Ki-67 is a frequently used nuclear marker that can be evaluated by IHC (immunohistochemistry) staining. The cutoff point for Ki-67 to differentiate between luminal A-like and luminal-B like is not officially determined, but 14% and 20% are both widely accepted. Besides Ki-67 index, PR status is also used to predict prognosis: For patients with a Ki-67 index over 14%, PR < 20% indicates poorer prognosis than PR > 20% [24]. Although surrogate intrinsic subtyping cannot completely reproduce the PAM50 subtypes, it is clinically useful and practical. Studies compared the survival of patients with luminal A-like breast cancer who received chemotherapy and who did not, and found that a strong benefit of chemotherapy was not visible, neither in node-negative nor in node-positive patients [25-27].

Yet, the surrogate intrinsic subtyping is insufficient to support treatment choices in the rather complex clinical scenario of luminal breast cancer. Most importantly, it is difficult to distinguish between luminal A-like and luminal B-like subtypes in tumors with intermediate Ki-67 levels and therefore difficult to decide whether to apply chemotherapy or not [24, 28]. Enlightened by the molecular heterogeneity of breast cancer [9], the heterogeneity of gene expression was investigated to identify patients with HR+/HER2– EBC who have a low risk of recurrence and may therefore not benefit from chemotherapy [13, 14, 29-34].

1.2.2 Multigene assays as risk predictors

Several molecular assays have been successfully established and widely applied, namely Oncotype DX® (Genomic Health, Redwood City, CA, U.S.A), MammaPrint® (Agendia, Irvine, CA, U.S.A), Prosigna® (NanoString Technologies, Seattle, WA, U.S.A), EndoPredict® (Sividon Diagnostics GmbH, Cologne, Germany) [35]. Eligibility criteria are similar for these assays; they are suitable for patients with HR+/HER2– EBC that is lymph node-negative or has not more than three positive lymph nodes [34].

1.2.2.1 Oncotype DX®

Oncotype DX® evaluates expression of 16 target genes and 5 reference genes [36] via qRT-PCR in FFPE breast cancer samples [37]. The selected 16 target genes have important functions in tumor proliferation, invasion, and estrogen signaling. A recurrence score (RS) is offered based on the relative expression of these 16 genes to predict the risk of breast cancer recurrence in ten years [36, 38, 39].

During development of the Oncotype DX® algorithm in breast cancer, Paik et al. selected 250 candidate genes from published resources that were associated with disease outcomes and then reduced them to 16 genes, which were most closely correlated with long-term distant recurrence free survival, and five reference genes. Based on the published trial NSABP-14 (National Surgical

Adjuvant Breast and Bowel Project), they then tested the original 21-gene assay in 668 samples of which 194 samples were from premenopausal patients (under the age of 50). They found that distant recurrence was significantly less in the low-risk patients than that in the high-risk ones. According to their multivariate Cox model, the predictive power of the 21-gene recurrence score (RS) is independent of age and tumor size (NSABP-14) [36].

Two years later, the same study group released another publication based on the NSABP-20 trial and suggested that RS not only determines the likelihood of breast cancer recurrence but also independtly predicts the magnitude of chemotherapy benefit. In this validation study, 289 out of 651 patients were under the age of 50 (NSABP-20) [39]. Later in 2010, the predictive power of RS regarding locoregional recurrence was tested as well based on both NSABP-14 and NSABP-20. In total, 895 patients, among whom 293 patients were below the age of 50 when they received treatment, were included in this study. A significant association between RS and locoregional recurrence rate was found; multivariate Cox regression suggested that RS is a significant independent predictor of locoregional recurrence together with age [38]. Meanwhile, a case-control study, which calculated RS in 790 node-negative patients, among whom 209 patients were premenopausal, also verified the predictive value of RS for ten-year survival (Habel, et al.) [40]. From 2004 to 2015, the use of Oncotype® steadily increased in HR+ N0 patients and improved patient survival [41].

To further validate and refine the clinical utility of Oncotype DX® to reduce unnecessary use of chemotherapy in patients with HR+/ HER2–, axillary node-negative breast cancer, a prospective Trial Assigning Individualized Options for Treatment (TAILORx) was designed. According to the first report which was released in 2015, patients who had an RS of 0 to 10 had a sufficiently low five-year recurrence rate, even if they only received endocrine therapy. Among the analysed 1626 patients, 480 patients were premenopausal. Multivariate analysis showed that age is not associated with the recurrence rate (TAILORx) [42]. Similar evidence was also offered by the PlanB trial from West Germany Study Group (WSG) in 2017. Plan B was designed to investigate the potential of RS to help avoid chemotherapy overtreatment in HER2–, pN0 and pN1 EBC patients who had a high clinical risk of recurrence. As concluded in the report, patients with an RS ≤11 had an excellent five-year disease-free survival rate when treated with endocrine therapy alone (Plan B) [43]. Moreover, a retrospective study (109 out of studied 709 patients were younger than 50 years upon diagnosis) suggested that patients with micrometastases/1–3 positive nodes and RS ≤ 18 could safely omit chemotherapy (Stemmer, et al.) [44]. Clinical practice reports also validated the feasibility of using Oncotype DX® to reduce chemotherapy use in ER+/HER2− patients [45-47].

In 2018, an update from TAILORx was released. According to the results, ER+/HER2–/axillary node-negative patients older than 50 years who have a RS of less than 26 could omit chemotherapy safely. However, for patients who are younger than 50 years with an RS between 16 and 25, chemotherapy may be beneficial (TAILORx) [48]. Moreover, in 2021, the most recent update from TAILORx indicated that premenopausal women with one-three positive lymph nodes should receive chemotherapy, but postmenopausal patients could omit chemotherapy if they have an RS that is under 26 [49].

1.2.2.2 MammaPrint®

MammaPrint® evaluates the expression of 70 genes [14] via DNA microarray in fresh or freshly frozen breast cancer tissues or formalin-fixed, paraffin-embedded (FFPE) samples [37]. By measuring the relative expression of these genes, MammaPrint® stratifies tumors into a high or low-

risk group, which predicts patients' risk of recurrence within five years [50] overall survival [30], disease-free survival [29, 51], and benefit from adjuvant chemotherapy [52, 53].

MammaPrint® was initially developed in 2002 when the expression of 70 genes was measured by microarray analysis in 295 EBC patients, who were under the age of 53. The assay demonstrated the predictive power of the 70-gene signature regarding outcome with a higher accuracy than the traditional subtyping method (Van de Vijver, et al.) [29]. Twelve-year follow-up of this study presented a significant difference in long-term distant metastasis-free survival between patients defined as low-risk and those as high-risk [54]. The follow-up data of 302 patients who were all under the age of 61 from the TRANSBIG (Translational Breast International Group) consortium also suggested an independent prognostic value of Mammaprint® in node-negative premenopausal patients who had not received adjuvant systemic therapy (TRANSBIG) [51].

To study the clinical impact of MammaPrint® on AST (adjuvant systemic treatment) decision making, a prospective study was conducted, namely the microarRAy prognoSTics in breast cancER (RASTER) study. During risk estimation, a considerable discrepancy between clinical parameters and the MammaPrint® prediction was observed. Implementation of the MammaPrint® in daily clinical practice appeared feasible, since adding the MammaPrint® results to standard clinicopathological factors changed advice on adjuvant systemic treatment in nearly one-fifth of the patients [55]. For patients who used the 70-gene expression classifier to determine the regimen of adjuvant systemic treatment, RASTER then tracked their outcomes. In 2013, Drucker et al. reported the five-year follow-up results of RASTER and suggested that the low-risk group has a low distant recurrence rate and could omit chemotherapy without compromising outcomes. Although no multivariate analysis was performed, 292 of the 427 patients were younger than 50 years, so that premenopausal patients account for a substantial portion of the RASTER results (RASTER)[56].

To investigate the power of MammaPrint® to support clinicopathological tools while selecting patients for adjuvant chemotherapy, a randomized, phase 3 study MINDACT (Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy) was launched. The study enrolled 6693 women with EBC and determined their genomic (using MamaPrint®) and clinical risk (using Adjuvant! Online). Median age of the patients was 55 years. Among 1550 patients who had a high clinical risk plus a low genomic risk, the five-year distant metastasis rate of those who did not receive chemotherapy was 1.5% higher than those who received chemotherapy, suggesting that a considerable part of patients with high clinical risk received no significant benefit from chemotherapy. Therefore, the authors concluded that using MammaPrint® to guide treatment for patients with high clinical risk can reduce the application of unnecessary chemotherapy (MINDACT) [57]. Clinical practice also suggested that the use of MammaPrint® increased physicians' confidence in treatment decisions [58].

1.2.2.3 Prosigna®

Prosigna® measures the expression of 50 genes that are included in PAM50 gene signature [59, 60]. Prosigna® uses the Nanostring® nCounter mRNA detection system, examines FFPE samples [9], and generates a risk of recurrence (ROR) score which estimates patients' risk of distant recurrence over ten years based on the PAM50 gene signature, intrinsic subtype, tumor size, nodal status and proliferation score [61-63].

The algorithm of Prosigna® in breast cancer was developed based on PAM50 (prediction analysis of microarray) [63], which selected 50 genes that showed the highest correlation to each intrinsic

subtype from a list of 1906 “intrinsic” genes [20]. Wallden et al. tested the capacity of Prosigna® to define intrinsic subtypes and predict distant recurrences over ten years in ER+, node-negative patients treated with 5 years of adjuvant tamoxifen [59]. No age information of the patients was offered in this paper, but the original study, in which the samples were collected included a considerable number of premenopausal patients [64, 65]. Later, the prognostic value of Prosigna® was evaluated in 653 patients with HR+/HER2− EBC. Results showed that patients with a low ROR score have a high 15-year survival rate even without adjuvant therapy, suggesting that the ROR score has a high prognostic value (Oslo1). Among the 653 patients, 382 patients were under the age of 55 [31]. Clinical practice also suggested that applying Prosigna® in patients with ER+/HER2− EBC at low-to-intermediate risk of recurrence could change adjuvant therapy recommendation, increase physicians’ confidence in decision-making and improve patients’ emotional well-being [32].

Most recently, in 2020, a retrospective study analyzed the potential of PAM50 for predicting long-term breast-cancer survival (over 15 years of follow-up). In 1253 patients (median age 50 years), PAM50 was confirmed as an independent prognostic indicator for long-term, disease-free survival, irrespective of menopausal status (Pu, et al.). Yet, it is worth mentioning that this study adopted a NanoString nCounter analysis system with a custom-made codeset containing probes for 123 gene expression targets, rather than Prosigna® for direct analysis and it generated five intrinsic subtypes rather than four. So, results of this study do not directly represent the predictive power of Prosigna [66].

For predicting chemotherapy benefit in patients with ER+ EBC, Prosigna® is currently only recommended to predict the risk of recurrence in postmenopausal women treated with endocrine therapy alone [63, 64]. Evidence from a prospective, multicenter study suggested that the ROR score was correlated with the probability of distant recurrence at 10 years for postmenopausal patients with HR+ EBC who received no chemotherapy [67]. Meanwhile, the Austrian Breast & Colorectal Cancer Study Group (ABCSG)-8 trial also validated the ROR score for predicting the risk of distant recurrence in postmenopausal patients with ER+ EBC who were treated with adjuvant endocrine therapy [68].

1.2.2.4 EndoPredict®

EndoPredict® (EPclin®) measures the expression of eight proliferative and HR-associated genes and four reference genes via qRT-PCR in FFPE breast cancer samples [37]. The EndoPredict® algorithm generates a risk score based on the gene expression results, nodal status and tumor size, thus recognizing the patients who could safely forgo chemotherapy [69].

The original training set of EndoPredict® consisted of 964 ER+/HER2− tumors from patients treated by adjuvant tamoxifen. Among these patients, 245 patients were premenopausal [70]. EndoPredict® was verified to be an independent prognostic factor in the node-positive, chemotherapy-treated, ER+/HER2− EBC patients. In 2014, Martin, et al investigated the prognostic power in 555 ER+/HER2− tumors in the GEICAM 9906 trial which included both pre- and postmenopausal women. Patients in this trial were randomized to anthracycline-containing chemotherapy +/- paclitaxel followed by endocrine therapy; distant metastasis-free survival was the primary endpoint. The analysis suggested that EndoPredict® is prognostic not only in postmenopausal but also in premenopausal patients [71].

For predicting chemotherapy benefit, the initial clinical validation of EndoPredict® was based on two clinical aromatase inhibitor trials (Austrian Breast and Colorectal Cancer Study Group,

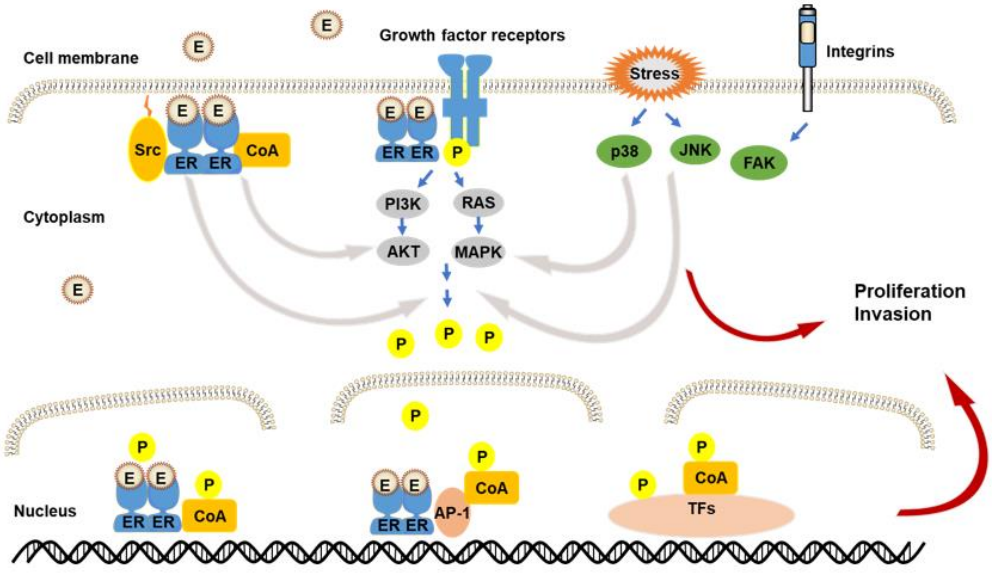
ABCSG-6 and ABCSG-8), which incorporated data exclusively from postmenopausal women. According to the report, the continuous EndoPredict® score is an independent predictor of distant recurrence in patients with ER+/HER2− EBC treated with endocrine therapy alone, regardless of nodal status [72]. Although no evidence implied that EndoPredict® would also be predictive for chemotherapy benefit in premenopausal patients, experience from clinical practice suggested that EndoPredict® could help to reduce chemotherapy use also in premenopausal patients with HR+/HER2−, T1–T2, and N0–N1 breast cancer and thus be practical in clinical routine [73].

1.2.3 Escalation of systemic treatment

Although most patients with luminal early breast cancer will fully recover after standard treatment, about one third of the patients will eventually develop distant metastasis. For patients who are at a high risk of developing distant metastasis, treatment “escalation” may be necessary. According to the definition made by the St. Gallen international expert consensus conference, “escalation” means to identify areas where optimal care may be achieved with “more” treatment [16]. For systemic treatment of HR+/HER2– EBC, “escalation” could be achieved through applying ovarian suppression in premenopausal patients, extending treatment duration in postmenopausal patients, and applying bisphosphonate in postmenopausal patients to prevent recurrence [16].

Meanwhile, resistance to current systemic treatment regimens, especially resistance to endocrine therapy is deemed as a major cause of poor prognosis [12, 22]. The underlying cause of endocrine resistance is complicated and could be associated with the following factors: 1) heterogeneity of ER expression, 2) changes in the metabolism of antiestrogens, 3) “late recurrence” phenotypes induced dormancy, 4) therapy induced adaptation, and 5) growth factors [12]. As for growth factors driving endocrine resistance, the most widely studied factors include the EGFR (epidermal growth factor receptor) superfamily, insulins/IGFs (insulin-like growth factors), MAPK (mitogen-activated protein kinase) and PI3K (phosphoinositide 3-kinases) /AKT (protein kinase B) /mTOR (mammalian target of rapamycin) signaling [12, 74].

ER-signaling is a complex biological pathway that controls a variety of functions such as proliferation, apoptosis, and angiogenesis. The function of ER signaling has a close association with proliferative-associated factors including EGFR, insulins/IGFs, MAPK, and PI3K/AKT/mTOR signaling (Figure 1-2) [75]. EGFR is a transmembrane glycoprotein, activation of which participated actively in cellular proliferation, differentiation, and survival. EGFR is abnormally activated by various mechanisms like receptor overexpression and mutation, and is associated with variety of human cancers [76]. The IGF system regulates multiple physiological processes, including mammalian development, metabolism, and growth. It is also widely implicated in cancer progression and identified as a clinically important therapeutic target [77]. MAPK signaling regulates a wide variety of cellular processes, including proliferation, differentiation, apoptosis, and stress responses. It is also actively involved in survival and development of tumor cells [78]. The PI3K/AKT/mTOR pathway is frequently activated in various human cancers and has been considered a promising therapeutic target [79]. Studies suggested that alteration of these elements can modulate ER activity or act as escape pathways to provide alternative proliferation and survival stimuli [75]



● E = estrogen; ER = estrogen receptor; CoA = coenzyme A; P = phosphate; PI3K = phosphoinositide 3-kinases; AKT = protein kinase B; RAS = rat sarcoma virus; MAPK = mitogen-activated protein kinase; JNK = c-Jun N-terminal kinase; FAK = focal adhesion kinase; AP-1 = activator protein 1; TFs = transcription factors

Based on Osborne et al. 2011

Figure 1-2 The mechanism of estrogen receptor (ER) in breast cancer [75]

Next to the extensively investigated proliferative pathways, CDK4/6 (cyclin-dependent kinases 4 and 6) has recently become a major target for treating endocrine resistant cancer. CDK4/6 is a key regulator of the cell cycle which functions through interfering with the phosphorylation of RB (retinoblastoma) protein. Inhibitors of CDK4/6 can induce cell-cycle arrest, invoke a senescence-like phenotype, modulate the mitogenic kinase signaling, and enhance immunity and improve patient survival [80].

For battling endocrine resistance in advanced breast cancer, targeted therapies are clinically available that modulate proliferation-related pathways (e.g., cyclin D-cyclin dependent kinase (CDK) 4/6-inhibitor of CDK4 (INK4)-retinoblastoma (Rb) pathway [81], PI3K (phosphoinositide 3-kinase)/AKT (proteinase B)/ mTOR (mammalian target of rapamycin) pathway [10, 82]). Recently, a first CDK 4/6 inhibitor, abemaciclib, has become available for patients with early luminal cancer based on results of the monarch-E trial [83]: further studies are ongoing to determine whether other CDK 4/6 inhibitors also bring survival benefit to patients with high risk of recurrence without causing significant harm to quality of life.

1.3 The necessity of establishing molecular tests for premenopausal patients

Although it is more often found in elderly women, breast cancer, and especially HR+/ HER2– breast cancer is associated with a worse prognosis in young patients [33, 84-86]. The following biological causes for recurrences have been reported for breast cancer in premenopausal patients: it is more commonly resistant to endocrine therapy [87] and frequently diagnosed with a higher histological grade [84, 88]; it presents more aggressive molecular patterns and could thus have a unique biology that requires novel therapeutic strategies [89-91].

Similar to postmenopausal patients, premenopausal patients with a low risk of recurrence could omit chemotherapy without compromising their overall survival [31, 92]. However, there are concerns that the gene expression assays may lead to false conclusions about the actual risk of recurrence in premenopausal patients. Firstly, most of the tests have been established mainly by using cancer samples from postmenopausal patients who are the majority of all breast cancer patients [93, 94] (Table 1-1). The risk calculation algorithm developed from postmenopausal cancers may fail to accurately reflect the recurrence risk in premenopausal patients. The second concern is that hormonal fluctuations in premenopausal patients may affect the risk scores calculated by some multigene tests. *In vitro* studies suggested that concurrent treatment with estrogen and progesterone in breast cancer cells could regulate growth factor pathways, result in switching the PAM50-determined intrinsic breast cancer subtype from luminal A to basal-like, and increase the Oncotype DX® recurrence score [95]. Therefore, the clinical utility of the four multigene profiling assays in premenopausal patients requires detailed evaluation and eventual adjustment of the prediction algorithm.

Table 1-1 Clinical validation of gene expression tests in premenopausal patients with HR+/HER2−, early breast cancer (EBC)

Study	Type	Proportion of Pre-P	LN	Focus
Oncotype DX®				
NSABP-B14 [36]	Retro	29.0% under age 50	N	Prognosis*
NSABP-B20 [48]	Retro	44.3% under age 50	N	Chemo-benefit**
NSABP-14 and NSABP-20 [38]	Retro	32.7% under age 50	N	Prognosis
Habel, et al. [40] +	Case-control	26.4% under age 50	N	Prognosis
TAILORx [42]	Pro	29.5%	N	Prognosis
Plan B [43]	Pro	The median age was 56	58.8% N	Prognosis
Stemmer, et al. [44]	Retro	15.3% under age 50	P	Chemo-benefit
TAILORx [48]	Pro	38.9%	N	Chemo-benefit
TAILORx [49]	Pro	33.2%	P	Chemo-benefit
MammaPrint®				
van de Vijver, et al. [29] +, Drukker, et al. [54] +	Retro	All under age 53	51.2% N	Prognosis
TRANSBIG [51]+	Retro	All under age 61	N	Prognosis

Table 1-1 (continued)

RASTER [56] +	Pro	68.3% under age 50	N	Prognosis
MINDACT [57] +	Pro	33.2% under age 50	79.0% N	Prognosis
Prosigna®				
Oslo1 [31] +	Retro	58.5% under age 55	64.1% N	Prognosis
EndoPredict®				
GEICAM 9906 [71]	Retro	54.0%	Positive	Prognosis

HR = Hormone Receptor; HER2 = Human Epidermal Growth Factor Receptor 2; EBC = Early-stage Breast Cancer; Pre-P = premenopausal patients; LN = lymph node; Retro = retrospective; Pro = prospective

+ This study contains part of HR− or HER2+ patients.

* In studies focusing on predicting ‘‘prognosis’’, the prognostic value of one assay was validated by comparing the prognosis of patients in different risk groups.

** In studies focusing on predicting ‘‘chemo-benefit’’, the predictive value of one assay was validated by comparing the prognosis of patients in certain risk group who received endocrine therapy alone and who received both endocrine therapy and chemotherapy.

New tools are in urgent need to more precisely predict risk of recurrence for premenopausal patients with HR+/ HER2– EBC. These tools will help clinicians to further individualize treatments, including decisions on chemotherapy use, type and duration of endocrine therapy, and to identify patients who may need novel strategies [86].

2. Material and Methods[1]

2.1 Patient

2.1.1 Study design

We performed a case-control study in premenopausal patients with luminal-like early breast cancer. Two hundred and seventy-eight premenopausal patients were included in the original cohort. All patients had been treated at the LMU breast center after 1998 and followed for more than ten years after the first surgery. They were originally diagnosed with early-stage invasive breast cancer, and their histological subtype was confirmed to be hormone receptor-positive (HR+) and human epidermal growth factor receptor 2-negative (HER2–). Treatment regimens were selected based on guidelines and according to investigators' choices. A total of 124 patients were lost to follow-up, and for 57, there was not enough tumor tissue left for RNA extraction. In total, we identified 97 premenopausal patients with sufficient follow-up duration of whom 48 patients had developed distant metastases during follow-up(M1) and 49 patients had not (M0; shortest follow-up was 10 years) (the study design is summarized in Figure 2-1) [96].

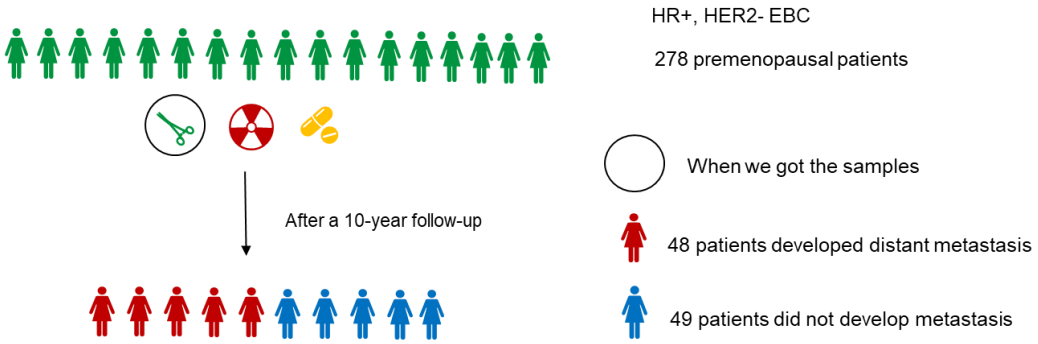


Figure 2-1 Study design [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

2.1.2 Ethical approval

This project is permitted by the Institutional Review Board of the Ludwig-Maximilians-University of Munich (LMU) (Germany) (Number: 19–745), and the original approval is submitted along with this dissertation.

2.1.3 Sample collection

Patients' clinical information was documented in the patient records. Tumor samples from the primary surgery were stored in formalin-fixed paraffin-embedded (FFPE) blocks and under room temperature. Histopathological diagnosis and classification were confirmed by two experienced pathologists at the LMU Institute of Pathology. HR and HER2 scoring were performed during routine diagnostics: HR+ was defined as ER and/or PR with immunohistochemistry (IHC) scores

[1] Part of the material and methods has been published by us in "Ni et al., Molecular Prognostic Factors for Distant Metastases in Premenopausal Patients with HR+/HER2- Early Breast Cancer. J Pers Med 2021, 11(*9*)*".* According to the regulation of the publisher, the authors retain the copyright and this article was licensed under an open access Creative Commons CC BY 4.0 license. The permission from the publisher is submitted with this dissertation.

of at least 1/100 following the international council on archives (ICA) standard or 1/12 according to the immunoreactive score (IRS) standard; HER2 was considered negative with IHC scores of 0–1+ or IHC 2+ and negative fluorescence in situ hybridization (FISH) analysis.

Sections from (FFPE) tumor blocks were prepared followed by hematoxylin-eosin (H&E) staining of one slide. The process of collecting samples is summarized in Figure 2-2. Microdissection was performed by specialists in the pathology institute on areas with a minimum percentage of 50% tumor cells from subsequent unstained sections and used for RNA preparation. Detailed information for every tumor sample can be found in Table 2-1 [96].

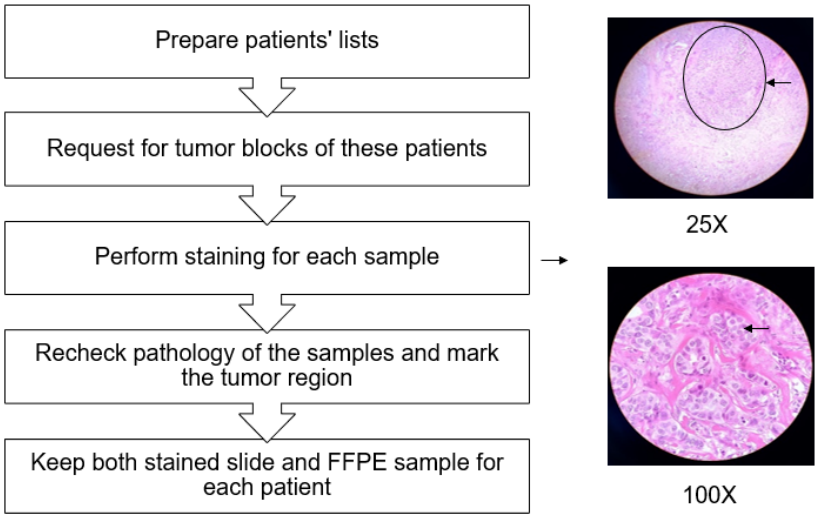


Figure 2-2 Work flow for sample collection

Table 2-1 Tumor content in FFPE samples [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

NO.	G	Size (cm×cm)	Percent	NO.	G	Size (cm×cm)	Percent
1	M0	0.8×0.8	60%	50	M1	0.4×0.1	70%
2	M0	0.8×1.0	80%	51	M1	1.0×0.5	80%
3	M0	0.5×0.3	80%	52	M1	1.6×0.5	90%
4	M0	0.8×0.5	70%	53	M1	0.5×0.5	70%
5	M0	0.6×0.6	60%	54	M1	0.7×0.7	70%
6	M0	1.0×0.8	90%	55	M1	0.4×0.4	90%
7	M0	0.6×0.6	80%	56	M1	1.5×1.0	80%
8	M0	0.6×0.6	70%	57	M1	1.0×1.0	90%
9	M0	0.5×0.6	70%	58	M1	0.5×0.4	80%
10	M0	1.0×0.5	50%	59	M1	0.5×0.5	70%
11	M0	1.0×0.6	90%	60	M1	0.5×0.3	80%
12	M0	1.0×0.7	70%	61	M1	0.5×0.2	70%

Table 2-1 (continued)

13	M0	0.8×0.5	80%	62	M1	1.0×0.7	90%
14	M0	0.6×0.5	70%	63	M1	0.8×0.2	50%
15	M0	0.6×0.5	60%	64	M1	1.0×0.9	70%
16	M0	0.6×0.6	60%	65	M1	0.8×0.5	50%
17	M0	1.0×0.5	80%	66	M1	2.0×0.5	80%
18	M0	0.7×0.5	60%	67	M1	1.0×0.5	70%
19	M0	0.6×0.5	60%	68	M1	0.5×0.5	80%
20	M0	0.8×0.8	80%	69	M1	0.6×0.6	80%
21	M0	0.6×0.5	80%	70	M1	0.5×0.5	80%
22	M0	0.6×0.6	70%	71	M1	1.0×0.6	70%
23	M0	0.6×0.5	90%	72	M1	1.0×0.8	70%
24	M0	1.0×0.9	70%	73	M1	1.1×1.0	80%
25	M0	0.6×0.6	60%	74	M1	1.0×0.6	70%
26	M0	0.7×0.5	80%	75	M1	0.2×0.2	60%
27	M0	0.5×0.5	90%	76	M1	1.2×0.5	90%
28	M0	1.0×0.5	80%	77	M1	2.0×0.5	70%
29	M0	0.7×0.5	80%	78	M1	0.8×0.5	80%
30	M0	0.7×0.6	50%	79	M1	1.5×0.5	70%
31	M0	0.6×0.3	60%	80	M1	1.0×1.0	60%
32	M0	1.0×0.5	80%	81	M1	0.5×0.5	90%
33	M0	0.8×0.2	70%	82	M1	0.7×0.2	70%
34	M0	0.7×0.7	70%	83	M1	1.0×0.9	80%
35	M0	0.5×0.2	50%	84	M1	0.5×0.2	60%
36	M0	0.5×0.5	70%	85	M1	1.0×0.7	70%
37	M0	1.0×0.5	70%	86	M1	0.5×0.5	80%
38	M0	1.0×0.5	60%	87	M1	1.0×0.8	50%
39	M0	0.5×0.2	80%	88	M1	0.5×0.5	80%
40	M0	1.2×1.0	90%	89	M1	0.5×0.5	70%
41	M0	0.6×0.6	70%	90	M1	0.6×0.6	70%
42	M0	0.4×0.4	60%	91	M1	0.5×0.4	50%

Table 2-1 (continued)

43	M0	0.7×0.7	80%	92	M1	0.5×0.7	80%
44	M0	0.7×0.7	70%	93	M1	0.8×0.6	80%
45	M0	1.0×1.0	80%	94	M1	0.2×0.2	50%
46	M0	0.5×0.3	60%	95	M1	0.8×0.4	80%
47	M0	1.0×0.5	70%	96	M1	0.6×0.5	60%
48	M0	1.0×1.2	90%	97	M1	1.5×0.6	70%
49	M0	1.0×0.3	70%				

* G = group; areas containing invasive tumor cells were circled and measured; the tumor percentage was counted according to following rules. Except for tumor cells, the other components include elastin, necrosis, white blood cells, vessels, and normal cells. 1) 50%. Half of the marked region consists of tumor cells. The quality of this slide is not satisfying, but usable. 2) 60%. This marked tumor region has a fair content of tumor cells. 3) 70%. This marked tumor region has a nearly high tumor cell content. 4) 80%. This marked tumor region has a high tumor cell content. 5) 90%. This marked tumor region has a very high tumor cell content. The marked area was then scratched down from 4-8 corresponding unstained FFPE tissue sections which are 2-5μm thick. The number of needed FFPE tissue sections were decided based on the size of the marked tumor area, an area which is less than 5×5 mm^2 needs 8 sections, an area which is more than 10×7 mm^2 needs 4 sections, and area which is between 5×5 mm^2 and10×7 mm^2 needs 6 sections.

2.2 Reagents

2.2.1 RNA extraction

1. QIAGENE RNeasy® FFPE kit (50 preps, NO.73504) (QIAGEN, Hilden, Germany)

1) RNeasy Min Elute spin columns (pink) (each in a 2ml collection tube)
2) Collection tubes (1.5ml)
3) Collection tubes (2ml)
4) Buffer RBC (red blood cell) (avoid bleach)
5) Buffer PKD (protein kinase D)
6) RNase-Free DNase Ⅰ (lyophilized)
7) RNase-Free Water (for use with RNase-Free DNase Ⅰ)
8) DNase Booster Buffer
9) Buffer RPE (R-Phycoerythrin) (concentrate)
10) RNase-Free Water

2. Deparaffinization solution (16ml, NO.19093) (QIAGEN, Hilden, Germany)
3. RNase-Free DNase Set (50 preps, NO.79254) (QIAGEN, Hilden, Germany)

1) 500 units RNase-free DNase Ⅰ
2) RNase-free buffer RDD (to dilute DNase Ⅰ) and RNase-free water

4. Filter tip (1000μl, 8×128, NO.990352) (QIAGEN, Hilden, Germany)
5. Rotor adapter (10×24, NO.990394) (QIAGEN, Hilden, Germany)
6. SafeSeal reaction tube (2ml, 250, NO.72.695.700) (SARSTEDT, Nümbrecht, Germany)
7. Reaction tube (1.5ml, 2×250, NO.72.690.001) (SARSTEDT, Nümbrecht, Germany)
8. Pipette tips (2.5μl, 10×96, NO.69005) (Biozym, Hessisch Oldendorf, Germany)

9. Pipette tips (10µl, 50×96, NO.VT0210X) (Biozym, Hessisch Oldendorf, Germany)
10. Pipette tips (100µl, 10×96, NO.VT0230) (Biozym, Hessisch Oldendorf, Germany)
11. Pipette tips (1250µl, 10×96, NO.VT0270) (Biozym, Hessisch Oldendorf, Germany)
12. Scalpel holder (13cm, NO.502) (BAYHA, Tuttlingen, Germany)
13. Scalpel blade (NO.23) (BAYHA, Tuttlingen, Germany)

2.2.2 Nanostring analysis

1. Nanostring Master Kit (9×12 reactions)(Nanostring technology, Seattle, WA, USA)

1) Reporter Codeset
2) Capture Codeset
3) Hybridization buffer
4) 12-tube hybridization strips
5) Cartridge
6) Sheath
7) Tips
8) Reaction plates

2. Pipette tips (2.5µl, 10×96, NO.69005) (Biozym, Hessisch Oldendorf, Germany)
3. Pipette tips (10µl, 50×96, NO.VT0210X) (Biozym, Hessisch Oldendorf, Germany)
4. Pipette tips (100µl, 10×96, NO.VT0230) (Biozym, Hessisch Oldendorf, Germany)
5. Pipette tips (1250µl, 10×96, NO.VT0270) (Biozym, Hessisch Oldendorf, Germany)
6. RNase-Free Water from RNase-Free DNase Set (50 preps, NO.79254) (QIAGEN, Hilden, Germany) (to dilute RNA)
7. Nanostring BC360® (Breast cancer 360) panel (an RLF file in one USB-stick, for further analysis) (Nanostring technology, Seattle, WA, USA)

2.3 Equipment

2.3.1 RNA extraction

1. Eppendorf® microcentrifuge 5417 (Eppendorf, Hamburg, Germany)
2. Eppendorf® Thermomixer comfort with 2ml block (Eppendorf, Hamburg, Germany)
3. QIAcube (NO.080001578) (QIAGEN, Hilden, Germany)
4. Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA)
5. Freezers (–80℃, –20℃, 4℃) (Eppendorf, Hamburg, Germany)/(Thermo Fisher Scientific, Waltham, MA, USA)

2.3.2 Nanostring

1. Eppendorf® centrifuge 5430 (Eppendorf, Hamburg, Germany)
2. Nanostring PTHMG001 nCounter® Prepstation (Nanostring technology, Seattle, WA, USA)
3. Nanostring PTHMG002 nCounter® Analyzer (NO.2-4-103) (Nanostring technology, Seattle, WA, USA)
4. Neolab® D-6020 Mini star centrifuge (neoLab Migge GmbH, Heidelberg, Germany)
5. Sunlab® Replacement rotor for 2×8 0.2 ml PCR-Strips D-8554 (Sunlab, Aschaffenburg, Germany)
6. Veriti® Thermal Cycler 96-well (Thermo Fisher Scientific, Waltham, MA, USA)

7. Freezers (–80℃, –20℃, 4℃) (Eppendorf, Hamburg, Germany)/(Thermo Fisher Scientific, Waltham, MA, USA)

2.4 Tools/Software

1. Naonostring® nSolver 4.0 (Nanostring technology, Seattle, WA, USA)
2. GSEA(gene set enrichment analysis) 4.1.0 (http://www.gsea-msigdb.org/gsea/index.jsp)
3. R studio (version 1.4.1106) (https://www.rstudio.com/)
4. R (version 4.0.4) (https://www.r-project.org/)
5. SPSS 23.0 (IBM, Armonk, NY, USA)
6. Graphpad prism 5 (GraphPad Software, Inc., San Diego, CA, USA)
7. STRING (search tool for the retrieval of interacting genes/proteins) tool (https://string-db.org/)
8. Kaplan-Meier Plotter Tool (https://kmplot.com/analysis/)

2.5 Methods

2.5.1 RNA extraction

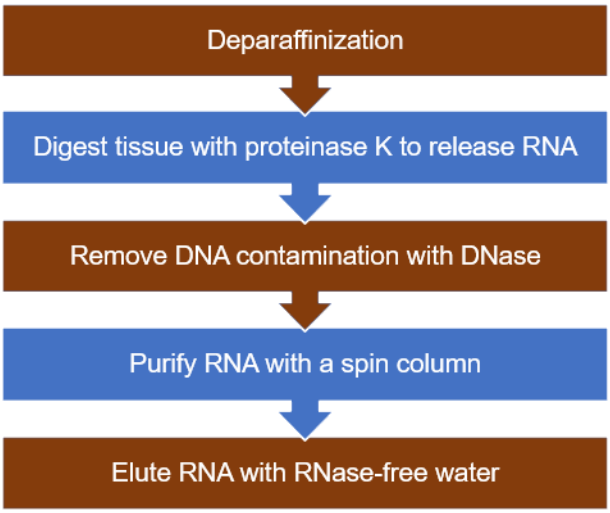


Figure 2-3 Overview of the RNA extraction procedure

The extraction was conducted following the protocol which was supplied along with the kit. The basic procedures include:

1. Prepare the reagents and equipment.
2. Scratch down the tumor tissue from the FFPE sections. Use the H&E-stained slide as a reference (the tumor region should be marked with a marker before wise), scratch down the tumor region from the sections with a scalpel (easier when the tip of the scalpel is soaked with deparaffinization solution).
3. Digest the tissue and release RNA. Detailed instructions could be found in QIAGEN RNeasy® FFPE handbook [97].

4. Place the sample tubes, rotor adapter, filter tips, DNase mix (for 12 samples: 145µl DNase mix + 232µl DNase Booster Buffer), and reagent bottle rack (Buffer RBC, 96–100% ethanol, Buffer RPE, RNase-free water) into the right position in the QIAcube instrument.
5. Set the QIAcube instrument as follows and start the extraction procedure.

Application	RNA
Kit	RNeasy® FFPE kit, protocol 1770
Sample material	1–2 FFPE tissue sections
Short protocol name	DNase digest
Version	1
Full protocol name	Purification of total RNA from up to two 10µm FFPE tissue sections
Elution volume	20µl
Software versions	Firmware version FIW-50-001-J_FW_MB.hex and PLC program version; FIW-50-002-G-PLC_MB.prs or higher; available from the QIAcube Web Portal

6. Remove the collection tubes from the QIAcube after accomplishment and measure the RNA concentration with Nanodrop 1000.
7. RNA quality control: RNA concentration should be more than 20ng/µl, A260/A280 should be within 1.8 and 2.0.

2.5.2 Nanostring profiling

Nanostring profiling was accomplished following the protocol provided along with the kits. Briefly, the profiling has the following three steps: hybridization, purification and immobilization, counting and analysis [98, 99] .

1. Set up the hybridization. Detailed instructions could be found in NanoString CodeSet Hybridization manual [100].
2. Run the mixed hybridization reactions on the nCounter® Prep Station
3. Scan the sample cartridge with the nCounter® Analyzer, and set the parameter as follows

Sensitivity	Very high
Panel	NS_BC_360_V2.0 rlf

4. Scanned raw data (RCC files) should be imported into the nSolver 4.0 for data quality control and normalization. Detailed instructions could be found in nSolver 4.0 Quick Start Guide and nSolver 4.0 User Manual that are included in the USB-stick provided along with the Master Kit. Briefly, the software will exclude outliers that have strange data of positive control, negative control and house keeper genes. As for normalization, genes in the TIS (tumor inflammation signature) signature are normalized using a ratio of the expression value to the geometric mean of the housekeeper genes used only for the TIS signature, genes in the PAM50 signature are normalized using a ratio of the expression value to the geometric mean of the housekeeper genes used only for the PAM50 signature, and other genes are normalized

using a ratio of the expression value to the geometric mean of all housekeeping genes on the panel.

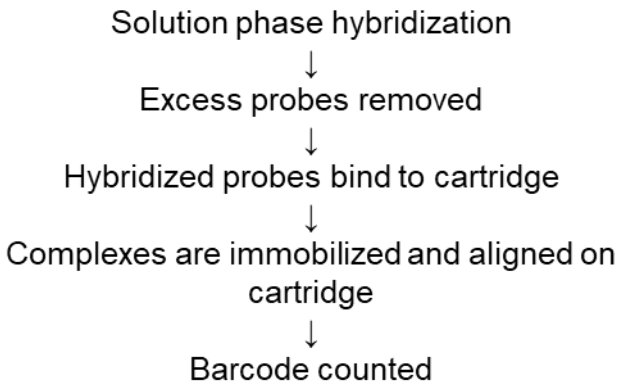


Figure 2-4 Methodology of Nanostring nCounter® technology [99]

2.5.3 Data analysis

1. Differential expression of BC360® signatures and genes

Among the 97 tumor samples, 85 passed the data quality control and were further analyzed. In total, 46 signatures and 758 genes were analyzed. Differential expression is fit on a per gene or per signature basis using a linear model for analyses without a blocking factor. The statistical model uses the expression value or signature score as the dependent variable and fits a grouping variable as a fixed effect to test for differences in the levels of that grouping variable.

Expression (gene or signature)= μ+Group+ε

P-values are adjusted within each analysis, gene or signature, and on the grouping variable level difference t-test using the Benjamini and Yekutieli False Discovery Rate (FDR) adjustment to account for correlations amongst the tests [101, 102]. All models are fit using the limma package in R [103].

2. PAM50 subtypes and Risk of Recurrence (ROR)

PAM50 subtype calls are the result of a three-step algorithm. The first step involves a scaling using two sets of scaling factors to bring the housekeeper and reference sample expression values into the scale necessary for the next step. This second step calculates the correlation between the observed scaled expression for the PAM50 genes and a centroid for each of the four subtypes resulting in a set of four correlation values for each sample. The remaining step is to identify the subtype correlation with the greatest value and set that subtype as the subtype call for that sample. ROR scores are the result of a multiple step algorithm. The first step involves a scaling using two sets of scaling factors to bring the housekeeper and reference sample expression values into the scale necessary for the next step. This second step calculates the correlation between the observed scaled expression for the PAM50 genes and a centroid for each of the four subtypes that is different than that for calling subtypes and results in a set of four correlation values for each sample. The next step is to calculate a proliferation score for each sample, followed by taking a weighted sum of the proliferation score and the four subtype correlations. The penultimate step is to calculate the weighted sum of this last score and a binned tumor size measure. This last score is then scaled to be between 0 and 100 [60, 62].

3. GSEA analysis

Gene Set Enrichment Analysis (GSEA) determines whether an a priori defined set of genes shows statistically significant, concordant differences between two groups (Table 2-2). The GSEA analysis of 751 genes (7 genes were not included in the GSEA database) between M1 and M0 patients was performed in the GSEA 4.1.0 software following the instructions. FDR q-value < 0.25 was considered significant (an FDR of 25% indicates that the result is likely to be valid 3 out of 4 times) [104].

Table 2-2 Tested GSEA sets [104]

Gene set	Annotation	Source
H	Hallmark gene sets	Coherently expressed signatures that represent well-defined biological states or processes
C1	Positional gene sets	Each human chromosome and cytogenetic band
C2	Curated gene sets	From online pathway databases, publications, and knowledge of domain experts
C3	Regulatory target gene sets	Based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites
C4	Computational gene sets	Defined by mining large collections of cancer-oriented microarray data
C5	Ontology gene sets	Consist of genes annotated by the same ontology term
C6	Oncogenic signature gene sets	Defined directly from microarray gene expression data from cancer gene perturbations
C7	Immunologic signature gene sets	Represent cell states and perturbations within the immune system

4. Analysis of differentially expressed genes

The gene expression between groups were further analyzed and compared by using limma package in R. The differentially expressed genes (DEGs) were selected out based on the following standard: $p < 0.05$ and abs (logFCsingle gene)> (mean(abs(logFC all genes))+2SD(abs(logFC all genes))). Heatmaps and volcano plots of Signatures/DEGs were created with the pheatmap and “ggplot2” functions in R. respectively. GO (Gene Ontology) analysis and KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis of DEGs were performed by using the Clusterprofiler package in R [105] to investigate the functions of the DEGs. STRING analysis was conducted online to investigate the functional interactions of the identified DEGs [96]. Crucial codes used in R were summarized in Appendix A.

5. Survival analysis

The “survfit” and “ggsurvplot” function in R were used to fit and plot Kaplan–Meier curves, respectively. And for survival curve plotting, the median of the observed gene expression or signature data was set as the cut point. The Cox proportional hazards regression model in the SPSS 23.0 software was used to carry out univariate and multivariate survival analyses. Statistically significant variables in univariate analyses were included in the multivariate analysis. For multivariate analysis, the method “Forward Stepwise (likelihood ratio)” was used to filter out the most significant factors in the survival model. There were three steps in the overall screening of “Forward Stepwise (likelihood ratio)”. At each step, the algorithm would select out the most significant prognostic factor to be included into the survival model (at step two and step three, the impact of the already selected one or two factors would be considered, therefore, the factor should have independent prognostic values from the factors that are already in the model). The correlation analysis between prognostic factors was carried out using the Spearman correlation model. A p-value < 0.05 was considered significant.

6. Other statistical analysis

Other comparative analyses between groups were performed using SPSS 23.0 software. For clinical parameters, linear variables were compared using Mann–Whitney test; non-linear variables were compared using Person Chi-square test or Fisher’s exact test. PAM50 subtype scores were compared with the Mann–Whitney test and plotted in GraphPad PRISM 5. All significance tests (where applicable) were two-tailed.

7. Online database validation

The Kaplan Meier plotter is a widely accepted tool to assess the correlation between the expression of genes and survival in more than 25,000 samples from various tumor types including breast cancer. Sources for the databases include GEO (gene expression omnibus), EGA (European genome-phenome archive), and TCGA (the cancer genome atlas) [106].

We used this tool to confirm the prognostic value of survival-related genes for relapse-free survival (RFS). The analyses were performed following the instructions, and the inclusion criteria for patients’ selection were: ER+ (IHC)/HER2− (array), had at least 10 years follow-up, received adjuvant systemic treatment (the menopausal status was unavailable). The survival-related genes were tested both as single genes and as a signature (the mean expression of genes was calculated). During the test process, mRNA gene chip data of 2301 breast cancer patients across 55 independent databases was included [106].

3. Results[2]

3.1 Patient characteristics

3.1.1 Follow-up information

The study design was summarized in the material and method part. In brief, our study included ninety-seven premenopausal patients who were treated for invasive HR+/HER2− early breast cancer. For patients who did not develop metastases during the follow-up period (M0 group, n = 49), follow-up time was between 121 and 191 months, with a median follow-up of 149 months. For patients who developed metastases (M1 group, n = 48), distant metastasis-free survival (DMFS) was between 7 and 184 months, with a median survival of 54 months [96] The survival swimmer plot shows the follow-up time or DMFS for each patient (Figure 3-1).

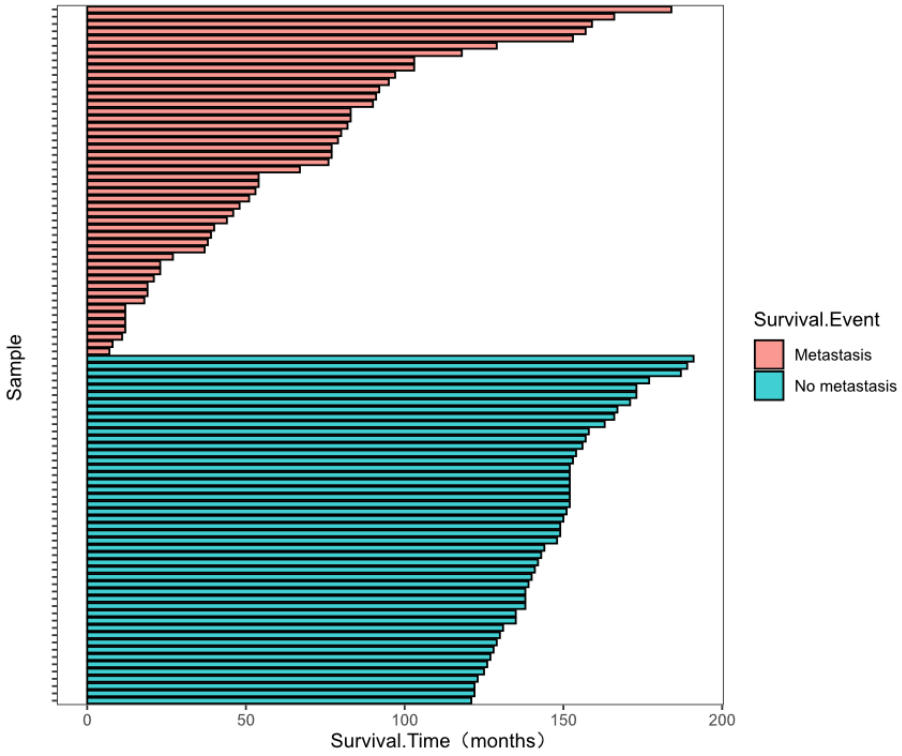


Figure 3-1 Survival swimmer plot (each line represents one patient)

3.1.2 Patient information

Clinical patient characteristics was summarized in Table 3-1. According to the analysis, patients who developed metastasis has larger tumor size (p =0.037), higher tumor grade (p =0.040), higher tumor stage (p =0.019), and more lymph node involvement (p =0.002). Patients received

[2] Part of the results has been published by us in "Ni et al., Molecular Prognostic Factors for Distant Metastases in Premenopausal Patients with HR+/HER2- Early Breast Cancer. J Pers Med 2021, 11(*9)".* According to the regulation of the publisher, the authors retain the copyright and this article was licensed under an open access Creative Commons CC BY 4.0 license. The permission from the publisher is submitted with this dissertation.

treatment according to applicable guidelines: all patients received surgery and endocrine therapy. Application of chemotherapy or radiotherapy was decided based on patients' clinical risk of recurrence and surgical choice [96].

Table 3-1 Patient clinical characteristics [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

Parameters		**M1 (*n* = 49)**		**M0 (*n* = 48)**		***p*-value**	**Test-method**
Age at diagnosis (y)		43(30-50)		47(29-50)		0.292	Mann-Whitney U test
Tumor size (cm)		2.3(0.2-6.3)		1.7(0.3-8.0)		0.037	Mann-Whitney U test
Grade	1	1	2.1%	6	12.2%	0.040	Pearson Chi-Square test
	2	26	54.2%	31	63.3%		
	3	21	43.8%	12	24.5%		
pT	1	17	35.4%	31	63.3%	0.019	Pearson Chi-Square test
	2	25	52.1%	13	26.5%		
	3	6	12.5%	5	10.2%		
pN	0	13	27.1%	32	65.3%	0.002	Fischer's exact test
	1	19	39.6%	11	22.4%		
	2	11	22.9%	4	8.2%		
	3	5	10.4%	2	4.1%		
Surgery	Lumpectomy	31	64.6%	35	71.4%	NA	NA
	Mastectomy	17	35.4%	14	28.6%		
ALND	Yes	43	89.6%	24	49.0%	NA	NA
	No	5	10.4%	25	51.0%		
Radio-therapy	Yes	39	86.7%	44	100.0%	NA	NA
	No	6	13.3%	0	0.0%		
Chem-other-apy	Yes	41	89.1%	29	60.4%	NA	NA
	No	5	10.9%	19	39.6%		
Taxane	Yes	23	62.2%	11	47.8%	NA	NA
	No	14	37.8%	12	52.2%		
Antra-cycline	Yes	32	86.5%	22	100.0%	NA	NA
	No	5	13.5%	0	0.0%		
Endo-crine therapy	TAM	31	81.6%	33	75.0%	NA	NA
	AI + GnRHa	2	5.3%	3	6.8%		
	sequence of both	5	13.2%	8	18.2%		

M1: Metastasis, M0: No metastasis, pT: pathological tumor stage, pN: pathological node status, ALND: axillary lymph node dissection, TAM: tamoxifen, AI: aromatase inhibitor, GnRHa: gonadotropin-releasing hormone agonist.

3.1.3 Survival relevance of clinical parameters

Among the clinical parameters, pN and pT were associated with survival. Specifically, compared to negative lymph node, positive lymph node is associated with poorer survival (p = 0.001); compared to stage one, stage higher than one is associated with poorer survival (p = 0.033). Tumor side, size, and histological type are not significantly associated with survival. Considering that the treatment option is highly dependent on the clinical parameters, no survival analysis was carried out specifically by treatment group (Table 3-2).

Table 3-2 Univariate survival analysis of clinical parameters

Factors	*p*-value	HR	95%CI	
			lower	upper
pN (positive vs negative)	0.001	3.2	1.6	6.4
pT (more than one vs one)	0.033	2.0	1.1	3.7
Tumor size	0.6	1.1	0.9	1.3
Tumor grade (three vs less than three)	0.1	1.7	0.9	3.1

3.2 PAM50 analysis

3.2.1 Subtype distribution

Eight-five of the 97 tumor samples passed data quality control and were analyzed. PAM50 (prediction analysis of microarray) analysis confirmed that 81 of the 85 tumor samples had luminal subtype. There was a higher proportion of luminal A tumors in the overall cohort and in each subgroup (M0, M1). A slightly higher proportion (not statistically significant) of luminal A subtypes (M0, 64%; M1, 56%) and lower luminal B subtypes (M0, 33%; M1, 37%) was noticed in patients who had not developed distant metastasis (Table 3-3) [96].

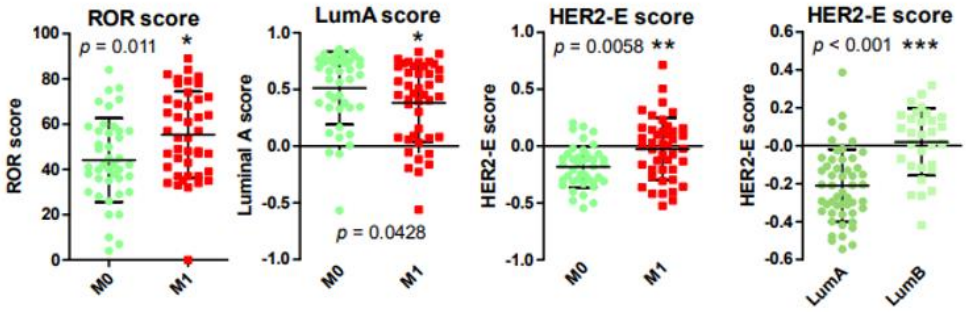
Table 3-3 Subtypes' distribution

Subtype	M1(n=43)		M0 (n=42)		*p*-value	Test method
Luminal A (n = 51)	24	55.8%	27	64.3%	0.425	Pearson Chi-Square test (Luminal A vs non -Luminal A)
Luminal B (n = 30)	16	37.2%	14	33.3%		
Basal (n=2)	1	2.3%	1	2.4%		
HER2-enriched (n=2)	2	4.7%	0	NA		

3.2.2 PAM50 subtype scores and risk of distant metastasis

As explained in the Materials & Methods part, PAM50 subtyping not just generates an overall risk of recurrence (ROR) score, but also provides scores of all molecular subtypes for each sample. Therefore, we analyzed whether the ROR score and subtype (luminal A, luminal B, HER2-E, basal) scores are associated the risk of developing metastasis. It turned out that, in our premenopausal patients, patients who developed distant metastasis had higher ROR scores (p = 0.01),

similar to the established trend in postmenopausal patients. Patients who developed distant metastasis had lower Luminal A scores ($p = 0.04$) and higher HER2-E scores ($p = 0.006$). Moreover, luminal B specimens had significantly higher HER2-E score than luminal A ones ($p < 0.001$) (Figure 3-2A) [96]. Subtype scores of luminal B and basal were not significantly different between M1 and M0 patients. Besides, higher ROR and HER2 scores and lower luminal A score were associated with shorter survival ($p = 0.002$, $p < 0.001$, $p = 0.037$) (Figure 3-2B). We also evaluated whether HER2-E score or luminal A score are correlated with ROR score. Of note, ROR scores had a negative correlation with luminal A scores ($p < 0.001$) and a positive correlated with HER2-E score ($p < 0.001$) (Figure 3-2C) [96].

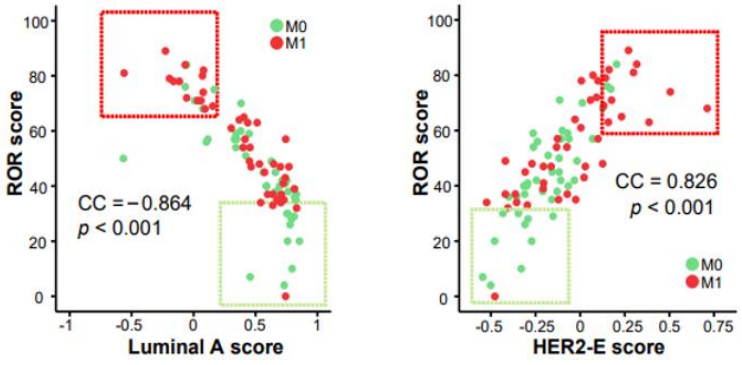


A

Univariate survival analysis of subtype scores

			95%CI	
Factors	***p*-value**	**HR**	**lower**	**upper**
ROR score	0.002	1.03	1.01	1.05
HER2-E score	<0.001	22.0	4.9	97.8
Luminal A score	0.037	0.43	0.2	0.95
Luminal B score	0.14	2.6	0.7	9.3
Basal score	0.53	1.4	0.5	3.7

B



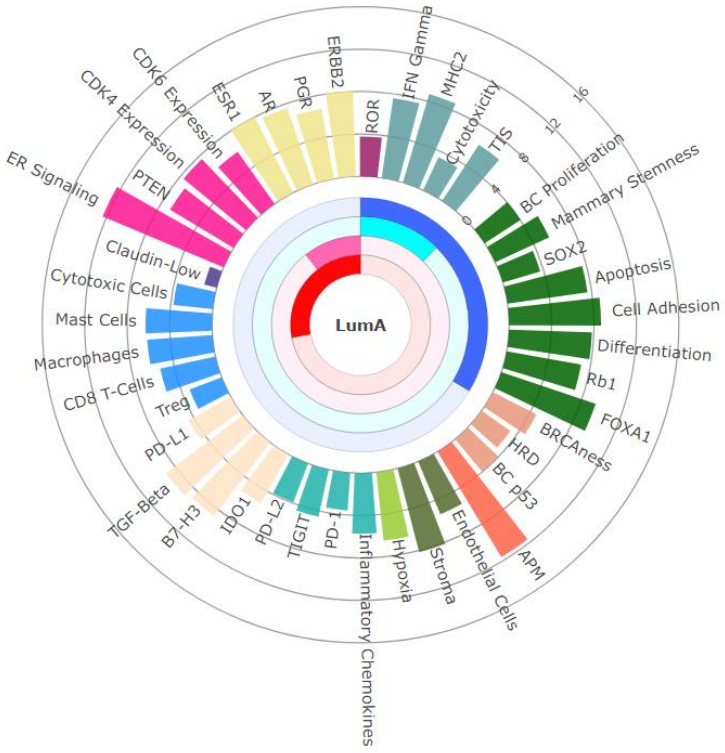
C

Figure 3-2 Differential analysis of the subtype scores [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

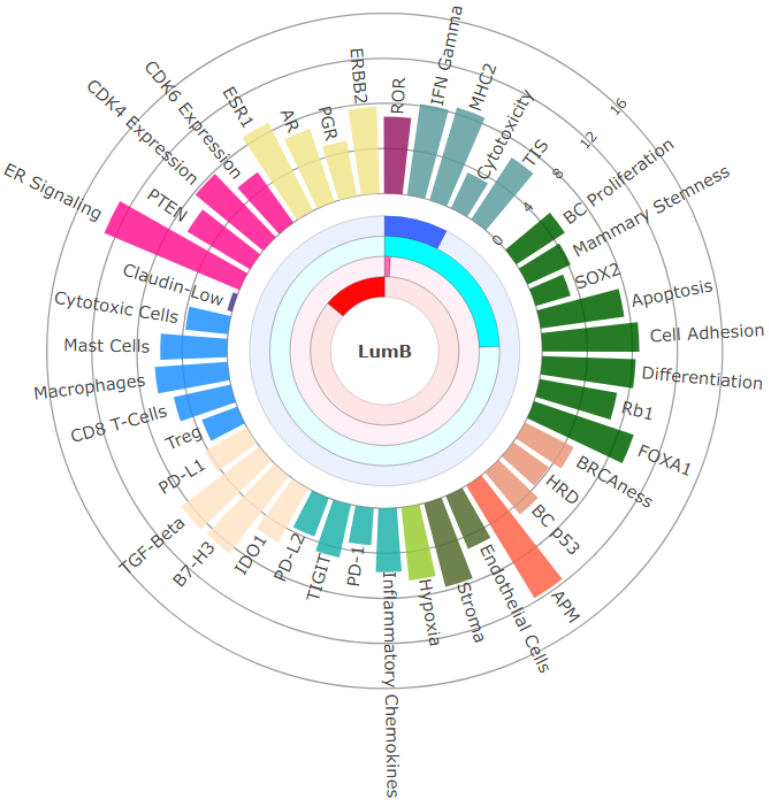
3.3 BC360® signature analysis

3.3.1 Signature score overview for each subtype

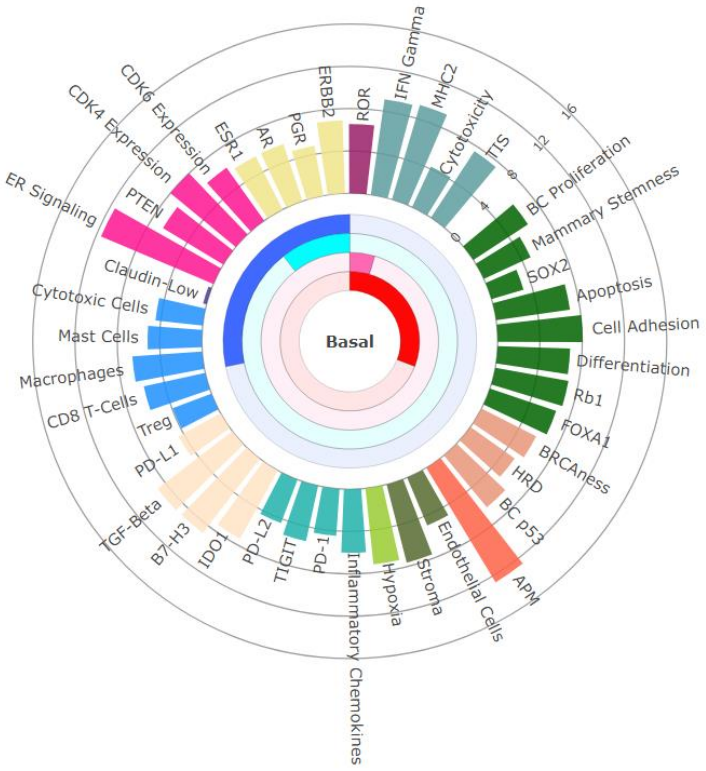
Based on the signature expression scores of samples, the overall signature score was calculated for each subtype. Besides, the correlation between subtypes was calculated (Figure 3-3). According to the calculation, luminal A subtype and luminal B subtype are closest to each other and most irrelevant to basal subtype. The reading of the wheel plots follows the following principle: the BC360® signatures scores are in the outer cycle of the wheel, and the scores are between 0 to 16; the overall subtype scores are in the inner cycle (the length of the line represents the subtype score, and the direction of the line represents whether the correlation is positive or negative, where a clockwise line represents a positive correlation and an anti-clockwise line represents a negative correlation). Of note, the wheel plots only showed the overview of each subtype, and no statistical comparison was performed.



A


LumB
ERBB2
ROR
IFN Gamma
MHC2
Cytotoxicity
TIS
BC Proliferation
Mammary Stemness
SOX2
Apoptosis
Cell Adhesion
Differentiation
Rb1
FOXA1
BRCAness
HRD
BC p53
APM
Endothelial Cells
Stroma
Hypoxia
Inflammatory Chemokines
TIGIT
PD-1
PD-L2
IDO1
B7-H3
TGF-Beta
PD-L1
Treg
CD8 T-Cells
Macrophages
Mast Cells
Cytotoxic Cells
Claudin-Low
PTEN
ER Signaling
CDK4 Expression
CDK6 Expression
ESR1
AR
PGR
0
4
8
12
16

B


Basal
ERBB2
ROR
IFN Gamma
MHC2
Cytotoxicity
TIS
BC Proliferation
Mammary Stemness
SOX2
Apoptosis
Cell Adhesion
Differentiation
Rb1
FOXA1
BRCAness
HRD
BC p53
APM
Endothelial Cells
Stroma
Hypoxia
Inflammatory Chemokines
TIGIT
PD-1
PD-L2
IDO1
B7-H3
TGF-Beta
PD-L1
Treg
CD8 T-Cells
Macrophages
Mast Cells
Cytotoxic Cells
Claudin-Low
PTEN
ER Signaling
CDK4 Expression
CDK6 Expression
ESR1
AR
PGR
0
4
8
12
16

C

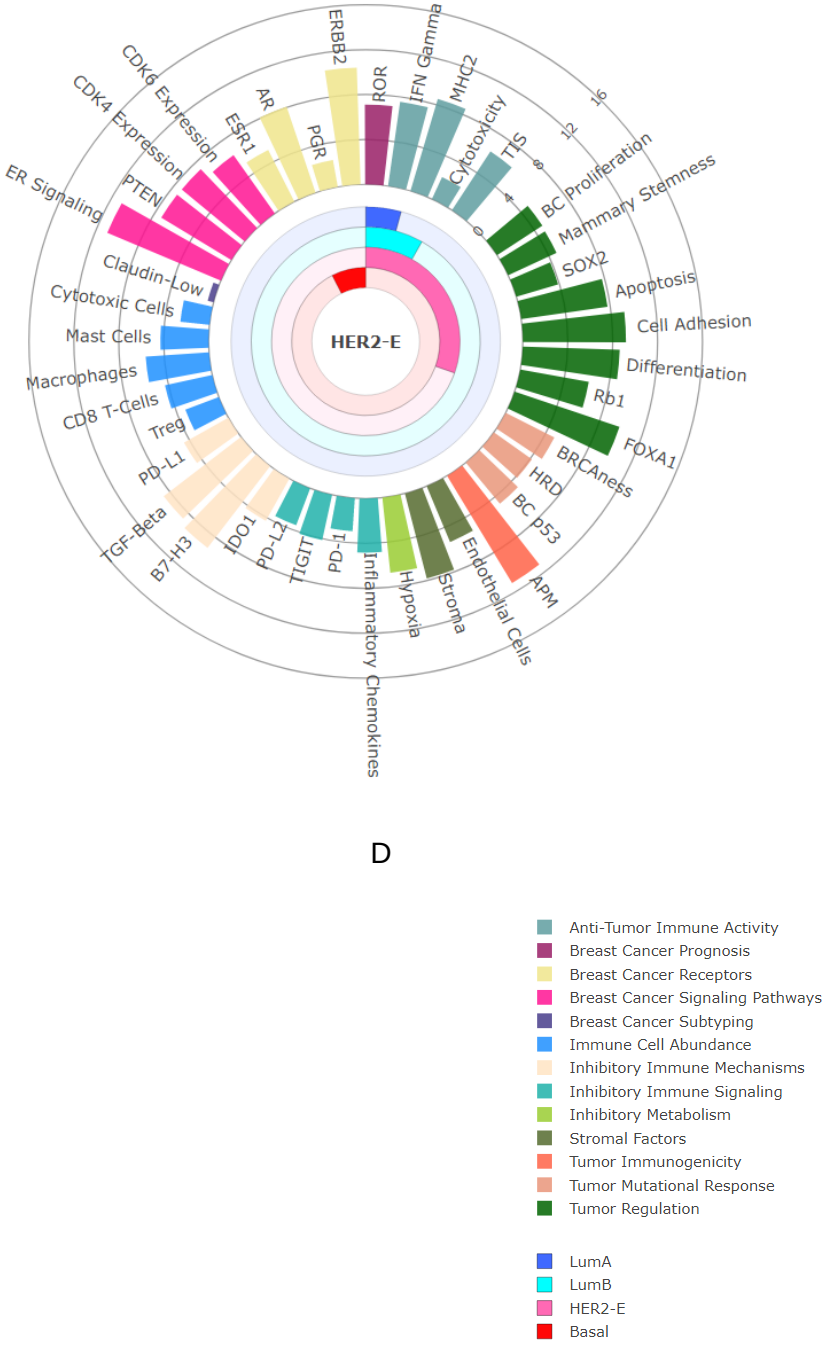


Figure 3-3 The overview of signature analysis for each subtype

3.3.2 BC360® signature expression

In order to investigate metastasis-related signatures, we compared the expression of BC360® signatures between the two group of patients. Four significantly different signatures include: patients who developed metastasis had higher expression of ROR (p = 0.006), and lower expression of claudin-low (p = 0.04), mammary stemness (p = 0.02), and PGR (progesterone receptor) (p = 0.02) (Figure 3-4) [96]. The other signatures that were not statistically significant include: ESR1, differentiation, FOXA1, AR, ERBB2, BC proliferation, APM, HRD, BC P53, macrophages, IFN-γ, CDK expression, hypoxia, cell adhesion, IDO1, PD-L2, MHC2, Treg, BRCAness, SOX2, TGF-β,

TIS, apoptosis, inflammatory chemokines, cytotoxicity, Rb1, mast cells, B7-H3, TIGIT, ER signaling, cytotoxic cells, endothelial cells, PTEN, CD8 T cells, PD-1, stroma, PD-L1, CDK6 expression (signatures were listed according to the fold change difference between M1 and M0).

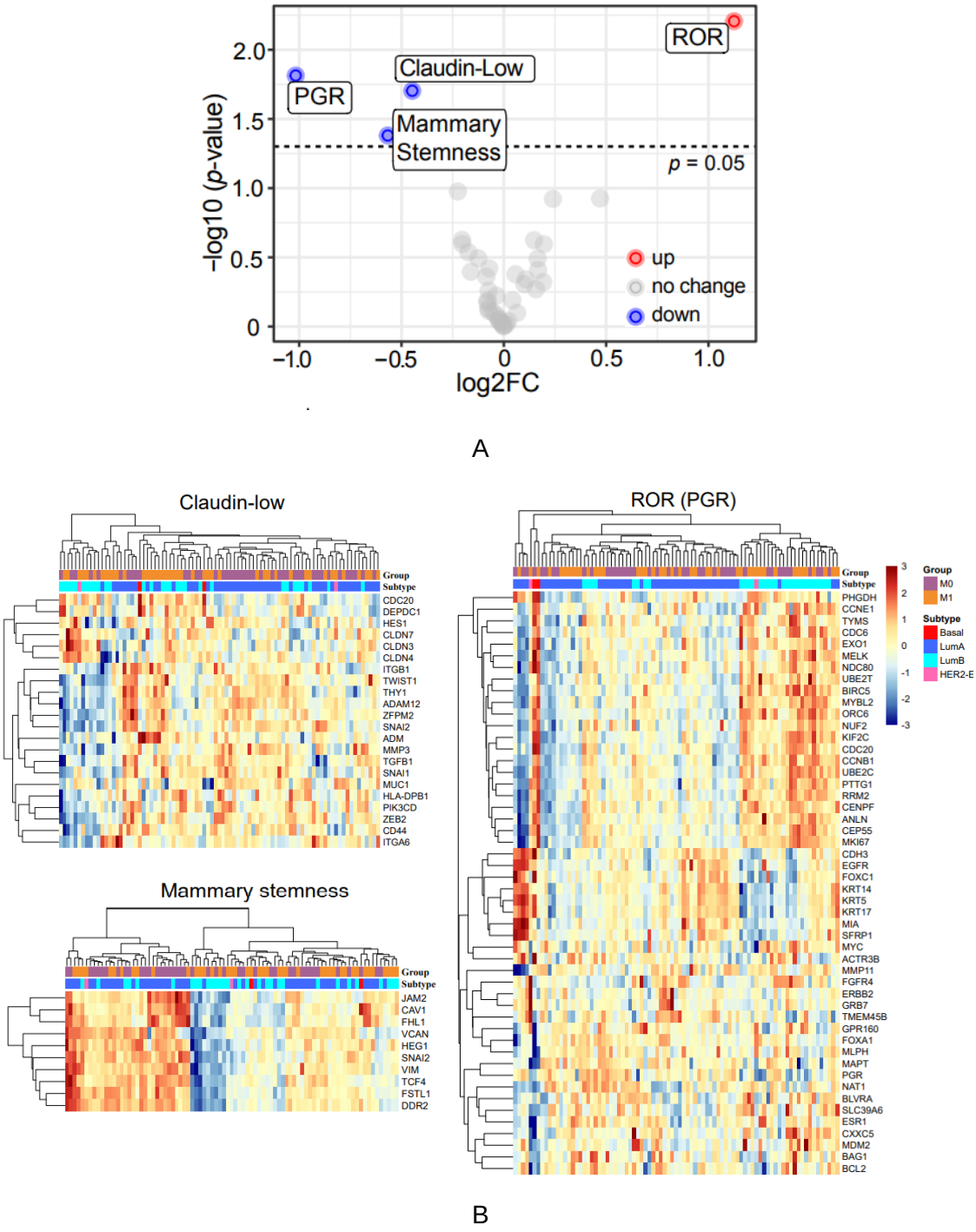


Figure 3-4 The differentially expressed BC360® signatures [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

3.3.3 The four significant signatures were also associated with survival

Through deferential analysis, we figured out four BC360® signatures that had different expression between M0 and M1 patients. To confirm the importance of these four signatures, we then carried out univariate survival analysis. It turned out, the signatures that were distant metastasis-related

were also survival-related. Specifically, shorter survival was observed in patients with higher expression of ROR (p = 0.002), and lower expression of claudin-low (p = 0.04), mammary stemness (p = 0.04), and PGR (p = 0.02) (Figure 3-5) [96].

Signatures	Differential expression		Univariate survival analysis 95%CI			
	Log2FC	*p*-value	HR	lower	upper	*p*-value
ROR	1.125	0.006	1.027	1.01	1.045	0.002
PGR	–1.017	0.02	0.847	0.736	0.976	0.02
Claudin-low	–0.448	0.02	0.667	0.451	0.985	0.04
Mammary stemness	–0.567	0.04	0.795	0.639	0.99	0.04

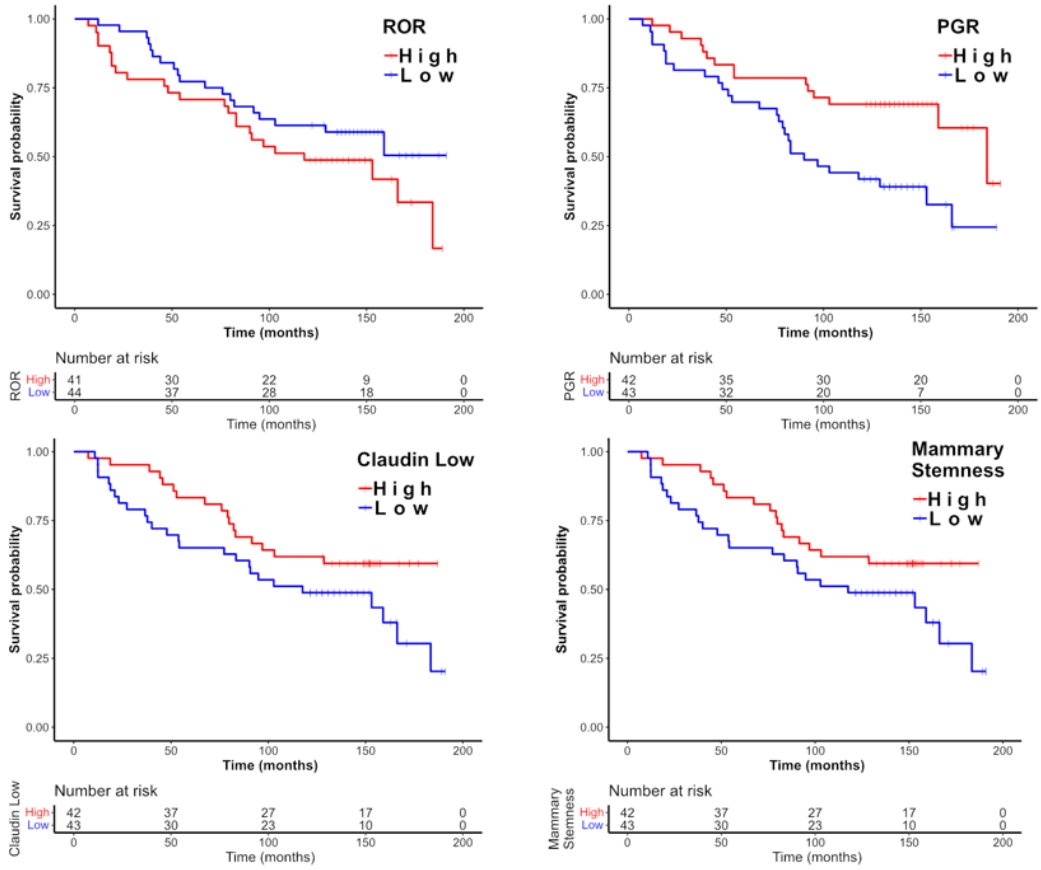
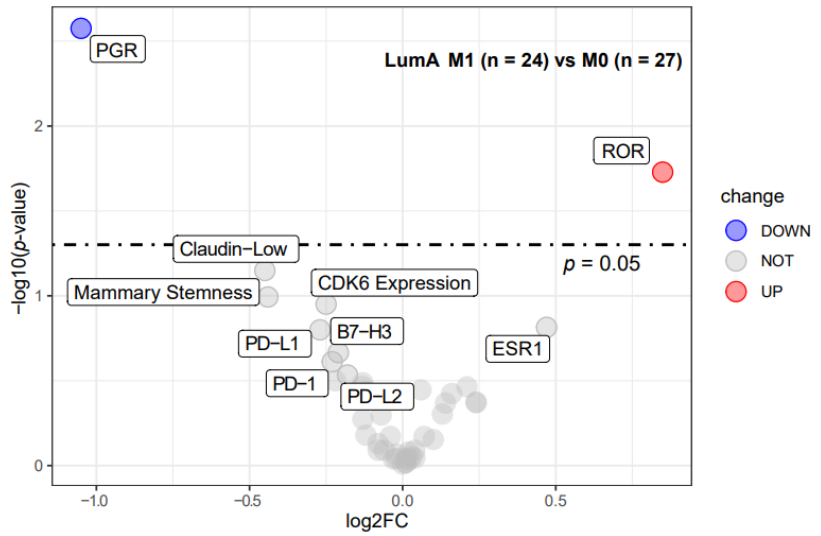


Figure 3-5 Survival analysis of four significant signatures [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

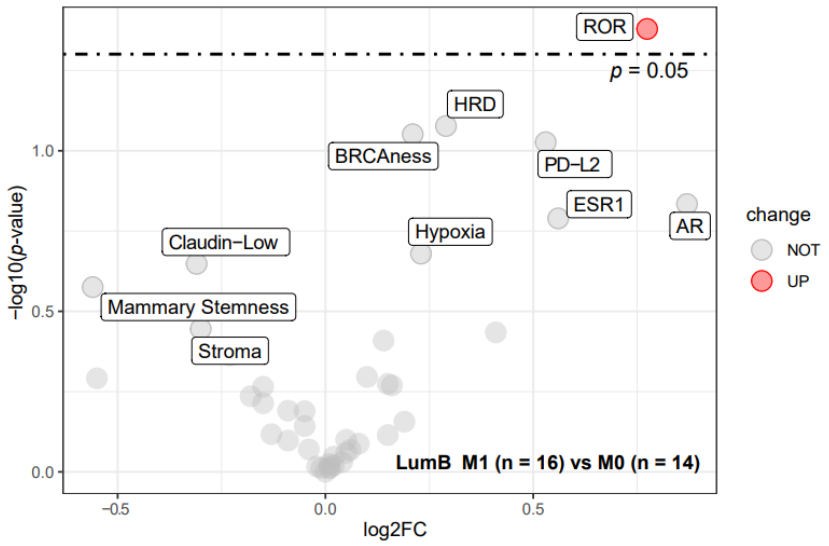
3.3.4 Subgroup analysis of BC360® signatures

We were curious if the signatures are consistently associated with distant metastasis and survival in luminal A and luminal B patients and carried out subgroup analyses to investigate the fact. The results showed that, both in luminal A and in luminal B subgroups, patients that developed distant metastasis had higher expression of ROR (p = 0.02, p = 0.04). Besides, only in the luminal A subgroup, patients that developed distant metastasis had lower expression of PGR than patients who did not (p = 0.002) (Figure 3-6A,3-6B) [96]. Univariate survival analysis of four significant signatures was also performed in patient subgroups. The results showed that ROR was significantly associated with survival in both luminal A and luminal B subgroups of patients (p = 0.017, p = 0.042); PGR was associated with survival in luminal A patients (p = 0.002) (Figure 3-6C) [96].

Claudin-low and mammary stemness were neither metastasis-related nor survival-related in subgroups.



A



B

Univariate survival analysis of four significant signatures in subgroups

	LumA M1 vs LumA M0				LumB M1 vs LumB M0			
Signatures	*p*-value	HR	95%CI lower	95%CI upper	*p*-value	HR	95%CI lower	95%CI upper
ROR	0.017	1.048	1.008	1.089	0.042	1.054	1.002	1.108
PGR	0.002	0.651	0.497	0.852	0.822	0.975	0.781	1.217
Claudin-low	0.216	0.769	0.507	1.166	0.242	0.552	0.203	1.496
Mammary Stemness	0.198	0.731	0.454	1.177	0.23	0.802	0.56	1.149

C

Figure 3-6 Subgroup analysis of BC360® signatures [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

3.4 GSEA analysis

GSEA was carried out to test for associations between tested 751 genes (7 tested genes were not included in the platform) and defined gene sets. Hallmark gene sets are the core gene sets of GSEA analysis and were tested first. In our patients, we found that gene expression of M1 patients has enrichment in the HALLMARK_MTORC1_SIGNALING (FDR q = 0.241). Subgroup analyses showed that the enrichment was more obvious in luminal B patients (FDR q = 0.047) and not significant in luminal A M1 patients (FDR q = 0.5421) (Figure 3-7) [96]. Besides Hall mark gene sets, C1–C7 gene sets (the introduction of the gene sets was covered in the method part) were also analyzed. The significant enrichments were summarized in Table 3-4. Briefly, gene expression of luminal A M1 patients has enrichment in C3 (regulatory target gene sets) and C6 (oncogenic signature gene sets); gene expression of luminal B M1 patients has enrichment in C2 (curated gene sets); and gene expression of luminal B M0 patients has enrichment in C3 (regulatory target gene sets) and C7 (immunologic signature gene sets). Limited by that the GSEA analyses had neither quantitative nor qualitative report for each patient, no survival analysis was carried out.

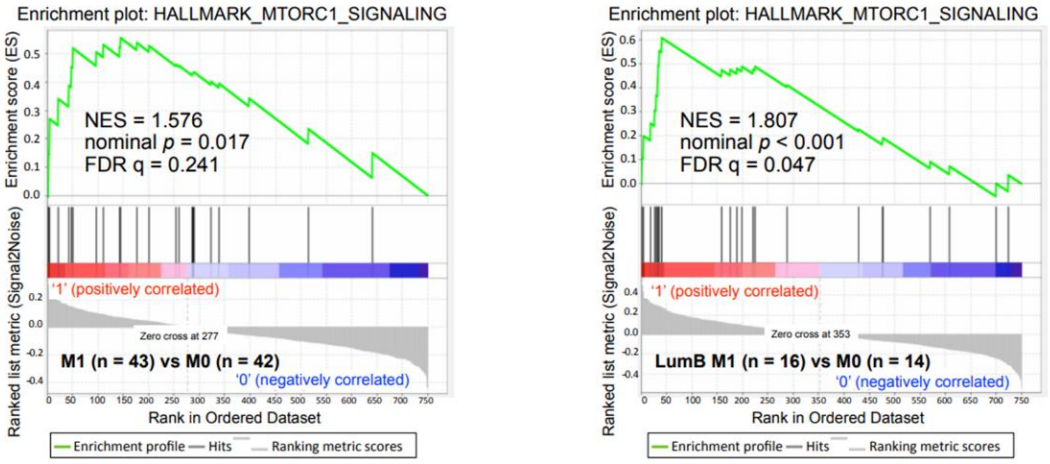


Figure 3-7 GSEA analysis of hallmark gene sets in all patients/subgroups [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

Table 3-4 Overview of significant gene set enrichments

Patients	In	Sets	Gene Set	NES	Nominal *p*	FDR q
All	M1	H	Hallmark_MTORC1_Signaling	1.58	0.017	0.241
	M0	C6	BMI1_DN_MEL18_DN.V1_DN	–1.66	0.008	0.247
Luminal A	M1	C3	MIR9983_3P	1.88	<0.001	0.233
		C6	IL15_UP.V1_UP	1.82	<0.001	0.151
			E2F1_UP.V1_DN	1.74	0.002	0.193

Table 3-4 (continued)

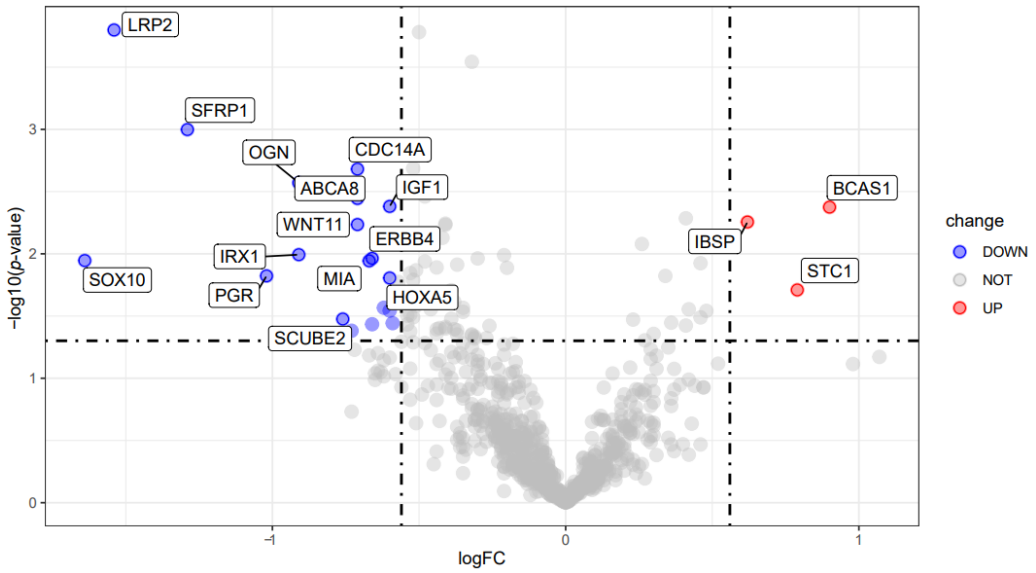
Luminal B	M1	H	Hall-mark_MTORC1_Sig-naling	1.81	<0.001	0.047
		C2	LY_AGING_PREMA-TURE_DN	1.88	<0.001	0.117
	M0	C3	MIR4755_5P	–2	<0.001	0.079
			MIR5006_3P	–2	<0.001	0.040
		C7	GSE25677_MPL_VS_R848_STIM_BCELL_DN	–1.97	0.002	0.165
			GSE35685_CD34POS_CD38NEG_VS_CD10POS_BONE_MAR-ROW_UP	–1.95	0.002	0.143
			GSE6259_33D1_POS_VS_DEC205_POS_SPLENIC_DC_DN	–1.91	<0.001	0.165
			GSE39820_TGF-BETA3_IL6_VS_TGF-BETA3_IL6_IL23A_TREATED_CD4_TCELL_UP	–1.89	<0.001	0.182
			GSE26488_WT_VS_VP16_TRANS-GENIC_HDAC7_KO_DOUBLE_POSI-TIVE_THYMO-CYTE_UP	–1.84	0.002	0.228

3.5 Analysis of differentially expressed genes

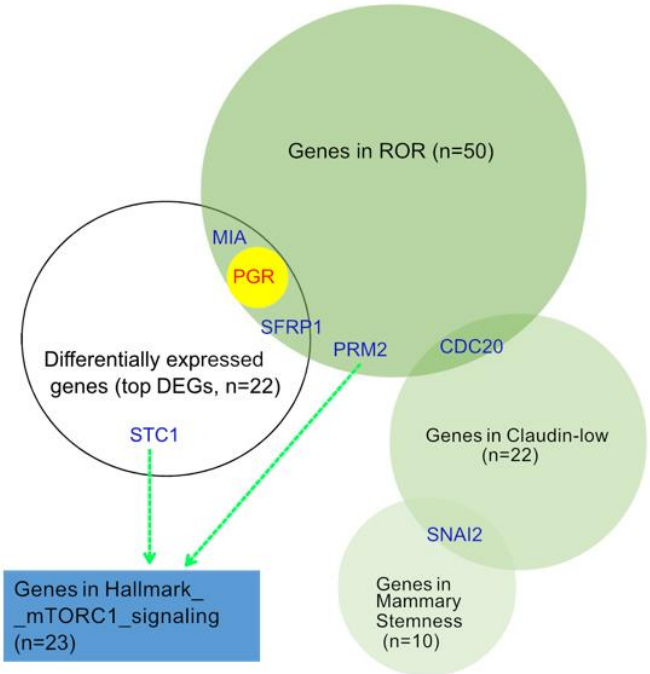
3.5.1 Differential expression of single genes

Except for established signatures which consist of several or dozens of genes, analysis of single gene was also performed to investigate the single prognostic genes for HR+/HER2− premenopausal EBC, as significant single genes might help to develop novel and more specific signatures for this special group of patients. As suggested in the method part, the expression of 758 genes were normalized and compared through the limma-t test, and 22 genes passed the cutoff (log2 FC > 0.586 and $p < 0.05$) for differentially expressed genes (Figure 3-8A). The overlap between

the differentially expressed genes and the genes included by the significant signatures was summarized. As could be seen in Figure 3-8B, not much overlap was observed. Three of the DEGs, namely PGR, MIA and SFRP1, were included by ROR signature, and one DEG, namely STC1, was involved in the mTORC1 signaling pathway [96].



A



B

Figure 3-8 The differentially expressed genes between two groups of patients [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

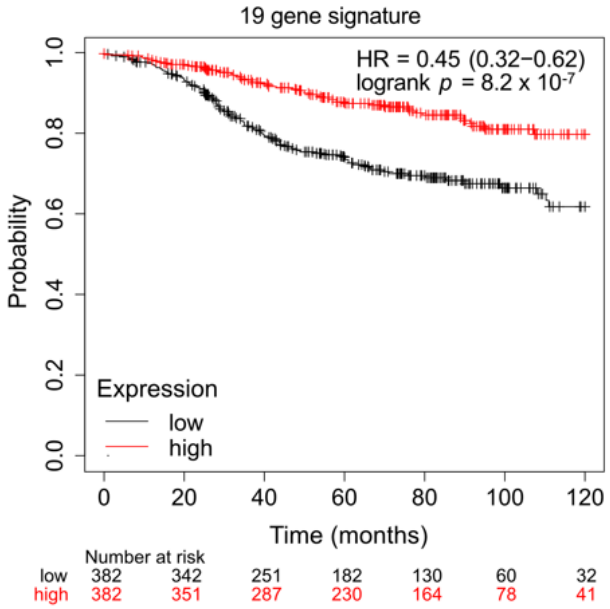
3.5.2 Survival relevancy of the DEGs

Univariate survival analysis was carried out for each DEG to confirm their significancy. As the results suggested, among the 22 DEGS, 19 were survival-related (Figure 3-9A). In order to further validate our findings, we used Kaplan-Meier Plotter tool to test the survival relevancy of the 19 genes as a whole signature and as single genes in patients from online databases. Excitingly, the nineteen-gene signature was survival-related in the larger cohort from other databases (Figure 3-9B), and 15 out of 19 genes were survival-related as single genes (Figure 3-9C) [96].

Significant univariate survival analysis of DEGs

Gene	Gene description	Univariate survival analysis			
				95%CI	
		p-value	HR	lower	upper
LRP2	Low-density lipoprotein receptor-related protein 2	<0.001	0.693	0.583	0.823
SFRP1	Secreted frizzled-related protein 1	0.002	0.749	0.625	0.899
CDC14A	Cell division cycle 14A	0.005	0.623	0.448	0.867
ABCA8	ATP binding cassette subfamily A member 8	0.005	0.697	0.54	0.899
IBSP	Integrin binding sialoprotein	0.006	1.495	1.122	1.992
OGN	Osteoglycin	0.008	0.757	0.616	0.93
BCAS1	Breast carcinoma amplified sequence 1	0.009	1.33	1.074	1.648
IGF1	Insulin-like growth factor 1	0.01	0.682	0.503	0.925
WNT11	Wnt family member 11	0.01	0.705	0.537	0.928
IRX1	Iroquois homeobox 1	0.01	0.767	0.625	0.94
MIA	Melanoma inhibitory activity	0.01	0.676	0.495	0.923
PGR	Progesterone receptor	0.02	0.847	0.736	0.976
SOX10	SRY-Box transcription factor 10	0.02	0.882	0.793	0.982
PTGER3	Prostaglandin E receptor 3	0.02	0.778	0.625	0.967
ERBB4	Erb-B2 receptor tyrosine kinase 4	0.03	0.803	0.661	0.975
SCUBE2	Signal peptide, CUB domain, and EGF Like domain containing 2	0.04	0.844	0.716	0.994
HOXA5	Homeobox A5	0.04	0.74	0.56	0.979
ZBTB16	Zinc finger and BTB domain containing 16	0.04	0.822	0.68	0.993
THBS4	Thrombospondin 4	0.048	0.791	0.627	0.998

A



B

Validation in online database for single genes

Gene	Affymetrix ID	NoP	HR (95%CI)	logrank *p*
LRP2	205710_at	2301	0.65 (0.54–0.77)	4.6×10^{-7}
SCUBE2	219197_s_at	2301	0.66 (0.56–0.79)	2.3×10^{-6}
SFRP1	202035_s_at	2301	0.67 (0.57–0.8)	5.6×10^{-6}
ZBTB16	205883_at	2301	0.68 (0.57–0.8)	6.9×10^{-6}
MIA	206560_s_at	2301	0.68 (0.58–0.81)	1.3×10^{-5}
OGN	218730_s_at	2301	0.69 (0.58–0.81)	1.4×10^{-5}
ABCA8	204719_at	2301	0.69 (0.58–0.82)	2×10^{-5}
PGR	208305_at	2301	0.71 (0.6–0.84)	6.7×10^{-5}
IGF1	209540_at	2301	0.71 (0.6–0.84)	8.3×10^{-5}
SOX10	209842_at	2301	0.71 (0.6–0.84)	8.6×10^{-5}
ERBB4	206794_at	2301	0.79 (0.66–0.93)	0.0054
IRX1	230472_at	764	0.67 (0.49–0.91)	0.011
PTGER3	208169_s_at	2301	0.81 (0.68–0.96)	0.014
HOXA5	213844_at	2301	0.82 (0.69–0.97)	0.02
WNT11	206737_at	2301	0.83 (0.7–0.98)	0.032

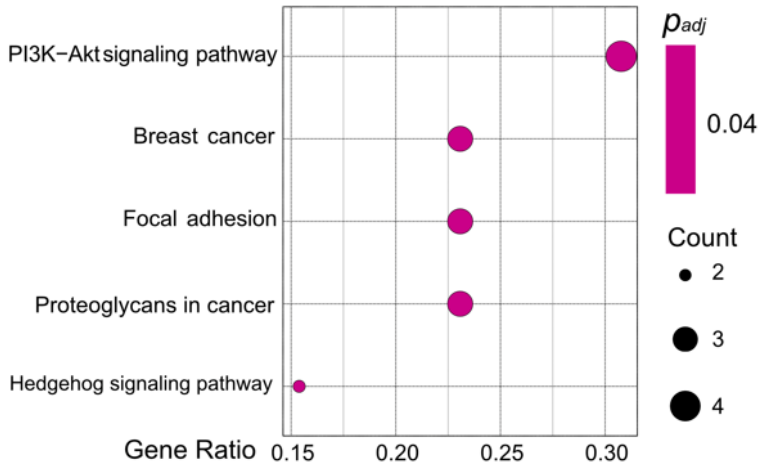
*NoP = Number of patients

C

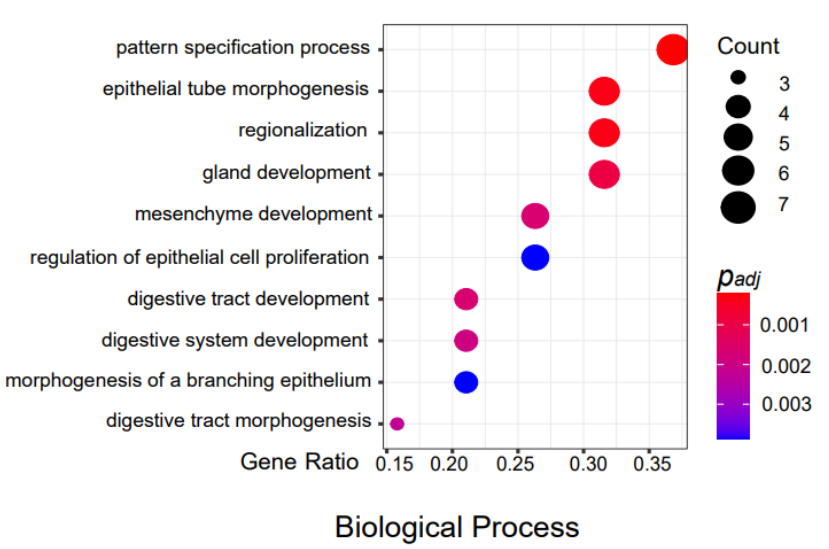
Figure 3-9 Survival analysis of the differentially expressed genes [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

3.5.3 Functional analysis of the 19 survival-related DEGs

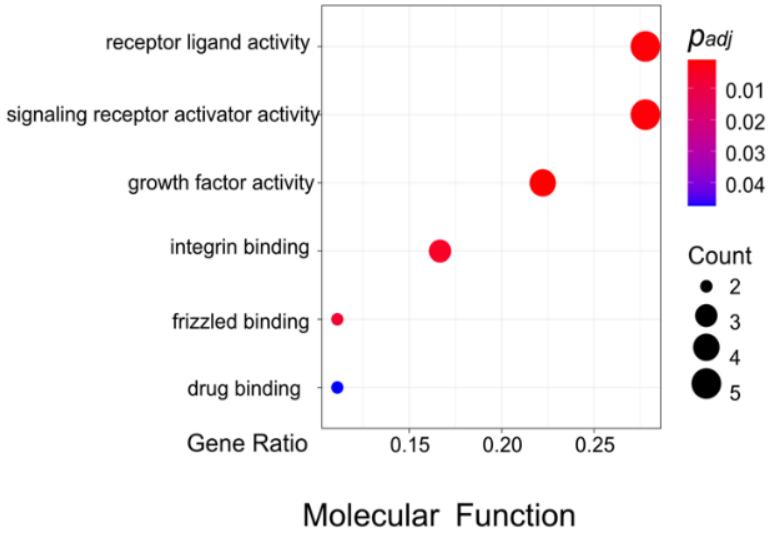
In order to understand the functions of the survival-related DEGs, we carried out GO, KEGG and STRING analyses. KEGG analysis suggested that the 19 DEGs were significantly associated with PI3K-Akt (phosphatidylinositol 3-kinase/protein kinase B) signaling, breast cancer, focal adhesion, proteoglycans in cancer, and hedgehog signaling pathway (Figure 3-10A). GO analysis suggested that several biological processes and molecular functions were involved (Figure 3-10B，Figure 3-10C). Functional interactions between the 19 DEGs were determined by STRING analysis, and several functional interactions were noticed (Figure 3-10D). The involved BC360 pathways of the 19 DEGs were summarized in Table 3-5 and Table 3-6.



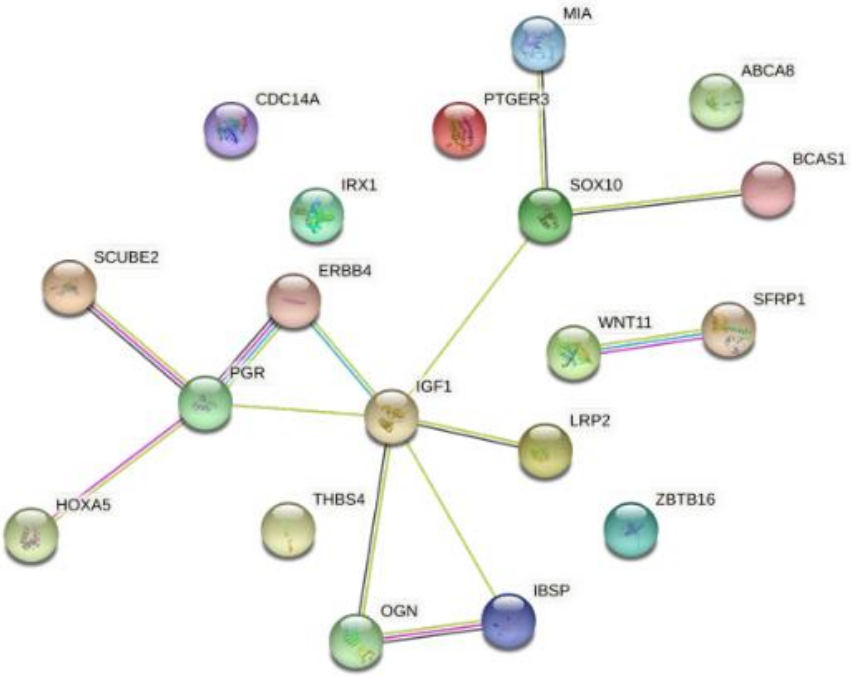
A



B



C



D

Figure 3-10 Functional analysis of nineteen survival-related DEGs [96] (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

Table 3-5 The involved BC360® pathways of the19 survival-related DEGs

Gene	Expression (M1 vs M0)	BC360® pathway
LRP2	DOWN	Hedgehog

Table 3-5 (continued)

SFRP1	DOWN	EMT, Subtypes, Triple negative Biology, Wnt
CDC14A	DOWN	Proliferation
OGN	DOWN	Proliferation, Triple negative biology
ABCA8	DOWN	Triple negative biology
IGF1	DOWN	EMT, MAPK, PI3K, Proliferation, Transcriptional misregulation, Triple negative biology
BCAS1	UP	EMT
IBSP	UP	Adhesion and migration, PI3K, Proliferation
WNT11	DOWN	Hedgehog, Wnt
IRX1	DOWN	Triple negative biology
ERBB4	DOWN	Triple negative biology
SOX10	DOWN	Triple negative biology
MIA	DOWN	Subtypes, Triple negative biology
PGR	DOWN	ER signaling, Subtypes
HOXA5	DOWN	Transcriptional misregulation
THBS4	DOWN	Adhesion and migration, PI3K, Proliferation, TGF-beta
PTGER3	DOWN	ER signaling
SCUBE2	DOWN	ER signaling, Triple negative biology
ZBTB16	DOWN	Transcriptional misregulation

Table 3-6 The distribution of 19 DEGs in BC360® pathways

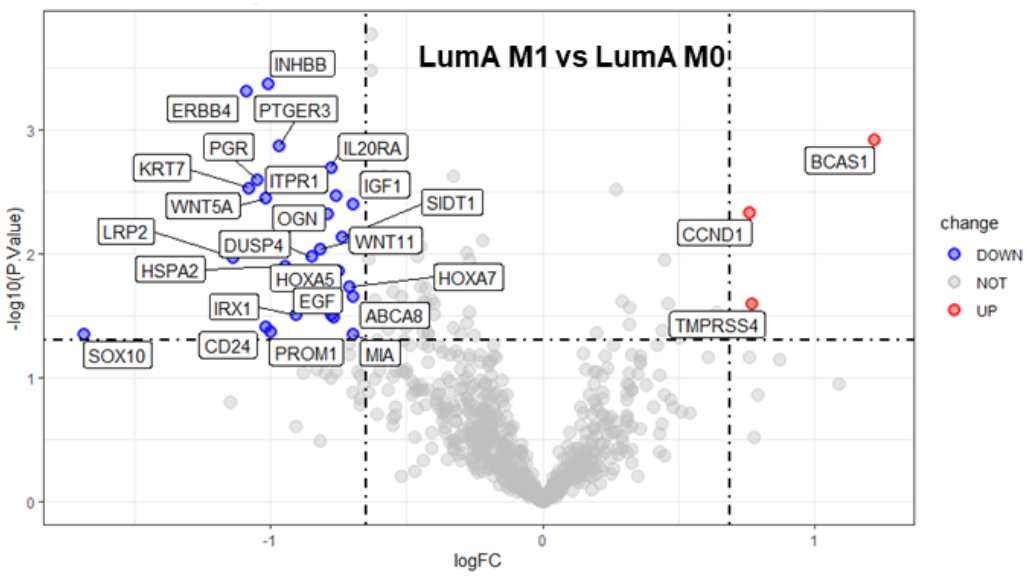
BC360® Pathway (n=24)	**All Genes**	**DEGs**	**Percentage (DEGs/All Genes)**
Triple Negative Biology	50	9	18.0%
ER Signaling	27	3	11.1%
Hedgehog	20	2	10.0%
Transcriptional Misregulation	63	3	4.8%
Subtypes	70	3	4.3%
Wnt	51	2	3.9%

Table 3-6 (continued)

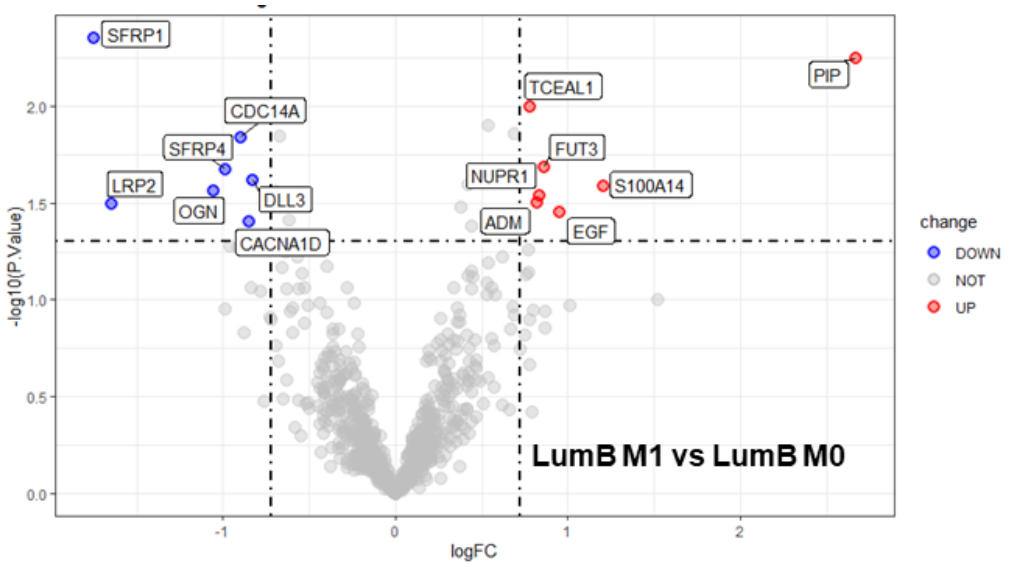
Proliferation	144	5	3.5%
PI3K	96	3	3.1%
Adhesion and Migration	83	2	2.4%
TGF-beta	57	1	1.8%
MAPK	100	1	1.0%
Angiogenesis	34	0	0
Antigen Presentation	21	0	0
Apoptosis	9	0	0
Cytokine and Chemokine Signaling	50	0	0
DNA Damage Repair	143	0	0
Epigenetic Regulation	18	0	0
Immune Infiltration	34	0	0
Internal Reference Gene	18	0	0
JAK-STAT	47	0	0
Notch	22	0	0
Stromal Markers	6	0	0
Tumor Metabolism	15	0	0
SUM	1263	37	2.9%

3.5.4 Differentially expressed genes in luminal A and luminal B subgroups

As was noticed in the signature analysis, the underlying biology of distant metastasis was inconsistent between luminal A and luminal B patients. Therefore, we carried out subgroup analyses where applicable, and the differentially expressed genes were also separately compared in luminal A and luminal B patients. As could be seen in Figure 3-11, twenty-eight DEGs were screened out in luminal A patients, and 14 DEGs were screened out in luminal B patients. DEGs in Luminal A patients have more similarity to the overall DEGs, and DEGs in luminal B have much less similarity. Only two DEGs were consistent in overall analysis and two subgroup analyses: LRP2 and OGN. The DEGs have prominent difference between luminal A and luminal B patients, rendering separate analyses in two subgroups necessary.



A



B

Figure 3-11 The differentially expressed genes in subgroups

3.6 Multivariate survival analysis

The survival relevance of the signatures and single genes were determined. Considering that pN and pT also have a significant impact on survival, we wanted to investigate all factors in one run. Therefore, a multivariate survival analysis including pN and pT, ROR score, HER2-E score, luminal A score, four significant signatures (ROR, PGR, claudin-low, mammary stemness), and 19 DMFS-related DEGs was performed. “Forward Stepwise (likelihood ratio)” was used to find the most representative factors (Table 3-7). The results suggested that besides pN (p = 0.016), the single genes LRP2 (p < 0.001) and PTGER3 (p = 0.017) were most representative prognostic factors (Table 3-8). In order to evaluate whether expression of LRP2 was correlated with other prognostic factors, we performed Pearson’s correlation analysis. It turned out that LRP2 was

correlated with most of the prognostic factors (18/25) (Table 3-9), especially HER2-E score, ROR score, SFRP1, CDC14A and ABCA8 ($p < 0.001$).

Table 3-7 Prognostic factors included in the multivariate analysis (the starting block)

Factor	Score	*p*-value	Factor	Score	*p*-value
pT (>1 vs 1)	4.7	0.030	BCAS1	6.9	0.009
pN (>0 vs 0)	12.4	<0.001	IBSP	7.6	0.006
ROR score	10.0	0.002	WNT11	6.0	0.015
HER2-E score	17.1	<0.001	IRX1	6.5	0.011
Luminal A score	4.4	0.035	ERBB4	5.0	0.025
Claudin-low	4.3	0.039	SOX10	5.5	0.020
Mammary stemness	4.2	0.040	MIA	6.1	0.013
PGR	5.4	0.020	PGR	5.4	0.020
LRP2	18.1	<0.001	HOXA5	4.4	0.035
SFRP1	9.6	0.002	THBS4	3.9	0.047
CDC14A	8.0	0.005	PTGER3	5.2	0.023
OGN	7.2	0.007	SCUBE2	4.1	0.042
ABCA8	7.8	0.005	ZBTB16	4.1	0.042
IGF1	6.2	0.013			

Table 3-8 Multivariate survival analysis of 28 prognostic factors (method: forward stepwise)

				95%CI	
Step	**Factors**	***p*-value**	**HR**	**lower**	**upper**
Step1	LRP2	<0.001	0.69	0.58	0.82
Step 2	LRP2	<0.001	0.67	0.56	0.80
	PTGER3	0.004	0.69	0.53	0.89
Step 3	LRP2	<0.001	0.71	0.60	0.85
	PTGER3	0.017	0.74	0.57	0.95
	pN (>0 vs 0)	0.016	2.4	1.2	5.1

Table 3-9 Correlation analysis of LRP2 and other prognostic factors

Factor	CC	*p*-value	Factor	CC	*p*-value
pN (>0 vs 0)	−0.17	0.106	BCAS1	0.08	0.498
pT (>1 vs 1)	−0.20	0.062	IBSP	−0.08	0.464
HER2-E score	−0.49**	<0.001	WNT11	0.27*	0.014
Luminal A score	0.32**	0.003	IRX1	0.30**	0.005
ROR score	−0.46**	<0.001	ERBB4	0.28**	0.01
PGR	0.32**	0.003	SOX10	0.36**	0.001
Mammary stem-ness	0.20	0.070	MIA	0.36**	0.001
Claudin-low	0.20	0.063	PGR	0.32**	0.003
SFRP1	0.45**	<0.001	HOXA5	0.24*	0.027
CDC14A	0.41**	<0.001	THBS4	0.34**	0.001
OGN	0.32**	0.003	PTGER3	0.03	0.795
ABCA8	0.41**	<0.001	SCUBE2	0.03	0.759
IGF1	0.35**	0.001	ZBTB16	0.29**	0.006

CC = correlation coefficient; * $p < 0.05$, ** $p < 0.01$. Negative values indicate negative correlations, and positive values indicate positive correlations.

Of note, for multivariate analysis, when different factors were included in the starting block, the final output may vary. We also tried including pN, pT, four differentially expressed signatures, 22 DEGs into the multivariate analysis without considering luminal A scores and HER2-E scores. And in this approach, LRP2 was also calculated as an independent prognostic factor, along with pN, IBSP, SCUBE2 [96] (Table 3-10). Therefore, the importance of LRP2 was confirmed, and the complexity of the interactions between the prognostic factors should be noticed and carefully interpreted.

Table 3-10 Multivariate survival analysis of pN, pT, and differentially expressed signatures/genes (method: forward stepwise) (Reproduced with permission from Ni et al. *J Pers Med*, 2021)

				95%CI	
Step	**Factors**	***p*-value**	**HR**	**lower**	**upper**
Step 3	pN	0.007	2.75	1.33	5.68
	LRP2	<0.001	0.73	0.62	0.86
	IBSP	0.03	1.44	1.05	1.97
	SCUBE2	0.04	0.82	0.68	0.98

4. Discussion

4.1 Clinical risk factors for developing distant metastasis in HR+/HER2− early breast cancer

Early breast cancer is curable, yet distant metastasis may develop and lead to inevitable death. Distant metastasis is the most common type of breast cancer recurrence [107], and it marks the time point at which the patient has entered the late stage and can no longer be cured [28]. Therefore, we selected distant metastasis as the endpoint of our follow-up, which is also common practice of clinical trials [29, 56, 108]. Besides, we set the minimum follow-up period as 10 years, because even though the peak for cancer relapse is 2-3 years, in particular in luminal EBC, relapses can occur much later as well and consequently, the common follow-up period is between 5 and 15 years [15, 107, 109]. Normally, the percentage of early breast cancer that develops distant metastasis is around 30% [10, 15]. In the final cohort of our case control study, we included 49 patients who had developed distant metastasis during follow-up and 48 who did not.

Though the exact drivers for distant metastases are not known, widely accepted prognostic factors for developing distant metastasis include tumor size, nodal status, histological grade, molecular subtype, and appropriate treatment [10, 11, 107, 110]. As shown by the patient characteristics in our study, tumor size, grade, tumor stage (based on tumor size and nodal status) are associated with risk of distant metastasis. It is generally accepted that nodal status and tumor size are interacting prognostic factors and nodal status is a more powerful prognostic factor than tumor size [111, 112], as a small tumor with extensive lymph node involvement holds a higher risk of metastasis than a larger tumor with less lymph node involvement [113]. Limited by the sample size, we did not divide patients into detailed subgroups, as this approach needs a much larger sample size to show the statistical significance. Instead, we generally divided the patients into the following subgroups: patients with positive lymph nodes and patients without; patients at stage I and patients at stage > I, to perform univariate survival analysis for investigating their prognostic value. According to the results, nodal status and tumor stage are both associated with survival, but nodal status has a much higher impact with higher significance (lower *p*-value and higher hazard ratio). Therefore, in our cohort, nodal status is the most important prognostic factor among the common clinical parameters.

HR+/HER2− breast cancer has a relatively favorable prognosis compared to triple-negative or HER2+ breast cancer, and the treatment routine is generally clear: A combination of local and systemic treatment [10, 28]. For local treatment, lumpectomy and mastectomy are two mainstays and should be selected based on the patient's clinical risk, contradictions, and personal preference. If lumpectomy was the choice, radiotherapy must be prescribed to limit local recurrence. If mastectomy was chosen, radiotherapy may be spared based on tumor extent [10, 11, 114]. In our cohort, 71.4% of patients who did not develop metastasis received lumpectomy plus radiotherapy, and the statistical results suggested that the surgery method did not influence metastasis development. Therefore, our small cohort confirmed that it is safe to conduct lumpectomy plus radiotherapy for patients with a low risk of recurrence. As for axillary lymph node management, axillary lymph node dissection (ALND) is only suggested for high-risk patients, and selective sentinel lymph node biopsy (SLNB) is the current standard of care for node-negative breast cancer. Even for positive SLNB, a low axillary disease burden could be sufficiently treated by radiotherapy [11, 114]. In our cohort, in the patient who did not develop distant metastasis, 51% of patients did

not have axillary lymph node dissection after SLNB. Moreover, among the 30 patients without further ALND, only 5 developed metastases. Considering that replacing ALND with SLNB (with or without radiotherapy) is associated with better post-surgery life quality [114], de-escalation of lymph node management should be advocated in patients with a low risk of recurrence.

For systemic treatment, the mainstay is endocrine therapy. Although established later than chemotherapy, endocrine drug therapy has taken over the leading role as it directly targets the underlying cause of HR+ breast cancer and is thus more efficient and less toxic [3, 115]. For premenopausal patients, tamoxifen is the standard of care [11, 114], although aromatase inhibitor plus gonadotropin-releasing hormone agonist (GnRHa) was reported to render more survival benefit [116]. In young patients at high risk of recurrence, pros and cons should be thoroughly considered before selecting aromatase inhibitors [114]. In our patient cohort, most patients (93%) were treated with tamoxifen, and the selection of endocrine therapy did not significantly influence the survival result. Yet, the small sample size limits the power of the conclusion. As for chemotherapy, considering that it does not bring substantial survival benefits for most patients with HR+ EBC, prudent avoidance of chemotherapy should be implemented [4, 114]. In our patient cohort, 39.6% of the patients without metastases did not receive chemotherapy. Of the patients with distant metastasis, most patients had received chemotherapy (89.1%). Thus, even though patients at higher clinical risk had been prescribed chemotherapy, they still experienced distant metastases. And for patients with lower risk of recurrence, omitting chemotherapy did not their harm survival chances. The most commonly used chemotherapy regimen contained anthracyclines and/or taxanes [114, 115].

A combination of anti-cancer medications with different mechanisms improves efficiency, loweres the possibility of resistance, and reduces toxicity [10, 114]. According to the basic statistical analysis of our cohort, the fact whether the chemotherapy contained anthracyclines or taxanes did not influence development of distant metastasis. Yet, as our study is retrospective and did not match patients regarding their (neo-)adjuvant chemotherapy, we cannot offer definitive conclusions regarding the best chemotherapy regime. Besides endocrine therapy and chemotherapy, targeted therapy and immunotherapy are potential systemic treatment choices that may help to further individualize the treatment in order to improve survival and reduce toxicity [4, 10]. But these additional therapies are not first choice for HR+/HER2− breast cancer, and therefore not evaluated by our study. Briefly, our patients were all treated by standard of care according to national guidelines, and no significant impact of any treatment was noticed.

Based on the analysis of clinical factors and treatment of our patient cohort, we concluded that nodal status and tumor stage are prognostic factors for survival of premenopausal patients, and that nodal status has the strongest prognostic power. By retrospective analysis of treatment modalities, we found that - at least in our cohort - escalated treatment (mastectomy, ALND, chemotherapy) may not be necessary for some patients and not effective enough for some patients to avoid distant metastasis.

Furthermore, analysis of the clinical risk factors revealed the necessity for researching into molecular risk factors. As seen in our cohort, 35% of pT1 tumors developed distant metastasis, and 28.8% of tumors with no lymph node involvement developed metastasis. For these patients, aggressive tumor biology may be the crucial driver for distant metastasis [10, 111].

4.2 Molecular subtype and risk of developing distant metastasis

Molecular subtypes based on the PAM50 test are well-established prognostic markers and predictors of treatment benefit. It is generally considered that luminal subtypes have a better prognosis than HER2-enriched or basal subtypes [10, 109]. The luminal subtype can be further divided into luminal A and luminal B subtypes, and the latter one is considered to be associated with a poorer prognosis [10, 21, 109]. Clinically, molecular subtyping is not routinely performed. Instead, surrogate subtyping based on the immunohistochemical expression of hormone receptors, HER2 and Ki-67 index is more common in clinical practice [10, 21, 65]. Our cohort included patients based on expression of HR and HER2, and an excellent coherence between surrogate subtyping and molecular subtyping was observed: 95.3% of the included HR+/HER2− tumors belong to luminal subtypes. Limited by the incomplete Ki-67 index data, a differentiation between luminal A-like and luminal B-like could not be done for the surrogating subtyping, which therefore could be compared with the molecular subtyping results.

In our patient cohort, the risk of developing distant metastasis was not significantly different between luminal A and luminal B tumors. Among the 51 patients with luminal A subtype, 47% had developed metastasis, and among the 30 patients of luminal B subtype, 53.3% had developed metastasis. Although the actual metastasis rate in our case-control study is not fully representative, the difference between the group was not notable. Meanwhile, it is clinically interesting that both patients with a HER2-enriched molecular subtype developed early distant metastasis, specifically 7 months and 11 months. Reviewing their clinical risk, we found they were not the patients with the highest clinical risk of recurrence: one patient had pT1 and pN1, and the other pT2 and pN1 disease. They all had received lumpectomy, ALND, radiotherapy, chemotherapy, and endocrine therapy. Although they had clinically been diagnosed with HR+/HER2− tumors at moderate risk of recurrence, they had the worst prognosis. Tumor biology was likely the underlying cause. Yet, since only two patients with HER2-enriched (HER2-E) molecular subtype were found, no statistical conclusions could be made. In brief, referring to PAM50 molecular subtypes in our patient cohort, the survival difference between luminal A and luminal B patients were small; but prognosis was poor for those patients with HER2-E subtype.

The molecular subtypes were not prognostic in our patients, but we found that the subtype scores, especially the HER2-E subtype scores, were significantly associated with the risk of distant metastasis and survival. Basically, this means that even for patients with luminal-like tumors, the prognosis is poor if their tumor biology is closer to the HER2-E subtype. The widely applied score based on the PAM50 algorithm is the risk of recurrence (ROR) score, which is currently recommended for clinical use in post-menopausal patients [31, 68]. In our premenopausal cohort, ROR score is prognostic for distant metastasis as well. Interestingly, our survival analysis suggested that HER2-E score has more significant prognostic power (lower *p*-value and higher hazard ratio) than ROR score. Further correlation analysis suggested that the three scores that are survival-related are correlated (luminal A score, HER2-E score, and ROR score). This is easy to understand considering they are all calculated based on expression of the same set of genes. As shown in the dot plots, patients with a HER2-E score higher than 0.25 were at a high risk of recurrence (all 7 patients developed distant metastasis), and patients with a ROR score lower than 30 were at relatively low risk (only 1 patient experienced distant metastasis out of 11 patients). Therefore, combined consideration of ROR score and HER2-E score has the potential to better stratify the patients. As far as we know, no other report has mentioned the prognostic value of the HER2-E

subtype score in HR+/HER2− patients. Therefore, validation of this observation in larger cohorts is necessary.

The in-depth evaluation of the molecular subtypes showed that even in luminal-like breast cancer, HER2-E biology exists and strongly influences patient survival. By reviewing related research, we found this is not a new finding. While reviewing HER2 amplification status in breast cancer, Daemen et al. noticed that HER2 amplification was not limited to breast cancer with HER2-E subtype but present in all breast cancer subtypes [117]. The association between HER2 amplification and early recurrence in HR+/HER2− breast cancer was also recently reported by Yamashita et al [118]. Besides, research into borderline HER2 expressing tumors also tried to further stratify the current HR+/HER2− breast cancer [119]. Taking together these investigations and our results, we suggest evaluation of HER2-E biology in luminal breast cancer to predict risk of recurrence.

4.3 BC360® signatures and the risk of developing distant metastasis

Based on the expression of 758 genes, expression of 46 breast cancer-related signatures was calculated for each patient. These signatures represent different aspects of cancer biology: anti-tumor immune activity, breast cancer prognosis, breast cancer receptors, breast cancer signaling pathways, breast cancer subtyping, immune cell abundance, inhibitory immune mechanisms, inhibitory immune signaling, inhibitory metabolism, stromal factors, tumor immunogenicity, tumor mutational response, and tumor regulation [99]. The wheel plot depicted the signatures scores for each subtype of patients and the correlation between the subtypes. Subtypes were not prognostic in our patient cohort. Therefore, the BC360® signatures of each subtype were not further compared statistically. The correlation between subtypes was not the emphasis of our study, but this information deepened our understanding of the subtypes: the subtypes are not completely independent, they had either positive or negative connection (e.g., luminal B subtype has a positive correlation with HER2-E subtype, while luminal A subtype has a negative correlation with HER2-E).

Four signatures were correlated with distant metastasis and survival in our cohort: ROR (as a signature, only the gene expression was calculated), PGR (progesterone receptor), claudin-low, and mammary stemness. Biologically, ROR belongs to “breast cancer prognosis”, PGR belongs to “breast cancer receptors”, claudin-low belongs to “breast cancer subtyping” and mammary stemness belongs to “tumor regulation”. ROR has the most significant *p*-value, while *p*-values of claudin-low and mammary stemness were only borderline significant. In order to avoid false discovery [120], the adjusted *p*-value was calculated for each signature [101] and no significance was noticed. Therefore, caution must be taken while interpreting the results and further validation with larger cohorts remains necessary. Moreover, according to the subgroup analysis, ROR expression remains prognostic in both luminal A and luminal B subgroups, and expression of PGR is more prognostic in the luminal A subgroup than in all patients.

ROR, as a signature, is the basis of the ROR score, which combines both the gene expression of ROR signature and clinical parameters [31, 68]. ROR score predicts the risk of distant recurrence in postmenopausal patients with ER+ EBC who were treated with adjuvant endocrine therapy [7, 68]. In our premenopausal cohort, ROR remains prognostic. But as discussed in the last chapter, ROR score was not as prognostic as HER2-E score. Therefore, whether ROR score is the best choice for predicting survival of premenopausal patients warrants further investigation.

The PGR signature includes one single gene PGR encoding the Progesterone Receptor (PR), which mediates the physiological effects of progesterone. Low PR expression is an established indicator for poor prognosis [10], and high PR expression is an indicator for better survival [121]. The prognostic power of PGR in luminal A patients needs to be stressed: PGR may not predict prognosis in luminal B patients.

The claudin-low subtype is defined by low expression of cell-cell adhesion genes, high expression of epithelial-mesenchymal transition genes, and stem cell-like/less differentiated gene expression patterns [122]. It is a complex additional phenotype which is poorly understood. There are different opinions regarding the prognostic impact of the claudin-low subtype: even though some research suggested that the claudin-low subtype related to poor prognosis, a recent cohort analysis suggested that claudin-low subtype was not an indicator of good or poor survivals [123]. The mammary stemness signature measures a cluster of EMT (Epithelial to mesenchymal transition) genes and stem cell relevant genes. Previous research suggested that breast stem cells are to blame for cancer relapse[124]. Considering the borderline *p*-values in our analysis of claudin-low and mammary stemness, and the discordant results in prior research, we believe that the current evidence is not enough for fully understanding their prognostic value.

The analysis of BC360 signatures helped us to connect our output with the established cancer biology. Higher expression of ROR is linked to poorer survival in our premenopausal cohorts, while higher expression of PGR is a favorable prognostic factor in luminal A breast cancer. Further validation remains necessary.

4.4 GSEA pathways and the risk of developing distant metastasis

Since its development, gene set enrichment has been widely used in bioinformatic analysis to explore the pathways involved in disease development [104, 125, 126]. We performed GSEA analysis with the same purpose as performing the BC® 360 signature analysis: To understand the underlying biology driving distant metastasis in our premenopausal cohort.

Overall, the most important finding lay in hallmark gene sets. In our cohort, mTORC1 (mammalian target of rapamycin complex 1) signaling was enriched in patients who developed distant metastasis, and enrichment was more prominent in luminal B patients. The phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT)/ mTOR pathway is crucial for breast cancer growth, progression, and insensitivity to endocrine interventions[10, 127, 128]. And mTORC1 signaling regulates many biological processes, including cell cycle progression, growth, and metabolism [129]. As a mTORC1 signaling inhibitor, everolimus is a validated medication for treating advanced breast cancer [11, 130]. The use of everolimus as second-line medication after disease progression is a standard of care [11] and subject of trials for therapy optimization [128, 131, 132]. Meanwhile, effort has been made to investigate the benefit of everolimus in earlier lines of treatment as a maintenance therapy, yet no survival benefit was noticed between patients who received everolimus and who did not [133]. As our results suggested, the enrichment of mTORC1 signaling is associated with distant metastasis and could be detected before the start of adjuvant treatment. Therefore, stratification of patients according to mTOC1 signaling enrichment may help to select patients who might benefit from an early combination of everolimus. Of note, mTORC1 activation marker was suggested as a predictive factor for everolimus benefit in advanced breast cancer [134]. Whether evaluation of mTOC1 signaling could lead to a more effective use of everolimus

in breast cancer requires a cautious judgement of pros and cons, and further exploration. Another enriched pathway is “BMI1_DN_MEL18_DN.V1_DN” belonging to “oncogenic gene sets”. This gene set includes genes that were downregulated in DAOY cells (medulloblastoma) upon knock-down of BMI1 (B lymphoma Mo-MLV insertion region 1 homolog) and PCGF2 (polycomb group ring finger 2) [135]. Since no clear former evidence was found regarding this pathway and breast cancer, we offer here first-time evidence, and wait for further validation of the observed correlation.

As discussed in an earlier chapter, even though luminal B subtype is considered as a more unfavorable prognostic marker than luminal A subtype [10], this difference was not obvious in our cohort. Since the metastasis rate was high in both luminal A and luminal B subtype, we were wondering if luminal A and luminal B have the same drivers for distant metastasis. Therefore, where applicable, after we carried out the analysis in all patients, we repeated the analyses in the luminal A and luminal B subgroup separately. Normally, a smaller sample size leads to weaker statistical significance. Yet, we found more significant results in subgroup analysis. In luminal A tumors, pathways regarding miR-9983-3p, IL-2 and IL-15, and E2F1 (retinoblastoma-associated protein 1) were significantly associated with distant metastasis. In luminal B tumors, besides mTORC1 signaling, a pathway regarding pre-mature aging was related to distant metastasis. Moreover, in luminal B tumors without distant metastasis, besides pathways regarding two microRNAs (miR4755-5p, miR5006-3p), an extra abundant enrichment of immunological gene sets was observed (five gene sets in total). Though no prior research regarding a connection between these specific pathways and breast cancer was found, immunological reaction was suggested as a crucial prognostic marker [136]. According to bioinformatic analyses, immunological signatures are not generally favorable or unfavorable prognostic factors [136], but presence of lymphocytic infiltration is generally considered as intense antitumor responses and a favorable prognostic marker [137]. In our luminal B cohort, several immunological pathways were protective against distant metastasis, and further investigation is crucial in understanding their actual function and potential application.

BC360® signature analysis offered us a possibility to understand the underlying biology of distant metastasis, GSEA further widened the tested range of the biological signatures. The advantage of GSEA results is that the false discovery rate of their *p*-values was also within the significant range, and therefore the conclusion is more reliable. The disadvantage is that except for the hallmark gene sets, the connection between other gene sets and breast cancer was not clear, and therefore requires further investigation. In brief, GSEA analysis confirmed the importance of mTORC1 signaling in predisposition for distant metastasis, especially in luminal B breast cancer. The underlying drivers of distant metastasis were not identical between luminal A and luminal B tumors: Immunology played a more significant role in distant metastasis of luminal B breast cancer.

4.5 Single gene markers and the risk of distant metastasis

Unlike signatures, which calculate the expression of several or dozens of genes at the same time, single-gene markers are more specific. Considering the lack of molecular signatures developed based on premenopausal patients [33, 91], single-gene markers could be more sensitive prognostic markers and more representative for our cohort. Therefore, we used the standard method [105] to explore prognostic single genes. In brief, in our study, twenty-two genes were defined as differentially expressed genes (DEGs), nineteen of which were significantly related to survival. Further database validation verified the prognostic value of fifteen DEGs as single markers and

the nineteen-gene signature as an integrated marker. Most of the DEGs were not covered by the prognostic signatures, which confirmed the necessity of researching single genes.

Among the nineteen survival-related DEGs, higher expression of two genes was linked to poor survival: IBSP and BCAS1, and low expression of the other seventeen genes were associated with poor survival: LRP2, SCUBE2, SFRP1, ZBTB16, MIA, OGN, ABCA8, PGR, IGF1, SOX10, ERBB4, IRX1, PTGER3, HOXA5, WNT11, CDC14A, and THBS4. Interactions between genes were also established in other analyses: IGF1 has a functional connection to ERBB4, PGR, OGN, IBSP LRP2, and SOX10; PGR has a functional connection to SCUBE2, ERBB4, HOXA5, and IGF1; SOX10 has a functional connection to MIA, BCAS1, and IGF1. Functional analyses suggested that these genes were significantly involved in the PI3K-AKT signaling pathway, breast cancer, focal adhesion, proteoglycans in cancer, and hedgehog pathway. As for the BC360® pathways of the genes, the significantly involved pathways are triple-negative biology (SFRP1, OGN, ABCA8, IGF1, IRX1, ERBB4, SOX10, MIA, SCUBE2), ER signaling (PGR, PTGER, SCUBE2), and hedgehog pathway (LRP2, WNT11). Of note, functional analysis only suggested functional enrichment of the 19 DEGs rather than functional involvement of pathways in distant metastasis, and therefore did not define the upper or lower regulation of certain pathways in cancers that developed distant metastasis.

4.5.1 The prognostic value of LRP2 in breast cancer

LRP2 (LDL receptor related protein 2), also known as megalin or glycoprotein 330, encodes a membrane glycoprotein belongs to the low-density lipoprotein receptor protein family [138]. This protein is typically present at the apical surface of the epithelial cells of embryonic and adult tissues, and the latter includes central nerve system, kidney, lung, thyroid, gallbladder and mammary gland [138]. The presence of LRP2 is crucial for normal differentiation of embryos [139] and the aberration of LRP2 has been recognized in many clinical situations, including diabetic nephropathy, Lowe syndrome, Dent disease, Alzheimer's disease, and gallstone disease [138]. Besides, higher expression of LRP2 has recently been reported as a favorable prognostic factor in renal cell carcinoma [140]. In our study, LRP2 had a lower expression in patients who developed metastasis and higher LRP2 levels were significantly associated with better survival, which implies that LRP2 is a favorable prognostic factor in premenopausal breast cancer.

Former direct evidence regarding the prognostic value of LRP2 in breast cancer is rare, but connections between LRP2 and breast cancer have been investigated. One connection was established through the vitamin D metabolism. Besides being a modulator of calcium homeostasis and osteosynthesis, vitamin D is also a regulator of the immune, muscular, and nervous systems [141]. Studies suggested that vitamin D deficiency is related to higher risk of breast cancer [141, 142], and the vitamin D receptor (VDR) is a potential tumor suppressor [143]. Therefore, mechanisms regarding metabolism of vitamin D were suggested to be influential in breast cancer biology [142, 143]. To be noted, LRP2 functions in the uptake and activation processes of vitamin D in mammary cells [139, 144], and the single nucleotide polymorphism of the LRP2 gene influence breast cancer risk in premenopausal women [145]. Nevertheless, due to the lack of direct evidence, further investigations are necessary to determine whether LRP2 influences the distant metastasis of breast cancer in premenopausal patients through influencing vitamin D metabolism. Besides, LRP2 is noticed as a critical regulator in the sonic hedgehog pathway, which is important in developmental processes and tumorigenesis [146, 147]. Other connections lie in the regulation of LRP2 expression. First, the peroxisome proliferator activated receptor (PPAR) α agonist is a tumor suppressor in breast cancer [148] and could increase the LRP2 expression [138]. Second,

PGR is a favorable prognostic factor in breast cancer [10], and one study suggested that LRP2 is a target gene of progesterone and PGR [149]. Third, retinoic acid prevents migration of breast cancer cells [150] and could induce expression of LRP2 [144]. Fourth, AGTR1 (angiotensin II receptor type 1) overexpression promotes proliferation of breast cancer [151] and was suggested as a new maker to select patients that may be better targeted with an AGTR antagonist [152]. One study suggested that an AGTR antagonist has a protective effect regarding LRP2 expression [138]. Considering the scarcity of direct evidence, further validation of the connections between LRP expression and breast cancer is necessary.

Our study highlighted the highly significant prognostic value of LRP2 in premenopausal patients with HR+/HER2− breast cancer, and further studies remain necessary to validate its prognostic value, investigate the underlying mechanism, and develop treatment approaches.

4.5.2 The prognostic value of SCUBE2 in breast cancer

SCUBE2 (signal peptide, CUB domain and EGF like domain containing 2) encodes a secreted, membrane-associated protein which was reported as a breast tumor suppressor [153, 154], a favorable prognostic factor [155] and is included in both MammaPrint® and Oncotype DX®, which are widely used diagnostic molecular tests to predict the risk of recurrence for breast cancer [153, 156]. Studies suggested that the anti-progression effect of SCUBE2 relies on reversing the epithelial-mesenchymal transition [157], antagonizing bone morphogenetic protein, suppressing the β-catenin pathway [154], and mediating the hedgehog signaling pathway [158]. Although the prognostic value of SCUBE2 is clear, the prognostic power of SCUBE2 in premenopausal patients has not yet not been elucidated. According to our study, SCUBE2 was a favorable prognostic factor in premenopausal patients with HR+/HER2− breast cancer.

4.5.3 Other single prognostic gene markers in breast cancer

The tumor suppressing capacity and prognostic value of SFRP1, ZBTB16, OGN, PGR, PTGER3, and HSPA2 in breast cancer have been examined by other studies. SFRP1(secreted frizzled-related protein 1) encodes a secreted protein that interacts with the cell membrane and Wnt proteins [159]. SFRP1 protein inhibits the Wnt signaling pathway, which is critical in developmental processes and involved in tumorigenesis [159]. Decreased expression of SFRP1 is a predictor for poor prognosis in breast cancer patients [160]. ZBTB16 (zinc finger and BTB domain containing 16) encodes a zinc finger transcription factor in the nuclear that functions in the cell cycle, transcription process, and protein dimerization transformation [161]. ZBTB16 is under-expressed or silenced in multiple cancer tissues, including breast cancer [161]. According to a recent study, decreased ZBTB16 expression is associated with poor survival, and restoration of ZBTB16 expression inhibits proliferation, cell cycling, and invasion of breast cancer cells [161]. OGN (osteoglycin) encodes a secreted proteoglycan that is associated with bone metabolism, endocrine regulation [162], and tumorigenesis [163]. According to a recent study, low expression of OGN was observed in breast cancer samples compared to normal breast tissues and in tumors with poor prognosis. Moreover, overexpression of OGN could inhibit the malignant biology of breast cancer cells and reverse EMT through PI3K/Akt/mTOR signaling pathway [163]. The prognostic value of PGR was discussed earlier in the signature chapter. In brief, PGR is a validated favorable prognostic factor in breast cancer [10]. PTGER3 (prostaglandin E2 receptor 3) or EP3 encodes a G-coupled protein that locates in the membrane and mediates the function of PGE2 (prostaglandin E2), which actively participates in the physiological processes, inflammation, and tumorigenesis.

One recent study established a connection between EP3 expression and patient survival: Positive EP3 expression was linked to good prognosis [164]. HSPA2 (heat shock protein family A member 2) encodes a molecular chaperone that interact with unfolded, mis-folded, and semi-native proteins. Recent studies advocated the favorable prognostic value of high HSPA2 expression in breast cancer [165, 166].

The tumor suppressing capability of ERBB4, IRX1, SFRP4 in breast cancer has previously been investigated by other studies. ERBB4 (epidermal growth factor receptor 4) encodes a membrane receptor that regulates embryonic development [167]. As an exception in the family, ERBB4 was recognized as a tumor suppressor in multiple cancers, including breast cancer. In breast cancer cells, overexpression of ERBB4 could impair proliferation and even induce apoptosis [167]. IRX1 (Iroquois Homeobox 1) encodes a transcription factor that is crucial in embryonic development [168]. The tumor suppressing effect of IRX1 was first validated in gastric cancer [169]. One recent study examined the tumor suppressing function of IRX in breast cancer: Down-regulation of IRX in ductal carcinoma compared to normal breast tissue was seen, and overexpression of IRX in breast cancer cells significantly hindered cell proliferation [170]. SFRP4 (secreted frizzled-related protein 4) encodes a protein that belongs to the same family as SFRP1. SFRP4 is a Wnt antagonist and could induce apoptosis and inhibit angiogenesis in cancer [171]. Activation of SFRP4 helps to alleviate chemoresistance of breast cancer cells [171] and inhibits the viability of breast cancer stem cells [172]. But the prognostic value of SFRP4 has not yet been investigated.

The tumor suppressing capability of ABCA8 and CDC14A in other cancers was investigated by prior studies. ABCA8 (ATP binding cassette subfamily A member 8) encodes a membrane associated protein functioning as a transmembrane transporter for organics, such as efflux of cholesterol and drugs. Down-regulation of ABCA8 was noticed in hepatocellular carcinoma (HCC) [173] and breast cancer tissues [174]. ABCA8 was suggested as a favorable prognostic factor for patients with HCC and low expression of ABCA8 induced EMT via the ERK (Extracellular-signal regulated kinase)/ZEB1(Zinc finger E-box binding homeobox 1) signaling pathway [173]. Yet, no studies have investigated the prognostic value of ABCA8 expression in breast cancer so far. CDC14A (cell division cycle 14A) encodes a cellular phosphatase that drives mitotic exit in Saccharomyces cerevisiae. In human cells, CDC14A was found to target p53 and to interact with Cdk1/cyclin B and epithelial protein in order to participate in carcinogenesis [175, 176]. Reduced CDC14A levels were associated with the risk of colorectal carcinoma and poor patient prognosis [176]. The prognostic value of CDC14A has not yet been validated in breast cancer.

The tumor driving potential of IBSP, BCAS1 and STC1 has been evaluated by prior studies. IBSP (integrin binding sialoprotein) encodes a secreted glycoprotein that was first discovered in mineralized tissues and subsequently found to be aberrantly expressed in various kinds of malignancies [177]. Increased expression of IBSP (also known as BSP, bone sialoprotein) was reported to be associated with a increased risk of bone metastasis in breast cancer [177, 178], especially in estrogen receptor positive tumors [179]. BCAS1 (brain enriched myelin associated protein 1, also known as NABC1, novel amplified breast cancer 1) encodes a cytoplasmic protein that amplifies in breast cancer cells [180]. BCAS1 was considered as an oncogene in breast cancer [180], but its functioning mechanism and prognostic value remain unclear. STC1 (Stanniocalcin-1) encodes a secreted glycoprotein that functions as a hormone in regulating calcium and phosphate homeostasis [181]. The oncogenic role of STC1 in breast cancer has been previously investigated and high STC1 levels were associated with poor patient outcome [182, 183].

Regarding these genes, our study offered further evidence for premenopausal patients with HR+/HER2− breast cancer. Our result and prior studies all support the potential utility of theses single gene markers as prognostic indicators.

Inconsistent evidence regarding MIA, IGF1, SOX10, HOXA5, WNT11, and THBS4 was found. MIA (melanoma inhibitory activity), as the name suggested, encodes a protein that can be secreted to the extracellular space and that is highly prognostic in malignant melanoma [184]. Recent studies believed that MIA is an oncogene and could promote distant metastasis and immune suppression in malignant melanoma and lung cancer [185, 186], but no evidence regarding MIA's functions in breast cancer has yet been published. IGF1 (insulin-like growth factor 1) encodes a secreted protein that can bind to IGF-1R (insulin-like growth factor 1 receptor type 1) and activate downstream transduction events. IGF1 was generally considered as an oncogene in multiple cancers, including breast cancer [187]. But the prognostic value of IGF1 regarding breast cancer recurrence is inconsistent and therefore under discussion [188]. SOX10 (Sry-related high-mobility-group/HMG box 10) encodes a transcription factor that regulates embryonic development and cell death. SOX10 is a frequently used marker for recognizing melanoma and triple negative breast cancer [189]. The prognostic value of SOX10 in luminal breast cancer is not established. HOXA5 (homeobox A5) encodes a transcription factor that is crucial in embryogenesis and involved in multiple cancers, including breast cancer [190]. Discrepancy has been noticed in deciphering the contribution of HOXA5 to breast cancer evolvement. Altough one study acknowledged the anti-tumor potential of HOXA5 [191], another study found that HOXA5 propels endocrine resistance in breast cancer [190]. HOXA5 was confirmed as a tumor suppressor in gastric cancer, and low expression of HOXA5 was associated with poorer survival [192]; yet, the prognostic value of HOX5A remains unestablished. WNT11 (Wnt family member 11) mainly encodes a secreted glycoprotein that is associated with the extracellular matrix and plays an important part in the development process and tumorigenesis [193]. The exact role of WNT11 in cancer has not yet been defined: wihile one study suggested that WNT11 is an oncogene, others noticed a tumor suppressing effect of WNT11 [193]. No direct evidence has validated the prognostic value of WNT11 in breast cancer. THBS4 (thrombospondin 4) encodes an extracellular calcium-binding protein that interacts with extracellular matrix and functions in embryogenesis, wound healing, and cancer development [194]. One study suggested that, in normal breast, THBS4 has the most abundant expression in the basement membrane surrounding ducts, fibroblasts and endothelial cells. During tumor progression, THBS4 was up-regulated in the extracellular matrix. Therefore, the author concluded that THBS4 may contribute to the invasion of breast cancer cells [194]. Nevertheless, in this case, THBS4 may also present to repair the lesions caused by cancer. Due to lack of direct evidence, the actual function of THBS4 and its prognostic value in invasive breast cancer is not yet clear. Regarding these genes, the inconsistency of their roles in cancer reflects the complexity of their functioning mechanisms. Further basic and clinical research would be necessary to clarify their actual contributions in cancer development and progression.

Among the 22 metastasis-associated genes, 19 genes showed significant survival relevance and were defined as DMFS-related DEGs. Furthermore, the prognostic values of fifteen of the DMFS-related DEGs in HR+/HER2− breast cancer (of note, both postmenopausal and premenopausal patients were included using online databases) were further validated by online databases. Through literature review, we found supportive evidence for 13 out of the 19 genes, while the other six genes showed more complex functions in carcinogenesis. Briefly, our sample-based case-cohort study offered direct evidence for the prognostic values of 19 genes in premenopausal patients with HR+/HER2− breast cancer: Both uniqueness and similarity exist in the molecular drivers for distant metastasis of breast cancer in premenopausal vs. postmenopausal patients.

Nevertheless, further validation is necessary for developing molecular prognostic tests for premenopausal patients with HR+/HER2− breast cancer.

4.6 Summary of the prognostic factors

As could be foreseen, there are certain correlations between prognostic factors: Some prognostic factors may be the underlying drivers for changes in other factors, and the factors could influence each other in the complex processes of distant metastasis. Most importantly, the prognostic impact of novel prognostic factors must be compared with that of the classic prognostic factors (pN, pT, ROR) to confirm their clinical value [36]. Therefore, multivariate survival analysis that included all significant DMFS-related prognostic factors (pN, pT, DMFS-related PAM50 subtype scores, DMFS-related BC360® signature scores, nineteen DMFS-related DEGs) was performed in our study as a final summary of all prognostic factors.

The analysis details of the multivariate survival analysis was introduced in the Methods chapter and the output in the Results section. Briefly, LRP2 was selected as the most significant prognostic factor. Besides, PTGER3 and pN have independent prognostic values and could further refine the survival model. Meanwhile, multivariate analysis did not show interactions of the prognostic factors and could not explain why other significant prognostic factors were not selected in the final survival model.

We therefore investigated the correlation between LRP2 and other prognostic factors. As correlation analysis suggested, expression of LRP2 was not correlated with pN or pT, but was correlated with most of the molecular prognostic factors (18/23). LRP2 had the most significant survival relevance among these factors and therefore was selected out as a representative factor. Based on both the multivariate and correlation analysis, we assumed that the lowered expression of a gene set that was LRP2-centered and included SFRP1, CDC14A, OGN, ABCA8, IGF1, IRX1, ERBB4, SOX10, MIA, PGR, THBS4, ZBTB16 was associated with a high risk of distant metastasis. Besides, the highly significant negative correlation between LRP2 and HER2-E score/ROR score confirmed the potential of LRP2 as a prognostic marker. As for BCAS1, IBSP, SCUBE2, and PTGER3, they were not correlated with LRP2, and therefore may represent other independent drivers in distant metastasis.

Regarding the mechanism of LRP2's functions in breast cancer, no direct evidence has yet been published, but indirect evidence indicates LRP2 may function through regulating vitamin D metabolism [139] and the hedgehog signaling pathway [146]. As our correlation analysis suggested, LRP2 has a significant correlation with the HER2-E score (negatively), SFRP1, CDC14A, and ABCA8. Therefore, the functions of LRP2 might relate to HER2 biology (negatively), Wnt signaling (negatively), membrane transport, and cell cycle regulation. But direct evidence remains necessary in deciphering the exact roles of LRP2 in breast cancer.

The prognostic impact of the molecular factors was summarized by multivariate analysis and the result confirmed the importance of the molecular prognostic factors, especially that of LRP2. However, more validation studies are necessary to establish a reliable molecular signature to predict prognosis or even chemotherapy benefit in premenopausal patients with HR+/HER2- breast cancer. Besides, the exact functions of theses DMFS-related molecular factors seem more complex than what we already know. Therefore, basic research is warranted to decipher important connections and their exact roles in carcinogenesis.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries**. *CA Cancer J Clin* 2021, **71**(3):209-249.
2. Faguet GB: **A brief history of cancer: age-old milestones underlying our current knowledge database**. *Int J Cancer* 2015, **136**(9):2022-2036.
3. Sainsbury R: **The development of endocrine therapy for women with breast cancer**. *Cancer Treat Rev* 2013, **39**(5):507-517.
4. Zurrida S, Veronesi U: **Milestones in breast cancer treatment**. *Breast J* 2015, **21**(1):3-12.
5. Kellenberger E: **The evolution of molecular biology**. *EMBO Rep* 2004, **5**(6):546-549.
6. DeVita VT, Jr., Rosenberg SA: **Two hundred years of cancer research**. *N Engl J Med* 2012, **366**(23):2207-2214.
7. Litton JK, Burstein HJ, Turner NC: **Molecular Testing in Breast Cancer**. *Am Soc Clin Oncol Educ Book* 2019, **39**:e1-e7.
8. Collins FS, Doudna JA, Lander ES, Rotimi CN: **Human Molecular Genetics and Genomics - Important Advances and Exciting Possibilities**. *N Engl J Med* 2021, **384**(1):1-4.
9. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**(6797):747-752.
10. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, Ruddy K, Tsang J, Cardoso F: **Breast cancer**. *Nat Rev Dis Primers* 2019, **5**(1):66.
11. NCCN: **NCCN, Clinical Practice Guidelines in Oncology, Breast Cancer**, Version 4.2021
12. Clarke R, Tyson JJ, Dixon JM: **Endocrine resistance in breast cancer – An overview and update**. *Molecular and Cellular Endocrinology* 2015, **418**:220-234.
13. Andre F, Ismaila N, Henry NL, Somerfield MR, Bast RC, Barlow W, Collyar DE, Hammond ME, Kuderer NM, Liu MC *et al*: **Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: ASCO Clinical Practice Guideline Update-Integration of Results From TAILORx**. *J Clin Oncol* 2019:JCO1900945.
14. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT *et al*: **Gene expression profiling predicts clinical outcome of breast cancer**. *Nature* 2002, **415**(6871):530-536.
15. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, Nielsen TO, Gelmon K: **Metastatic behavior of breast cancer subtypes**. *J Clin Oncol* 2010, **28**(20):3271-3277.
16. Curigliano G, Burstein HJ, Winer EP, Gnant M, Dubsky P, Loibl S, Colleoni M, Regan MM, Piccart-Gebhart M, Senn HJ *et al*: **De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017**. *Ann Oncol* 2017, **28**(8):1700-1712.
17. Altun I, Sonkaya A: **The Most Common Side Effects Experienced by Patients Were Receiving First Cycle of Chemotherapy**. *Iran J Public Health* 2018, **47**(8):1218-1219.
18. EBCTCG: **Polychemotherapy for early breast cancer: an overview of the randomised trials**. *The Lancet* 1998, **352**(9132):930-942.
19. EBCTCG: **Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials**. *The Lancet* 2012, **379**(9814):432-444.

20. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al*: **Supervised risk predictor of breast cancer based on intrinsic subtypes**. *J Clin Oncol* 2009, **27**(8):1160-1167.

21. Prat A, Pineda E, Adamo B, Galvan P, Fernandez A, Gaba L, Diez M, Viladot M, Arance A, Munoz M: **Clinical implications of the intrinsic molecular subtypes of breast cancer**. *Breast* 2015, **24 Suppl 2**:S26-35.

22. Loi S, Piccart M, Sotiriou C: **The use of gene-expression profiling to better understand the clinical heterogeneity of estrogen receptor positive breast cancers and tamoxifen response**. *Crit Rev Oncol Hematol* 2007, **61**(3):187-194.

23. Duffy MJ, Harbeck N, Nap M, Molina R, Nicolini A, Senkus E, Cardoso F: **Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM)**. *Eur J Cancer* 2017, **75**:284-298.

24. Ades F, Zardavas D, Bozovic-Spasojevic I, Pugliano L, Fumagalli D, de Azambuja E, Viale G, Sotiriou C, Piccart M: **Luminal B breast cancer: molecular characterization, clinical management, and future perspectives**. *J Clin Oncol* 2014, **32**(25):2794-2803.

25. Taskaynatan H, Kucukzeybek Y, Alacacioglu A, Yildiz Y, Salman T, Oflazoglu U, Varol U, Bolat Kucukzeybek B, Kemal Atahan M, Oktay Tarhan M: **Is adjuvant chemotherapy necessary for Luminal A-like breast cancer?** *J BUON* 2018, **23**(4):877-882.

26. Kwak HY, Chae BJ, Eom YH, Hong YR, Seo JB, Bae JS, Jung SS, Song BJ: **Is adjuvant chemotherapy omissible in women with T1-2 stage, node-positive, luminal A type breast cancer?** *J Chemother* 2015, **27**(5):290-296.

27. Herr D, Wischnewsky M, Joukhadar R, Chow O, Janni W, Leinert E, Fink V, Stuber T, Curtaz C, Kreienberg R *et al*: **Does chemotherapy improve survival in patients with nodal positive luminal A breast cancer? A retrospective Multicenter Study**. *PLoS One* 2019, **14**(7):e0218434.

28. Harbeck N, Gnant M: **Breast cancer**. *Lancet* 2017, **389**(10074):1134-1150.

29. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ *et al*: **A gene-expression signature as a predictor of survival in breast cancer**. *N Engl J Med* 2002, **347**(25):1999-2009.

30. Bueno-de-Mesquita JM, Linn SC, Keijzer R, Wesseling J, Nuyten DS, van Krimpen C, Meijers C, de Graaf PW, Bos MM, Hart AA *et al*: **Validation of 70-gene prognosis signature in node-negative breast cancer**. *Breast Cancer Res Treat* 2009, **117**(3):483-495.

31. Ohnstad HO, Borgen E, Falk RS, Lien TG, Aaserud M, Sveli MAT, Kyte JA, Kristensen VN, Geitvik GA, Schlichting E *et al*: **Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up**. *Breast Cancer Res* 2017, **19**(1):120.

32. Wuerstlein R, Sotlar K, Gluz O, Otremba B, von Schumann R, Witzel I, Schindlbeck C, Janni W, Schem C, Bauerfeind I *et al*: **The West German Study Group Breast Cancer Intrinsic Subtype study: a prospective multicenter decision impact study utilizing the Prosigna assay for adjuvant treatment decision-making in estrogen-receptor-positive, HER2-negative early-stage breast cancer**. *Curr Med Res Opin* 2016, **32**(7):1217-1224.

33. Azim HA, Jr., Davidson NE, Ruddy KJ: **Challenges in Treating Premenopausal Women with Endocrine-Sensitive Breast Cancer**. *Am Soc Clin Oncol Educ Book* 2016, **35**:23-32.

34. McCart Reed AE, Kalita-De Croft P, Kutasovic JR, Saunus JM, Lakhani SR: **Recent advances in breast cancer research impacting clinical diagnostic practice**. *J Pathol* 2019, **247**(5):552-562.

35. Chang MC, Souter LH, Kamel-Reid S, Rutherford M, Bedard P, Trudeau M, Hart J, Eisen A, Molecular Oncology Advisory C: **Clinical utility of multigene profiling assays in early-stage breast cancer**. *Curr Oncol* 2017, **24**(5):e403-e422.

36. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T *et al*: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer**. *N Engl J Med* 2004, **351**(27):2817-2826.

37. Mosly D, Turnbull A, Sims A, Ward C, Langdon S: **Predictive markers of endocrine response in breast cancer**. *World J Exp Med* 2018, **8**(1):1-7.

38. Mamounas EP, Tang G, Fisher B, Paik S, Shak S, Costantino JP, Watson D, Geyer CE, Jr., Wickerham DL, Wolmark N: **Association between the 21-gene recurrence score assay and risk of locoregional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20**. *J Clin Oncol* 2010, **28**(10):1677-1683.

39. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J *et al*: **Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer**. *J Clin Oncol* 2006, **24**(23):3726-3734.

40. Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, Pho M, Baker J, Walker M, Watson D, Hackett J *et al*: **A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients**. *Breast Cancer Res* 2006, **8**(3):R25.

41. Zhang L, Hsieh MC, Petkov V, Yu Q, Chiu YW, Wu XC: **Trend and survival benefit of Oncotype DX use among female hormone receptor-positive breast cancer patients in 17 SEER registries, 2004-2015**. *Breast Cancer Res Treat* 2020, **180**(2):491-501.

42. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE, Jr., Dees EC, Perez EA, Olson JA, Jr. *et al*: **Prospective Validation of a 21-Gene Expression Assay in Breast Cancer**. *N Engl J Med* 2015, **373**(21):2005-2014.

43. Nitz U, Gluz O, Christgen M, Kates RE, Clemens M, Malter W, Nuding B, Aktas B, Kuemmel S, Reimer T *et al*: **Reducing chemotherapy use in clinically high-risk, genomically low-risk pN0 and pN1 early breast cancer patients: five-year data from the prospective, randomised phase 3 West German Study Group (WSG) PlanB trial**. *Breast Cancer Res Treat* 2017, **165**(3):573-583.

44. Stemmer SM, Steiner M, Rizel S, Geffen DB, Nisenbaum B, Peretz T, Soussan-Gutman L, Bareket-Samish A, Isaacs K, Rosengarten O *et al*: **Clinical outcomes in ER+ HER2 -node-positive breast cancer patients who were treated according to the Recurrence Score results: evidence from a large prospectively designed registry**. *NPJ Breast Cancer* 2017, **3**:32.

45. Thibodeau S, Voutsadakis IA: **The Oncotype Dx Assay in ER-Positive, HER2-Negative Breast Cancer Patients: A Real Life Experience from a Single Cancer Center**. *Eur J Breast Health* 2019, **15**(3):163-170.

46. Rizki H, Hillyar C, Abbassi O, Miles-Dua S: **The Utility of Oncotype DX for Adjuvant Chemotherapy Treatment Decisions in Estrogen Receptor-positive, Human Epidermal Growth Factor Receptor 2-negative, Node-negative Breast Cancer**. *Cureus* 2020, **12**(3):e7269.

47. Crolley VE, Marashi H, Rawther S, Sirohi B, Parton M, Graham J, Vinayan A, Sutherland S, Rigg A, Wadhawan A *et al*: **The impact of Oncotype DX breast cancer assay results on clinical practice: a UK experience**. *Breast Cancer Res Treat* 2020, **180**(3):809-817.

48. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE, Jr., Dees EC, Goetz MP, Olson JA, Jr. *et al*: **Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer**. *N Engl J Med* 2018, **379**(2):111-121.

49. Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, Lin NU, Perez EA, Goldstein LJ, Chia SKL *et al*: **21-Gene Assay to Inform Chemotherapy Benefit in Node-Positive Breast Cancer**. *N Engl J Med* 2021, **385**(25):2336-2347.

50. AGENDIA: **MammaPrint**, accessed 1 Oct 2021, < http://www.agendia.com/our-tests/mammaprint/ >

51. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, d'Assignies MS, Bergh J, Lidereau R, Ellis P *et al*: **Validation and clinical utility of a 70-gene prognostic**

signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006, **98**(17):1183-1192.

52. Straver ME, Glas AM, Hannemann J, Wesseling J, van de Vijver MJ, Rutgers EJ, Vrancken Peeters MJ, van Tinteren H, Van't Veer LJ, Rodenhuis S: **The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer**. *Breast Cancer Res Treat* 2010, **119**(3):551-558.

53. Knauer M, Mook S, Rutgers EJ, Bender RA, Hauptmann M, van de Vijver MJ, Koornstra RH, Bueno-de-Mesquita JM, Linn SC, van 't Veer LJ: **The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer**. *Breast Cancer Res Treat* 2010, **120**(3):655-661.

54. Drukker CA, van Tinteren H, Schmidt MK, Rutgers EJ, Bernards R, van de Vijver MJ, Van't Veer LJ: **Long-term impact of the 70-gene signature on breast cancer outcome**. *Breast Cancer Res Treat* 2014, **143**(3):587-592.

55. Bueno-de-Mesquita JM, van Harten WH, Retel VP, van 't Veer LJ, van Dam FS, Karsenberg K, Douma KF, van Tinteren H, Peterse JL, Wesseling J *et al*: **Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER)**. *Lancet Oncol* 2007, **8**(12):1079-1087.

56. Drukker CA, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H, Wesseling J, Roumen RM, Knauer M, van 't Veer LJ, Sonke GS *et al*: **A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study**. *Int J Cancer* 2013, **133**(4):929-936.

57. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, Pierga JY, Brain E, Causeret S, DeLorenzi M *et al*: **70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer**. *N Engl J Med* 2016, **375**(8):717-729.

58. Soliman H, Shah V, Srkalovic G, Mahtani R, Levine E, Mavromatis B, Srinivasiah J, Kassar M, Gabordi R, Qamar R *et al*: **MammaPrint guides treatment decisions in breast Cancer: results of the IMPACt trial**. *BMC Cancer* 2020, **20**(1):81.

59. Wallden B, Storhoff J, Nielsen T, Dowidar N, Schaper C, Ferree S, Liu S, Leung S, Geiss G, Snider J *et al*: **Development and verification of the PAM50-based Prosigna breast cancer gene signature assay**. *BMC Med Genomics* 2015, **8**:54.

60. Nielsen T, Wallden B, Schaper C, Ferree S, Liu S, Gao D, Barry G, Dowidar N, Maysuria M, Storhoff J: **Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens**. *BMC Cancer* 2014, **14**:177.

61. Jensen MB, Laenkholm AV, Nielsen TO, Eriksen JO, Wehn P, Hood T, Ram N, Buckingham W, Ferree S, Ejlertsen B: **The Prosigna gene expression assay and responsiveness to adjuvant cyclophosphamide-based chemotherapy in premenopausal high-risk patients with breast cancer**. *Breast Cancer Res* 2018, **20**(1):79.

62. Chia SK, Bramwell VH, Tu D, Shepherd LE, Jiang S, Vickery T, Mardis E, Leung S, Ung K, Pritchard KI *et al*: **A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen**. *Clin Cancer Res* 2012, **18**(16):4465-4472.

63. Veracyte: **Prosigna**, accessed 1 Oct 2021, < https://www.prosigna.com/en-gb/ >

64. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, Davies SR, Snider J, Stijleman IJ, Reed J *et al*: **A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer**. *Clin Cancer Res* 2010, **16**(21):5222-5232.

65. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS *et al*: **Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer**. *J Natl Cancer Inst* 2009, **101**(10):736-750.

66. Pu M, Messer K, Davies SR, Vickery TL, Pittman E, Parker BA, Ellis MJ, Flatt SW, Marinac CR, Nelson SH *et al*: **Research-based PAM50 signature and long-term breast cancer survival**. *Breast Cancer Res Treat* 2020, **179**(1):197-206.

67. Hequet D, Callens C, Gentien D, Albaud B, Mouret-Reynier MA, Dubot C, Cottu P, Huchon C, Zilberman S, Berseneff H *et al*: **Prospective, multicenter French study evaluating the clinical impact of the Breast Cancer Intrinsic Subtype-Prosigna(R) Test in the management of early-stage breast cancers**. *PLoS One* 2017, **12**(10):e0185753.

68. Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z, Mlineritsch B, Kwasny W, Knauer M, Singer C *et al*: **Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone**. *Ann Oncol* 2014, **25**(2):339-345.

69. Myriad genetics: **EndoPredict® Brest Cancer Prognosic Test**, accessed Oct 1 2021, < https://myriad.com/managed-care/endopredict/ >

70. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, Dietze O, Greil R, Jelen A, Sevelda P *et al*: **A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors**. *Clin Cancer Res* 2011, **17**(18):6012-6020.

71. Martin M, Brase JC, Calvo L, Krappmann K, Ruiz-Borrego M, Fisch K, Ruiz A, Weber KE, Munarriz B, Petry C *et al*: **Clinical validation of the EndoPredict test in node-positive, chemotherapy-treated ER+/HER2- breast cancer patients: results from the GEICAM 9906 trial**. *Breast Cancer Res* 2014, **16**(2):R38.

72. Filipits M, Dubsky P, Rudas M, Greil R, Balic M, Bago-Horvath Z, Singer CF, Hlauschek D, Brown K, Bernhisel R *et al*: **Prediction of Distant Recurrence Using EndoPredict Among Women with ER(+), HER2(-) Node-Positive and Node-Negative Breast Cancer Treated with Endocrine Therapy Only**. *Clin Cancer Res* 2019, **25**(13):3865-3872.

73. Villarreal-Garza C, Lopez-Martinez EA, Deneken-Hernandez Z, Maffuz-Aziz A, Munoz-Lozano JF, Barragan-Carrillo R, Ramos-Elias P, Moreno B, Diaz-Perez H, Pena-Curiel O *et al*: **Change in therapeutic management after the EndoPredict assay in a prospective decision impact study of Mexican premenopausal breast cancer patients**. *PLoS One* 2020, **15**(3):e0228884.

74. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, Cai Y, Bielski CM, Donoghue MTA, Jonsson P *et al*: **The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers**. *Cancer Cell* 2018, **34**(3):427-438 e426.

75. Osborne CK, Schiff R: **Mechanisms of endocrine resistance in breast cancer**. *Annu Rev Med* 2011, **62**:233-247.

76. Singh D, Attri BK, Gill RK, Bariwal J: **Review on EGFR Inhibitors: Critical Updates**. *Mini Rev Med Chem* 2016, **16**(14):1134-1166.

77. Chen YM, Qi S, Perrino S, Hashimoto M, Brodt P: **Targeting the IGF-Axis for Cancer Therapy: Development and Validation of an IGF-Trap as a Potential Drug**. *Cells* 2020, **9**(5).

78. Guo YJ, Pan WW, Liu SB, Shen ZF, Xu Y, Hu LL: **ERK/MAPK signalling pathway and tumorigenesis**. *Exp Ther Med* 2020, **19**(3):1997-2007.

79. Aoki M, Fujishita T: **Oncogenic Roles of the PI3K/AKT/mTOR Axis**. *Curr Top Microbiol Immunol* 2017, **407**:153-189.

80. Goel S, DeCristo MJ, McAllister SS, Zhao JJ: **CDK4/6 Inhibition in Cancer: Beyond Cell Cycle Arrest**. *Trends Cell Biol* 2018, **28**(11):911-925.

81. Hamilton E, Infante JR: **Targeting CDK4/6 in patients with cancer**. *Cancer Treat Rev* 2016, **45**:129-138.

82. Ellis H, Ma CX: **PI3K Inhibitors in Breast Cancer Therapy**. *Curr Oncol Rep* 2019, **21**(12):110.

83. Harbeck N, Rastogi P, Martin M, Tolaney SM, Shao ZM, Fasching PA, Huang CS, Jaliffe GG, Tryakin A, Goetz MP *et al*: **Adjuvant abemaciclib combined with endocrine therapy for high-risk early breast cancer: updated efficacy and Ki-67 analysis from the monarchE study**. *Ann Oncol* 2021, **32**(12):1571-1581.

84. Radecka B, Litwiniuk M: **Breast cancer in young women**. *Ginekol Pol* 2016, **87**(9):659-663.

85. Kim EK, Noh WC, Han W, Noh DY: **Prognostic significance of young age (<35 years) by subtype based on ER, PR, and HER2 status in breast cancer: a nationwide registry-based study**. *World J Surg* 2011, **35**(6):1244-1253.

86. Vaz-Luis I, Francis PA, Di Meglio A, Stearns V: **Challenges in Adjuvant Therapy for Premenopausal Women Diagnosed With Luminal Breast Cancers**. *Am Soc Clin Oncol Educ Book* 2021, **41**:1-15.

87. Ahn SH, Son BH, Kim SW, Kim SI, Jeong J, Ko SS, Han W, Korean Breast Cancer S: **Poor outcome of hormone receptor-positive breast cancer at very young age is due to tamoxifen resistance: nationwide survival data in Korea--a report from the Korean Breast Cancer Society**. *J Clin Oncol* 2007, **25**(17):2360-2368.

88. Suter MB, Pagani O: **Should age impact breast cancer management in young women? Fine tuning of treatment guidelines**. *Ther Adv Med Oncol* 2018, **10**:1758835918776923.

89. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-DeMore N, Perou CM: **Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes?** *J Clin Oncol* 2011, **29**(1):e18-20.

90. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, Zhang Y, Wang Y, Marcom PK, Marks JR, Febbo PG *et al*: **Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression**. *J Clin Oncol* 2008, **26**(20):3324-3330.

91. Liao S, Hartmaier RJ, McGuire KP, Puhalla SL, Luthra S, Chandran UR, Ma T, Bhargava R, Modugno F, Davidson NE *et al*: **The molecular landscape of premenopausal breast cancer**. *Breast Cancer Res* 2015, **17**:104.

92. Partridge AH, Pagani O, Abulkhair O, Aebi S, Amant F, Azim HA, Jr., Costa A, Delaloge S, Freilich G, Gentilini OD *et al*: **First international consensus guidelines for breast cancer in young women (BCY1)**. *Breast* 2014, **23**(3):209-220.

93. Bernhardt SM, Dasari P, Walsh D, Townsend AR, Price TJ, Ingman WV: **Hormonal Modulation of Breast Cancer Gene Expression: Implications for Intrinsic Subtyping in Premenopausal Women**. *Front Oncol* 2016, **6**:241.

94. Joerger M, Thurlimann B: **Chemotherapy regimens in early breast cancer: major controversies and future outlook**. *Expert Rev Anticancer Ther* 2013, **13**(2):165-178.

95. Need EF, Selth LA, Trotta AP, Leach DA, Giorgio L, O'Loughlin MA, Smith E, Gill PG, Ingman WV, Graham JD *et al*: **The unique transcriptional response produced by concurrent estrogen and progesterone treatment in breast cancer cells results in upregulation of growth factor pathways and switching from a Luminal A to a Basal-like subtype**. *BMC Cancer* 2015, **15**:791.

96. Ni H, Kumbrink J, Mayr D, Seiler A, Hagemann F, Degenhardt T, Sagebiel S, Wurstlein R, Kates R, Harbeck N *et al*: **Molecular Prognostic Factors for Distant Metastases in Premenopausal Patients with HR+/HER2- Early Breast Cancer**. *J Pers Med* 2021, **11**(9).

97. QIAGEN: **RNeasy FFPE Handbook**. 2014.

98. Decalf J, Albert ML, Ziai J: **New tools for pathology: a user's review of a highly multiplexed method for in situ analysis of protein and RNA expression in tissue**. *J Pathol* 2019, **247**(5):650-661.

99. nanoString: **nCounter® Breast Cancer 360™ Panel,** accessed Oct 1 2021, < https://nanostring.com/products/ncounter-assays-panels/oncology/breast-cancer-360/ >

100. NanoString: **MAN-10056-02 CodeSet Hybridization Setup**. 2018.

101. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I: **Controlling the false discovery rate in behavior genetics research**. *Behav Brain Res* 2001, **125**(1-2):279-284.

102. Chen X: **False discovery rate control for multiple testing based on discrete p-values**. *Biom J* 2020, **62**(4):1060-1079.

103. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic Acids Res* 2015, **43**(7):e47.

104. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci U S A* 2005, **102**(43):15545-15550.

105. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *OMICS* 2012, **16**(5):284-287.

106. Győrffy B: **Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer**. *Computational and Structural Biotechnology Journal* 2021, **19**:4101-4109.

107. Rugo HS: **The importance of distant metastases in hormone-sensitive breast cancer**. *The Breast* 2008, **17**:S3-S8.

108. Sestak I, Buus R, Cuzick J, Dubsky P, Kronenwett R, Denkert C, Ferree S, Sgroi D, Schnabel C, Baehner FL *et al*: **Comparison of the Performance of 6 Prognostic Signatures for Estrogen Receptor-Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial**. *JAMA Oncol* 2018, **4**(4):545-553.

109. Caan BJ, Sweeney C, Habel LA, Kwan ML, Kroenke CH, Weltzien EK, Quesenberry CP, Jr., Castillo A, Factor RE, Kushi LH *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* 2014, **23**(5):725-734.

110. Voogd AC, Nielsen M, Peterse JL, Blichert-Toft M, Bartelink H, Overgaard M, van Tienhoven G, Andersen KW, Sylvester RJ, van Dongen JA *et al*: **Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II breast cancer: pooled results of two large European randomized trials**. *J Clin Oncol* 2001, **19**(6):1688-1697.

111. Comen EA, Norton L, Massague J: **Breast cancer tumor size, nodal status, and prognosis: biology trumps anatomy**. *J Clin Oncol* 2011, **29**(19):2610-2612.

112. Foulkes WD, Reis-Filho JS, Narod SA: **Tumor size and survival in breast cancer--a reappraisal**. *Nat Rev Clin Oncol* 2010, **7**(6):348-353.

113. Wo JY, Chen K, Neville BA, Lin NU, Punglia RS: **Effect of very small tumor size on cancer-specific mortality in node-positive breast cancer**. *J Clin Oncol* 2011, **29**(19):2619-2627.

114. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, Zackrisson S, Senkus E, clinicalguidelines@esmo.org EGCEa: **Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-updagger**. *Ann Oncol* 2019, **30**(8):1194-1220.

115. **Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials**. *The Lancet* 2005, **365**(9472):1687-1717.

116. Francis PA, Pagani O, Fleming GF, Walley BA, Colleoni M, Lang I, Gomez HL, Tondini C, Ciruelos E, Burstein HJ *et al*: **Tailoring Adjuvant Endocrine Therapy for Premenopausal Breast Cancer**. *N Engl J Med* 2018, **379**(2):122-137.

117. Daemen A, Manning G: **HER2 is not a cancer subtype but rather a pan-cancer event and is highly enriched in AR-driven breast tumors**. *Breast Cancer Res* 2018, **20**(1):8.

118. Yamashita H, Ishida N, Hatanaka Y, Hagio K, Oshino T, Takeshita T, Kanno-Okada H, Shimizu AI, Hatanaka KC, Matsuno Y: **HER2 Gene Amplification in ER-positive HER2**

Immunohistochemistry 0 or 1+ Breast Cancer With Early Recurrence. *Anticancer Res* 2020, **40**(2):645-652.

119. Bhattacharjee A, Rajendra J, Dikshit R, Dutt S: **HER2 borderline is a negative prognostic factor for primary malignant breast cancer**. *Breast Cancer Res Treat* 2020, **181**(1):225-231.

120. Venet D, Dumont JE, Detours V: **Most random gene expression signatures are significantly associated with breast cancer outcome**. *PLoS Comput Biol* 2011, **7**(10):e1002240.

121. Kurozumi S, Matsumoto H, Hayashi Y, Tozuka K, Inoue K, Horiguchi J, Takeyoshi I, Oyama T, Kurosumi M: **Power of PgR expression as a prognostic factor for ER-positive/HER2-negative breast cancer patients at intermediate risk classified by the Ki67 labeling index**. *BMC Cancer* 2017, **17**(1):354.

122. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM: **Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer**. *Breast Cancer Res* 2010, **12**(5):R68.

123. Fougner C, Bergholtz H, Norum JH, Sorlie T: **Re-definition of claudin-low as a breast cancer phenotype**. *Nat Commun* 2020, **11**(1):1787.

124. Kotiyal S, Bhattacharya S: **Breast cancer stem cells, EMT and therapeutic targets**. *Biochem Biophys Res Commun* 2014, **453**(1):112-116.

125. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C *et al*: **Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap**. *Nat Protoc* 2019, **14**(2):482-517.

126. Cabanas Morafraile E, Perez-Pena J, Fuentes-Antras J, Manzano A, Perez-Segura P, Pandiella A, Galan-Moya EM, Ocana A: **Genomic Correlates of DNA Damage in Breast Cancer Subtypes**. *Cancers (Basel)* 2021, **13**(9).

127. Gnant M: **The role of mammalian target of rapamycin (mTOR) inhibition in the treatment of advanced breast cancer**. *Curr Oncol Rep* 2013, **15**(1):14-23.

128. Beck JT, Hortobagyi GN, Campone M, Lebrun F, Deleu I, Rugo HS, Pistilli B, Masuda N, Hart L, Melichar B *et al*: **Everolimus plus exemestane as first-line therapy in HR(+), HER2(-) advanced breast cancer in BOLERO-2**. *Breast Cancer Res Treat* 2014, **143**(3):459-467.

129. Laplante M, Sabatini DM: **mTOR signaling in growth control and disease**. *Cell* 2012, **149**(2):274-293.

130. Steelman LS, Martelli AM, Cocco L, Libra M, Nicoletti F, Abrams SL, McCubrey JA: **The therapeutic potential of mTOR inhibitors in breast cancer**. *Br J Clin Pharmacol* 2016, **82**(5):1189-1212.

131. Fan Y, Sun T, Shao Z, Zhang Q, Ouyang Q, Tong Z, Wang S, Luo Y, Teng Y, Wang X *et al*: **Effectiveness of Adding Everolimus to the First-line Treatment of Advanced Breast Cancer in Premenopausal Women Who Experienced Disease Progression While Receiving Selective Estrogen Receptor Modulators: A Phase 2 Randomized Clinical Trial**. *JAMA Oncol* 2021, **7**(10):e213428.

132. Moore HCF, Barlow WE, Somlo G, Gralow JR, Schott AF, Hayes DF, Kuhn P, Hicks JB, Welter L, Dy PA *et al*: **A Randomized Trial of Fulvestrant, Everolimus, and Anastrozole for the Front-line Treatment of Patients with Advanced Hormone Receptor-positive Breast Cancer, SWOG S1222**. *Clin Cancer Res* 2021.

133. Guarneri V, Giorgi CA, Cinieri S, Bengala C, Mariani G, Bisagni G, Frassoldati A, Zamagni C, De Rossi C, Amoroso V *et al*: **Everolimus plus aromatase inhibitors as maintenance therapy after first-line chemotherapy: Final results of the phase III randomised MAIN-A (MAINtenance Afinitor) trial**. *Eur J Cancer* 2021, **154**:21-29.

134. Treilleux I, Arnedos M, Cropet C, Wang Q, Ferrero JM, Abadie-Lacourtoisie S, Levy C, Legouffe E, Lortholary A, Pujade-Lauraine E *et al*: **Translational studies within the TAMRAD randomized GINECO trial: evidence for mTORC1 activation marker as a**

predictive factor for everolimus efficacy in advanced breast cancer. *Ann Oncol* 2015, **26**(1):120-125.

135. Wiederschain D, Chen L, Johnson B, Bettano K, Jackson D, Taraszka J, Wang YK, Jones MD, Morrissey M, Deeds J *et al*: **Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis**. *Mol Cell Biol* 2007, **27**(13):4968-4979.

136. Wang S, Xiong Y, Zhang Q, Su D, Yu C, Cao Y, Pan Y, Lu Q, Zuo Y, Yang L: **Clinical significance and immunogenomic landscape analyses of the immune cell signature based prognostic model for patients with breast cancer**. *Brief Bioinform* 2021, **22**(4).

137. Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, Loi S: **Clinical relevance of host immunity in breast cancer: from TILs to the clinic**. *Nat Rev Clin Oncol* 2016, **13**(4):228-241.

138. Marzolo MP, Farfan P: **New insights into the roles of megalin/LRP2 and the regulation of its functional expression**. *Biol Res* 2011, **44**(1):89-105.

139. Nykjaer A, Dragun D, Walther D, Vorum H, Jacobsen C, Herz J, Melsen F, Christensen EI, Willnow TE: **An endocytic pathway essential for renal uptake and activation of the steroid 25-(OH) vitamin D3**. *Cell* 1999, **96**(4):507-515.

140. Gonias SL, Karimi-Mostowfi N, Murray SS, Mantuano E, Gilder AS: **Expression of LDL receptor-related proteins (LRPs) in common solid malignancies correlates with patient survival**. *PLoS One* 2017, **12**(10):e0186649.

141. de La Puente-Yague M, Cuadrado-Cenzual MA, Ciudad-Cabanas MJ, Hernandez-Cabria M, Collado-Yurrita L: **Vitamin D: And its role in breast cancer**. *Kaohsiung J Med Sci* 2018, **34**(8):423-427.

142. Hossain S, Beydoun MA, Beydoun HA, Chen X, Zonderman AB, Wood RJ: **Vitamin D and breast cancer: A systematic review and meta-analysis of observational studies**. *Clin Nutr ESPEN* 2019, **30**:170-184.

143. Welsh J: **Function of the vitamin D endocrine system in mammary gland and breast cancer**. *Mol Cell Endocrinol* 2017, **453**:88-95.

144. Chlon TM, Taffany DA, Welsh J, Rowling MJ: **Retinoids modulate expression of the endocytic partners megalin, cubilin, and disabled-2 and uptake of vitamin D-binding protein in human mammary cells**. *J Nutr* 2008, **138**(7):1323-1328.

145. Anderson LN, Cotterchio M, Cole DE, Knight JA: **Vitamin D-related genetic variants, interactions with vitamin D exposure, and breast cancer risk among Caucasian women in Ontario**. *Cancer Epidemiol Biomarkers Prev* 2011, **20**(8):1708-1717.

146. Christ A, Herzog K, Willnow TE: **LRP2, an auxiliary receptor that controls sonic hedgehog signaling in development and disease**. *Dev Dyn* 2016, **245**(5):569-579.

147. Carballo GB, Honorato JR, de Lopes GPF, Spohr T: **A highlight on Sonic hedgehog pathway**. *Cell Commun Signal* 2018, **16**(1):11.

148. Tan Y, Wang M, Yang K, Chi T, Liao Z, Wei P: **PPAR-alpha Modulators as Current and Potential Cancer Treatments**. *Front Oncol* 2021, **11**:599995.

149. Oh SJ, Kim TH, Lim JM, Jeong JW: **Progesterone induces expression of Lrp2 in the murine uterus**. *Biochem Biophys Res Commun* 2013, **441**(1):175-179.

150. Flamini MI, Gauna GV, Sottile ML, Nadin BS, Sanchez AM, Vargas-Roig LM: **Retinoic acid reduces migration of human breast cancer cells: role of retinoic acid receptor beta**. *J Cell Mol Med* 2014, **18**(6):1113-1123.

151. Du N, Feng J, Hu LJ, Sun X, Sun HB, Zhao Y, Yang YP, Ren H: **Angiotensin II receptor type 1 blockers suppress the cell proliferation effects of angiotensin II in breast cancer cells by inhibiting AT1R signaling**. *Oncol Rep* 2012, **27**(6):1893-1903.

152. Rhodes DR, Ateeq B, Cao Q, Tomlins SA, Mehra R, Laxman B, Kalyana-Sundaram S, Lonigro RJ, Helgeson BE, Bhojani MS *et al*: **AGTR1 overexpression defines a subset of breast cancer and confers sensitivity to losartan, an AGTR1 antagonist**. *Proc Natl Acad Sci U S A* 2009, **106**(25):10284-10289.

153. Cheng CJ, Lin YC, Tsai MT, Chen CS, Hsieh MC, Chen CL, Yang RB: **SCUBE2 suppresses breast tumor cell proliferation and confers a favorable prognosis in invasive breast cancer**. *Cancer Res* 2009, **69**(8):3634-3641.

154. Lin YC, Chen CC, Cheng CJ, Yang RB: **Domain and functional analysis of a novel breast tumor suppressor protein, SCUBE2**. *J Biol Chem* 2011, **286**(30):27039-27047.

155. Esmaeili R, Mohammadi S, Jafarbeik-Iravani N, Yadegari F, Olfatbakhsh A, Mazaheri M, Kaviani A, Rezaee M, Majidzadeh AK: **Expression of SCUBE2 and BCL2 Predicts Favorable Response in ERalpha Positive Breast Cancer**. *Arch Iran Med* 2021, **24**(3):209-217.

156. Vieira AF, Schmitt F: **An Update on Breast Cancer Multigene Prognostic Tests-Emergent Clinical Biomarkers**. *Front Med (Lausanne)* 2018, **5**:248.

157. Lin YC, Lee YC, Li LH, Cheng CJ, Yang RB: **Tumor suppressor SCUBE2 inhibits breast-cancer cell migration and invasion through the reversal of epithelial-mesenchymal transition**. *J Cell Sci* 2014, **127**(Pt 1):85-100.

158. Qi X, Li X: **Mechanistic Insights into the Generation and Transduction of Hedgehog Signaling**. *Trends Biochem Sci* 2020, **45**(5):397-410.

159. Klopocki E, Kristiansen G, Wild PJ, Klaman I, Castanos-Velez E, Singer G, Stohr R, Simon R, Sauter G, Leibiger H *et al*: **Loss of SFRP1 is associated with breast cancer progression and poor prognosis in early stage tumors**. *Int J Oncol* 2004, **25**(3):641-649.

160. Veeck J, Niederacher D, An H, Klopocki E, Wiesmann F, Betz B, Galm O, Camara O, Durst M, Kristiansen G *et al*: **Aberrant methylation of the Wnt antagonist SFRP1 in breast cancer is associated with unfavourable prognosis**. *Oncogene* 2006, **25**(24):3479-3488.

161. He J, Wu M, Xiong L, Gong Y, Yu R, Peng W, Li L, Li L, Tian S, Wang Y *et al*: **BTB/POZ zinc finger protein ZBTB16 inhibits breast cancer proliferation and metastasis through upregulating ZBTB28 and antagonizing BCL6/ZBTB27**. *Clin Epigenetics* 2020, **12**(1):82.

162. Lee NJ, Ali N, Zhang L, Qi Y, Clarke I, Enriquez RF, Brzozowska M, Lee IC, Rogers MJ, Laybutt DR *et al*: **Osteoglycin, a novel coordinator of bone and glucose homeostasis**. *Mol Metab* 2018, **13**:30-44.

163. Xu T, Zhang R, Dong M, Zhang Z, Li H, Zhan C, Li X: **Osteoglycin (OGN) Inhibits Cell Proliferation and Invasiveness in Breast Cancer via PI3K/Akt/mTOR Signaling Pathway**. *Onco Targets Ther* 2019, **12**:10639-10650.

164. Semmlinger A, von Schoenfeldt V, Wolf V, Meuter A, Kolben TM, Kolben T, Zeder-Goess C, Weis F, Gallwas J, Wuerstlein R *et al*: **EP3 (prostaglandin E2 receptor 3) expression is a prognostic factor for progression-free and overall survival in sporadic breast cancer**. *BMC Cancer* 2018, **18**(1):431.

165. Huang Z, Duan H, Li H: **Identification of Gene Expression Pattern Related to Breast Cancer Survival Using Integrated TCGA Datasets and Genomic Tools**. *Biomed Res Int* 2015, **2015**:878546.

166. Klimczak M, Biecek P, Zylicz A, Zylicz M: **Heat shock proteins create a signature to predict the clinical outcome in breast cancer**. *Sci Rep* 2019, **9**(1):7507.

167. Segers VFM, Dugaucquier L, Feyen E, Shakeri H, De Keulenaer GW: **The role of ErbB4 in cancer**. *Cell Oncol (Dordr)* 2020, **43**(3):335-352.

168. Jung IH, Jung DE, Chung YY, Kim KS, Park SW: **Iroquois Homeobox 1 Acts as a True Tumor Suppressor in Multiple Organs by Regulating Cell Cycle Progression**. *Neoplasia* 2019, **21**(10):1003-1014.

169. Guo X, Liu W, Pan Y, Ni P, Ji J, Guo L, Zhang J, Wu J, Jiang J, Chen X *et al*: **Homeobox gene IRX1 is a tumor suppressor gene in gastric carcinoma**. *Oncogene* 2010, **29**(27):3908-3920.

170. He B, Chen J, Song W, Bai Y: **miR-646/TET1 mediated demethylation of IRX1 promoter upregulates HIST2H2BE and promotes the progression of invasive ductal carcinoma**. *Genomics* 2021, **113**(3):1469-1481.

171. Deshmukh A, Kumar S, Arfuso F, Newsholme P, Dharmarajan A: **Secreted Frizzled-related protein 4 (sFRP4) chemo-sensitizes cancer stem cells derived from human breast, prostate, and ovary tumor cell lines**. *Sci Rep* 2017, **7**(1):2256.

172. Mandal S, Gamit N, Varier L, Dharmarajan A, Warrier S: **Inhibition of breast cancer stem-like cells by a triterpenoid, ursolic acid, via activation of Wnt antagonist, sFRP4 and suppression of miRNA-499a-5p**. *Life Sci* 2021, **265**:118854.

173. Cui Y, Liang S, Zhang S, Zhang C, Zhao Y, Wu D, Wang J, Song R, Wang J, Yin D *et al*: **ABCA8 is regulated by miR-374b-5p and inhibits proliferation and metastasis of hepatocellular carcinoma through the ERK/ZEB1 pathway**. *J Exp Clin Cancer Res* 2020, **39**(1):90.

174. Chen L, Ye C, Huang Z, Li X, Yao G, Liu M, Hu X, Dong J, Guo Z: **[Differentially expressed genes and potential signaling pathway in Asian people with breast cancer by preliminary analysis of a large sample of the microarray data]**. *Nan Fang Yi Ke Da Xue Xue Bao* 2014, **34**(6):807-812.

175. Paulsen MT, Starks AM, Derheimer FA, Hanasoge S, Li L, Dixon JE, Ljungman M: **The p53-targeting human phosphatase hCdc14A interacts with the Cdk1/cyclin B complex and is differentially expressed in human cancers**. *Mol Cancer* 2006, **5**:25.

176. Chen NP, Uddin B, Hardt R, Ding W, Panic M, Lucibello I, Kammerer P, Ruppert T, Schiebel E: **Human phosphatase CDC14A regulates actin organization through dephosphorylation of epithelial protein lost in neoplasm**. *Proc Natl Acad Sci U S A* 2017, **114**(20):5201-5206.

177. Gordon JA, Sodek J, Hunter GK, Goldberg HA: **Bone sialoprotein stimulates focal adhesion-related signaling pathways: role in migration and survival of breast and prostate cancer cells**. *J Cell Biochem* 2009, **107**(6):1118-1128.

178. Zhang Y, He W, Zhang S: **Seeking for Correlative Genes and Signaling Pathways With Bone Metastasis From Breast Cancer by Integrated Analysis**. *Front Oncol* 2019, **9**:138.

179. Wu K, Feng J, Lyu F, Xing F, Sharma S, Liu Y, Wu SY, Zhao D, Tyagi A, Deshpande RP *et al*: **Exosomal miR-19a and IBSP cooperate to induce osteolytic bone metastasis of estrogen receptor-positive breast cancer**. *Nat Commun* 2021, **12**(1):5196.

180. Beardsley DI, Kowbel D, Lataxes TA, Mannino JM, Xin H, Kim W-J, Collins C, Brown KD: **Characterization of the novel amplified in breast cancer-1 (NABC1) gene product**. *Experimental Cell Research* 2003, **290**(2):402-413.

181. Chen F, Zhang Z, Pu F: **Role of stanniocalcin-1 in breast cancer**. *Oncol Lett* 2019, **18**(4):3946-3953.

182. Chang AC, Doherty J, Huschtscha LI, Redvers R, Restall C, Reddel RR, Anderson RL: **STC1 expression is associated with tumor growth and metastasis in breast cancer**. *Clin Exp Metastasis* 2015, **32**(1):15-27.

183. Brantley KD, Kjaersgaard A, Cronin-Fenton D, Yacoub R, Nielsen AS, Lauridsen KL, Hamilton-Dutoit S, Lash TL: **Stanniocalcin Expression as a Predictor of Late Breast Cancer Recurrence**. *Cancer Epidemiol Biomarkers Prev* 2018, **27**(6):653-659.

184. Bosserhoff AK, Buettner R: **Expression, function and clinical relevance of MIA (melanoma inhibitory activity)**. *Histol Histopathol* 2002, **17**(1):289-300.

185. Schmidt J, Riechers A, Bosserhoff AK: **MIA--a new target protein for malignant melanoma therapy**. *Histol Histopathol* 2013, **28**(4):421-426.

186. Gu QH, Li D, Xie ZH, Shen QB: **The clinical significance of MIA gene in tumorigenesis of lung cancer**. *Neoplasma* 2020, **67**(3):660-667.

187. Rigiracciolo DC, Nohata N, Lappano R, Cirillo F, Talia M, Scordamaglia D, Gutkind JS, Maggiolini M: **IGF-1/IGF-1R/FAK/YAP Transduction Signaling Prompts Growth Effects in Triple-Negative Breast Cancer (TNBC) Cells**. *Cells* 2020, **9**(4).

188. Al-Delaimy WK, Flatt SW, Natarajan L, Laughlin GA, Rock CL, Gold EB, Caan BJ, Parker BA, Pierce JP: **IGF1 and risk of additional breast cancer in the WHEL study**. *Endocr Relat Cancer* 2011, **18**(2):235-244.

189. Cimino-Mathews A: **Novel uses of immunohistochemistry in breast pathology: interpretation and pitfalls**. *Mod Pathol* 2021, **34**(Suppl 1):62-77.

190. Kim CY, Kim YC, Oh JH, Kim MH: **HOXA5 confers tamoxifen resistance via the PI3K/AKT signaling pathway in ER-positive breast cancer**. *J Cancer* 2021, **12**(15):4626-4637.

191. Teo WW, Merino VF, Cho S, Korangath P, Liang X, Wu RC, Neumann NM, Ewald AJ, Sukumar S: **HOXA5 determines cell fate transition and impedes tumor initiation and progression in breast cancer through regulation of E-cadherin and CD24**. *Oncogene* 2016, **35**(42):5539-5551.

192. Peng X, Zha L, Chen A, Wang Z: **HOXA5 is a tumor suppressor gene that is decreased in gastric cancer**. *Oncol Rep* 2018, **40**(3):1317-1329.

193. Uysal-Onganer P, Kypta RM: **Wnt11 in 2011 - the regulation and function of a non-canonical Wnt**. *Acta Physiol (Oxf)* 2012, **204**(1):52-64.

194. McCart Reed AE, Song S, Kutasovic JR, Reid LE, Valle JM, Vargas AC, Smart CE, Simpson PT: **Thrombospondin-4 expression is activated during the stromal response to invasive breast cancer**. *Virchows Arch* 2013, **463**(4):535-545.

Apendix A: Key codes used in R analysis

Packages	BiocManager::install('CLL') install.packages('corrplot') install.packages('gpairs') install.packages('vioplot') install.packages('ggplot2') install.packages('glue') library(CLL) library(corrplot) library(gpairs) library(vioplot) library(ggplot2)
DEG	logFC_cutoff <- with(DEG,mean(abs(logFC)) + 2*sd(abs(logFC))) DEG$change = as.factor(ifelse(DEG$P.Value < 0.05 & abs(DEG$logFC) > logFC_cutoff, ifelse(DEG$logFC > logFC_cutoff ,'UP','DOWN'),'NOT') this_tile <- paste0('Cutoff for logFC is ',round(logFC_cutoff,3), '\nThe number of up gene is ',nrow(DEG[DEG$change =='UP',]) , '\nThe number of down gene is ',nrow(DEG[DEG$change =='DOWN',])) nrDEG <- DEG rm(DEG) DEG= nrDEG[which(abs(nrDEG$logFC)> logFC_cutoff),] DEG= DEG[which(DEG$P.Value<0.05),] gene<- DEG$SYMBOL write.csv(DEG,"D:\\.csv")
Volcano plot	g = ggplot(data=nrDEG, aes(x=logFC, y=-log10(P.Value), color=change)) + geom_point(alpha=0.4, size=3.5, aes(color=change)) + ggtitle(this_tile) + theme(plot.title = element_text(size=15,hjust = 0.5))+ scale_color_manual(values=c("blue", "grey","red"))+ geom_vline(xintercept=c(-0.578,0.583),lty=4,col="black",lwd=0.8)

	+ geom_hline(yintercept = -log10(0.05),lty=4,col="black",lwd=0.8) + theme_bw() ## corresponding to the levels(res$change) print(g) g + geom_point(size = 3, shape = 1, data =DEG) + ggrepel::geom_label_repel(aes(label = SYMBOL), data = DEG, color="black")
Pheatmap	library(pheatmap) sample_info <-`85.patients.subtype.group` annotation_col = sample_info a1 <- `11.survival.significant.DEGs` rownames(annotation_col) = colnames(a1) pheatmap(a1, annotation_col = annotation_col)

Survival curves	install.packages(c("survival", "survminer"))
	library("survival")
	library("survminer")
	fit <- survfit(Surv(month, status) ~ ROR.score, data = ROR.score.85.patients.3)
	print(fit)
	install.packages('markdown')
	install.packages('G Gally')
	ggsurvplot(fit,
	pval = FALSE, conf.int = TRUE,
	risk.table = TRUE, # Add risk table
	risk.table.col = "strata", # Change risk table color by groups
	linetype = "strata", # Change line type by groups
	ggtheme = theme_bw(), # Change ggplot2 theme
	palette = c("#E7B800", "#2E9FDF"))

Acknowledgements

Three years in Munich is like a vivid dream for me, which is full of excitement and adventures. When I look back at this journey, I feel so lucky that I had such an opportunity to pursue my dream major at LMU. Although there were difficult times, I received immeasurable support from professors, colleagues, and friends to help me through. Therefore, I must present my sincere gratitude here.

I would like to first thank my supervisor, Prof. Nadia Harbeck. I thank her for accepting me as her MD student, for giving me such a precious opportunity to study breast cancer, for helping me at her best, for always being friendly, patient, and supportive. The scientific training is not the only thing that improved me, her personality, intelligence, and attitudes regarding scientific studies had a substantial impact on me. And I believe the most precious gift from Prof. Harbeck is that she led me to me a direction that I am fully interested in and hope to pursue as my future career. Here, I present my best wishes to her, and will always keep the sincerest gratitude in my heart.

Then, I would like to thank Ms. Hoffmann. I thank her for always being so helpful, friendly, and responsive. From the application till submitting the dissertation, she offered me countless help. She is like a bridge; I would not be able to arrive at the destination without her help. I would also like to thank Dr. Eggersmann and Dr. Kates for offering me professional and unconditional support. Besides, I much present my sincere thanks to Ms. Kurt. She is a strong, warm-hearted, and selfless woman. Although she won't be able to read the sentences, I will never forget about her.

Moreover, I sincerely thank the professors and colleagues that supported me through the challenges in the lab work. I would like to thank Prof. Mayr and Dr. Kumbrink for helping me throughout my lab works so efficiently and professionally. They showed me how to perform trustworthy scientific research, and I will always follow their guidance. I thank Prof. Jung for offering support in his lab and allowing me to join his group seminars for learning and discussions. I also must thank Ms. Sagebiel for being so professional and friendly while supporting my daily experiments. Besides, I thank all the colleagues in the pathology department that helped me with sample collection and sample preparation. Although I spent limited time in the clinics, I learned a lot. So, I would like to present my thanks to all the doctors and nurses that offered me opportunities to observe in the breast center, showed me the processes of daily patient care, and explained to me about their professions. All these experiences and memories will be an invaluable treasure to me.

Being admitted to LMU was the beginning of this journey. So, in the end, I would like to thank Prof. Baethmann, who recommended me to Prof. Harbeck. I know it was not his responsibility to help with the international students, but he helped us in such a perfect and responsible way. I believe his generous help definitely has impacted a lot of students, and I will always keep the sincere gratitude in my heart. Besides, I would also like to thank Dr. Zhang, the international office of LMU, the LMU-CSC (China Scholarship

Council) program, and the MMRS (Munich medical research school) office for all the support to help me accomplish my study.

Personally, I must thank the people who supported my daily life in Munich. Without them, it would be much harder for me to accomplish my study. First, I present my sincere thanks to KinderHaus Medienfabric for taking care of my son in such a devoted way. During the pandemic, everything was so complicated, but they tried their best to take care of the children in a responsible and stable way. It was their dedication that guaranteed my work time. Then, I would like to thank Ms. Rank for offering me lovely and secure accommodation in Munich, presenting me with the local life specialties, and helping me unconditionally. Moreover, let me thank all my precious friends in Munich: I thank Guyuan, Liangwei, Penglin, Aimee, Samuel, Liwei, Panpan, Lipan, Hahna, Ina, Veronika, Wangmeng, and Chunfang for offering me precious help and accompany.

I am glad I was not traveling alone on this journey. For this, I must thank my little boy, Bangyan, who brought me so much happiness that I could not imagine. It was because of him that I had a chance to discover Munich like a small child: We visited nearly every park and attraction in Munich, read a lot of German children's books and participated in the local festivals throughout the year. We had so much fun together in this lovely city, and I am glad he could share all the precious memories with me. Besides, I thank my husband, family, and friends for supporting me from distance. Although we were not together physically, I always got strength from their unconditional support.

Vielen Dank und alles Gute!

Affidavit

Affidavit

Ni, Hua

Surname, first name

Hans-Fischer Str.29

Street

80339, Munich, Germany

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Evaluation of molecular prognostic markers for premenopausal patients with HR+/HER2− early breast cancer

…….

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Beijing, China, 03.11.2022

place, date

Hua Ni

Signature doctoral candidate

List of publications

Regarding the dissertation:

1. H.Ni, J.Kumbrink, D.Mayr, N.Harbeck, T.Eggersmann, etc. Molecular Prognostic Factors for Distant Metastases in Premenopausal Patients with HR+/HER2− Early Breast Cancer. Journal of Personalized Medicine. 2021, 11, 835. [**2021** IF: 4.945, JCR Q1]

2. H.Ni, J.Kumbrink, D.Mayr, N.Harbeck, T.Eggersmann, etc. Molecular risk factors for distant metastases in premenopausal patients with HR+/HER2- EBC[J][**Abstract**], Annals of Oncology, 2021, 32: S44.

3. H.Ni, A.Kurt, J.Kumbrink, D.Mayr, N.Harbeck, T.Eggersmann, etc. Gene expression profiles in premenopausal women with HR+ HER2− early breast cancer[J] [**Abstract**], European Journal of Cancer, 2020, 138, S76.

Other projects:

4. H.Ni, L.Niu, H.Li, etc. Long non-coding RNA LINC00152 is up-regulated in ovarian cancer tissues and regulates proliferation and cell cycle of SKOV3 cells[J], Eur Rev Med Pharmacol Sci, 2019,23(22): 9803-9813.

5. L.Niu, H.Ni (co-first),H.Li, etc. miR-509-3p enhances platinum drug sensitivity in ovarian cancer[J], Gene, 2019,686: 63-67.

6. K.Dong, H.Ni, D.Shi, etc. ROS-mediated glucose metabolic reprogram induces insulin resistance in type 2 diabetes[J], Biochemical and Biophysical Research Communications, 2016, 476(4): 204-211.

7. H.Ni, L.Niu, H.Li, a, etc. Long non-coding RNA NR2F2-AS1 relates to occurrence and development of epithelial ovarian cancer[J], TUMOR, 2018(38), 581-589. [**in Chinese**]

8. H.Ni, L.Niu, H.Li.Research progress on long Non-coding RNA relating to epithelial ovarian cancer[J], J Int Obstet Gynecol, 2017, 44(3) :284-287. [**in Chinese**]

9. H.Ni, L.Niu, H.Li, etc. A clinical retrospective analysis of five cases of primary cervical small cell cancer[J], Chin J Min Inv Surg, 2017, 17(11): 1007-1010. [**in Chinese**]

10. H.Ni, L.Niu, H.Li, etc. Primary Small cell carcinomas of the gynecologic tract: A clinical retrospective analysis of eight cases[J], Chin J Cancer Prev Treat, 2016, 23(3): 194-197. [**in Chinese**]

Your contribution to the publication

Part of the dissertation has been published in “Ni et al., Molecular Prognostic Factors for Distant Metastases in Premenopausal Patients with HR+/HER2- Early Breast Cancer. J Pers Med 2021, 11(*9)”.* According to the regulation of the publisher, the authors retain the copyright and this article was licensed under an open access Creative Commons CC BY 4.0 license. The permission from the publisher is submitted with this dissertation.

For this publication, as the first author, I have completed the following works: sorting patients’ information, collecting samples, validating the histology of the samples, marking the tumor region, extracting RNA from the samples, running the Nanostring test, preparing the raw data, analyzing data, preparing tables and figures, and drafting the original manuscript.