High-throughput molecular techniques and personalized treatment for primary breast cancer

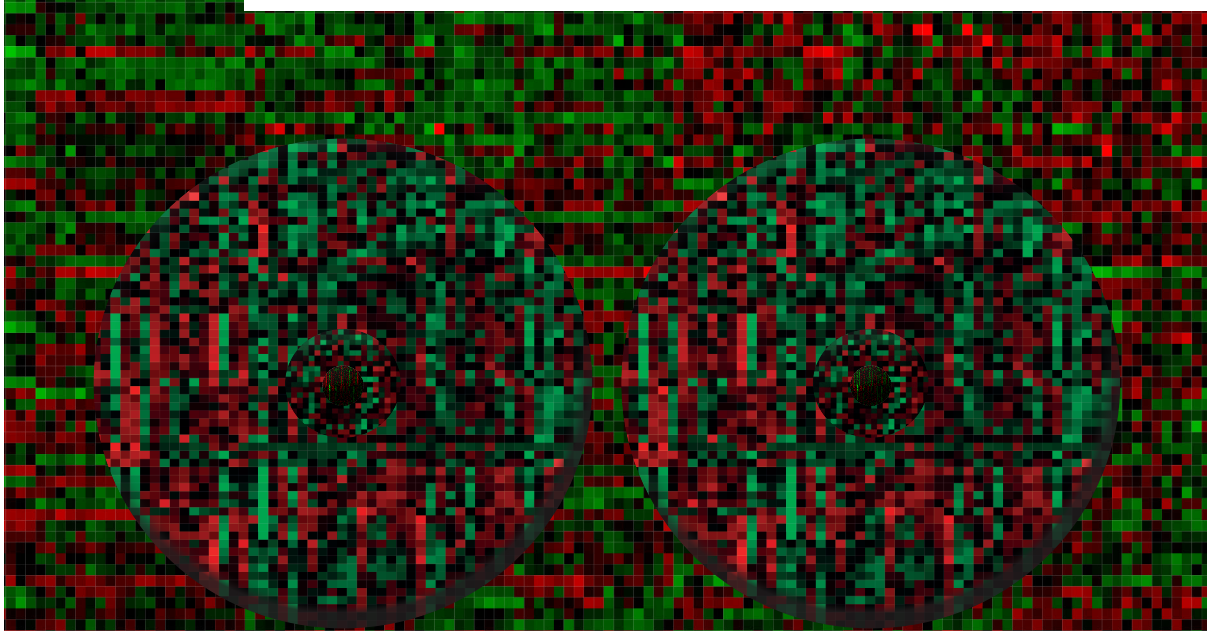
Sjöström, Martin

2018

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

High-throughput molecular techniques and personalized treatment for primary breast cancer

MARTIN SJÖSTRÖM
DIVISION OF ONCOLOGY AND PATHOLOGY | DIVISION OF SURGERY | LUND UNIVERSITY



High-throughput molecular techniques and personalized treatment for primary breast cancer

High-throughput molecular techniques and personalized treatment for primary breast cancer

Martin Sjöström
Faculty of Medicine
Division of Oncology and Pathology, Division of Surgery
Department of Clinical Sciences, Lund

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be publicly defended at the lecture hall in the Radiotherapy building,
Klinikgatan 5, Skåne University Hospital, Lund, Sweden.
January 26th, 2018, 9.00 am.

Faculty opponent
Professor Dr. John W.M. Martens, Department of Medical Oncology,
Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Organization:
LUND UNIVERSITY
Faculty of Medicine
Department of Clinical Sciences Lund
Division of Oncology and Pathology
Division of Surgery

Author: Martin Sjöström

Document name:
DOCTORAL DISSERTATION

Date of issue:
2018-01-26

Sponsoring organization

Title and subtitle: High-throughput molecular techniques and personalized treatment for primary breast cancer

Abstract: Breast cancer is the most common cancer among women worldwide. The heterogeneity between tumors is the basis for precision medicine. Better characterization of tumors, leading to clinical decision tools, is needed to further individualize therapy, and to avoid over-treatment and under-treatment. The work constituting this thesis is focused on personalized treatment of primary breast cancer using molecular high-throughput technologies, and consists of five studies.

In study I, we investigated the role of a new putative estrogen receptor, GPR30, in predicting the response to adjuvant endocrine therapy. We showed that endocrine therapy response may be independent of GPR30 expression. On the other hand, the lack of GPR30 expression in the plasma membrane of the cancer cells identified a long-term low-risk group of patients, suggesting a functional change in GPR30 signaling.

In study II, we further investigated the functional role of GPR30. We found that it is constitutively active and appears to be pro-apoptotic in its signaling, as well as prognostic for breast cancer outcome.

In study III, we evaluated the response to adjuvant radiotherapy after breast-conserving surgery in different breast cancer subtypes. Overall, the subtypes did not appear to be treatment predictive, but the HER2 amplified/over-expressed subtype had the lowest effect of postoperative radiotherapy. We further showed that a presumed low-risk group of patients, similar to patients currently enrolled in clinical trials of de-escalation, have a very good effect of adjuvant radiotherapy on ipsilateral breast tumor recurrences (IBTR), without systemic adjuvant treatment.

In study IV, we used gene expression analysis to find tumors that are responsive, or not responsive, to adjuvant radiotherapy. We created a targeted radiosensitivity gene panel, and Single Sample Predictors that were prognostic for IBTR. Combined, they showed promise in stratifying patients for treatment. The correlation of the classifiers with proliferation and immune response could explain the biology behind the models, and may also explain why ours and other classifiers of radiosensitivity perform differently in subgroups of breast cancer.

In study V, we used mass spectrometry-based proteomics to search for protein biomarkers of risk of distant recurrence. We created a method of combing shotgun non-targeted discovery mass spectrometry for candidate discovery, and targeted selected reaction monitoring (SRM) mass spectrometry for candidate validation. The method was applied to a cohort of breast tumors first enriched for N-glycopeptides. The workflow was established and a set of 5 candidate protein biomarkers were discovered, which were further orthogonally validated at the gene expression level.

In conclusion, the work in this thesis represents a small step towards better personalized treatment of primary breast cancer.

Key words: Breast cancer, treatment, high-throughput, gene expression, mass spectrometry, GPR30, GPER

Classification system and/or index terms (if any)

Supplementary bibliographical information | Language: English

ISSN and key title 1652-8220 Lund University, Faculty of Medicine Doctoral Dissertation Series 2018:8 | ISBN 978-91-7619-575-8

Recipient's notes | Number of pages | Price

Security classification

Signature  Date 2017-12-12

High-throughput molecular techniques and personalized treatment for primary breast cancer

Martin Sjöström
Faculty of Medicine
Division of Oncology and Pathology, Division of Surgery
Department of Clinical Sciences, Lund

LUND
UNIVERSITY

Supervisor: Dr. Emma Niméus, M.D., Ph.D.
Faculty of Medicine
Division of Surgery, Division of Oncology and Pathology
Department of Clinical Sciences, Lund

Co-supervisor: Professor Mårten Fernö, Ph.D.
Faculty of Medicine
Division of Oncology and Pathology,
Department of Clinical Sciences, Lund

Coverphoto by Martin Sjöström, arrangement by Jonas Palm. Heatmap of gene expression results.

Quod optimus medicus sit quoque philosophus – The best physician is also a philosopher. Title on work by Galenos, originally in Greek, 2nd century AD.

Faculty of Medicine
Department of Clinical Sciences Lund
Division of Oncology and Pathology
Division of Surgery

ISBN 978-91-7619-575-8
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2017

MADE IN SWEDEN

To my grandfather, and all cancer victims

Quod optimus medicus sit quoque philosophus

Contents

Populärvetenskaplig sammanfattning

Bröstcancer är den vanligaste cancerformen bland kvinnor och drabbar årligen omkring 8000 personer i Sverige. Prognosen har stadigt förbättras, framför allt tack vare bättre behandlingar, men fortfarande dör över 1500 kvinnor varje år på grund av sin sjukdom. Bröstcancer är inte en enda sorts sjukdom, utan består av många olika varianter där de individuella förändringarna i varje tumör styr hur aggressiv cancern är och vilka behandlingar som är lämpliga. En stor del av bröstcancerforskningen idag går ut på att karaktärisera olika typer av bröstcancer och ta reda på vilka behandlingar som passar bäst för varje enskild patient. Detta görs både för att öka chanserna till bot, och för att undvika att ge behandlingar som är verkningslösa, vilka istället enbart ger biverkningar och onödiga kostnader.

Grundbehandlingen av bröstcancer är att operera bort tumören, och därefter ges tilläggsbehandlingar för att förhindra återfall. Tilläggsbehandlingarna verkar genom att ta bort eventuella tumörceller som finns kvar i bröstet, eller redan har spridit sig i kroppen utan att märkas. Vanligen lägger man till strålbehandling och ytterligare hormonbehandling och målriktade läkemedel, beroende på tumörens egenskaper. Tidigare har man eftersträvat att operera bort tumören med stor marginal och omfattande operationer, men med modern tilläggsbehandling räcker det oftast med att operera bort en del av bröstet.

Arbetet i den här avhandlingen har som mål att använda nya molekylära metoder för att arbeta vidare med individanpassning av tilläggsbehandlingen, och består av av fem delstudier.

I den första delstudien undersökte vi en ny östrogenreceptor, GPR30, och vilken roll den har för behandlingsresistens mot hormonbehandling med antiöstrogenet tamoxifen. Trots tecken på att GPR30 medverkar till resistens i försök med cellinjer, ser vi inget samband mellan mängden receptorer i tumörerna hos patienter, och hur de svarar på tamoxifenbehandling. Däremot har de patienter som har mycket GPR30 specifikt på ytan av cancercellerna en sämre prognos, vilken kan tyda på att GPR30 bidrar till en aggressiv tumör.

I den andra delstudien fortsatte vi arbetet med GPR30 genom att undersöka hur receptorn fungerar. Vi upptäckte att den verkar kunna signalera utan att stimuleras av östrogen, och att den har en bromsande effekt på tumörceller. För mycket receptorer i cellen leder till s.k. programmerad celldöd (apoptos). Både i den första och andra delstudien fann vi att både de minst och de mest aggressiva tumörerna har mest av GPR30, vilket ytterligare talar för att något under cancerutvecklingen gör att GPR30 ändrar sin funktion, lokalisation i cellen och potentiellt bidrar till en aggressivare sjukdom. Det skulle kunna innebära att GPR30 är intressant som måltavla för riktad bröstcancerbehandling.

I delstudie tre undersökte vi hur de olika etablerade undergrupperna av bröstcancer svarar på strålbehandling som tilläggsbehandling. Idag ger man behandlingen till de flesta patienterna eftersom man inte vet vem som har nytta av den. Undantaget är en förmodad lågriskgrupp där man bedömer att den absoluta nyttan är så liten att nackdelarna med biverkningar är större. Vi samlade in tumörer från en tidigare studie där man randomiserade patienter till att få, eller inte få, strålbehandling. Vi kunde inte se någon skillnad i effekt av strålbehandling mellan olika undergrupper i frånvaro av systemisk och/eller målriktad behandling, förutom att tumörer som överuttrycker tillväxtfaktorreceptorn HER2 möjligen har sämre nytta av strålbehandling. Vidare undersökte vi den förmodade lågriskgruppen, som visade sig ha en mycket god effekt av strålbehandling avseende lokala återfall. Det väcker frågan om man kan använda strålbehandling istället för andra tilläggsbehandlingar.

I delstudie fyra fortsatte vi arbetet med att hitta sätt att individanpassa strålbehandlingen. Vi samlade in tumörer från patienter i Lund, Stockholm och Uppsala och analyserade dem med s.k. genexpressionanalys, i vilken man analyserar vilka gener som är aktiva eller avstängda. Vi valde ut gener till en riktad analyspanel, som även kan analysera tumörprover med sämre kvalitet, vilket är vanligt i klinisk rutin. Utifrån genuttrycken som vi mätte med den riktade panelen utvecklade vi en statistisk modell för att förutsäga risken för lokala återfall. Genom att kombinera modellerna med och utan strålbehandling kunde patienterna delas in i grupper som svarade bra eller dåligt på strålbehandlingen. Ytterligare analyser av biologin bakom modellerna kunde förklara varför våra modeller, och de som tidigare publicerats, fungerar olika bra i olika typer av bröstcancer.

I delstudie fem använde vi två typer av masspektrometri för att analysera glykosylerade proteiner hos brösttumörer som har hög eller låg risk för att få fjärråterfall trots hormonell behandling. En metod för att analysera det globala proteinuttrycket användes för att välja proteiner till en riktad panel. Tumörerna återanalyserades med den riktade panel och 9 proteiner var lovande som biomarkörer för risk att utveckla fjärråterfall, varav 5 kunde valideras i en separat kohort av patienter på genexpressionsnivå.

Sammanfattningsvis är den här avhandlingen ett litet steg på vägen mot en förbättrad skräddarsydd behandling av bröstcancer.

Abbreviations

AI	aromatase inhibitor
AUC	area under the curve
BCD	breast cancer death
CI	confidence interval
DDA	data-dependent acquisition
DDFS	distant disease-free survival
DIA	data-independent acquisition
EBCTCG	Early Breast Cancer Trialists' Collaborative Group
ER	estrogen receptor alpha
GPER	G protein-coupled estrogen receptor
GPR30	G protein-coupled receptor 30
HER2	human epidermal growth factor receptor 2
HR	hazard ratio
IBTR	ipsilateral breast tumor recurrence
IHC	immunohistochemistry
iTRAQ	isobaric tag for relative and absolute quantification
kTSP	k-top scoring pairs
LC	liquid chromatography
LFQ	label-free quantification
MS	mass spectrometry
MS/MS	tandem mass spectrometry
OS	overall survival
PM	plasma membrane
PR	progesterone receptor
ROC	receiver operating characteristics
RSI	radiosensitivity index
RSS	radiosensitivity signature
RT	radiotherapy
SERD	selective estrogen receptor downregulator
SERM	selective estrogen receptor modulator
SILAC	stable in culture labeling of amino acids
SRM	selected reaction monitoring
SSP	Single Sample Predictor
STO-3	Stockholm-3
TMT	tandem mass tags

Studies included in the thesis

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

II. G protein-coupled estrogen receptor is apoptotic and correlates with increased distant disease-free survival of estrogen receptor-positive breast cancer patients. *Broselid S*, Cheng B**, ***Sjöström M****, Lövgren K, Klug-De Santiago HL, Belting M, Jirström K, Malmström P, Olde B, Bendahl PO, Hartman L, Fernö M**, Leeb-Lundberg LMF**.*
Clin Cancer Res. **2013** Apr 1;19(7):1681-92

III. Response to Radiotherapy After Breast-Conserving Surgery in Different Breast Cancer Subtypes in the Swedish Breast Cancer Group 91 Radiotherapy Randomized Clinical Trial. ***Sjöström M****, Lundstedt D*, Hartman L, Holmberg E, Killander F, Kovács A, Malmström P, Niméus E, Werner Rönnerman E, Fernö M**, Karlsson P**.*
J Clin Oncol. **2017** Oct 1;35(28):3222-3229

IV. Identification and validation of single sample breast cancer radiosensitivity gene expression predictors. ***Sjöström M****, Staaf J, Edén P, Wärnberg F, Bergh J, Malmström P, Fernö M, Niméus E**, Fredriksson I**.*
Submitted manuscript.

V. A Combined Shotgun and Targeted Mass Spectrometry Strategy for Breast Cancer Biomarker Discovery. ***Sjöström M****, Ossola R*, Breslin T, Rinner O, Malmström L, Schmidt A, Aebersold R, Malmström J, Niméus E.*
J Proteome Res. **2015** Jul 2;14(7):2807-18.

(*contributed equally as first authors, **contributed equally as last authors)

Studies not included in the thesis

Changes in glycoprotein expression between primary breast tumour and synchronous lymph node metastases or asynchronous distant metastases.
*Kurbasic E, **Sjöström M**, Krogh M, Folkesson E, Grabau D, Hansson K, Rydén L, Waldemarson S, James P, Niméus E.*
Clin Proteomics. **2015** May 12;12(1):13.

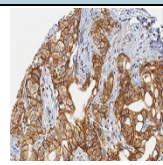
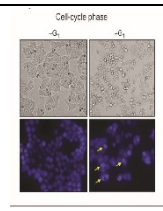
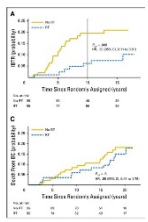
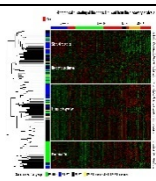
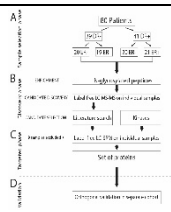
Stem cell biomarker ALDH1A1 in breast cancer shows an association with prognosis and clinicopathological variables that is highly cut-off dependent.
***Sjöström M**, Hartman L, Honeth G, Grabau D, Malmström P, Hegardt C, Fernö M, Niméus E.*
J Clin Pathol. **2015** Dec; 68(12):1012-9

Remarkable similarities of chromosomal rearrangements between primary human breast cancers and matched distant metastases as revealed by whole-genome sequencing. *Tang MH, Dahlgren M, Brueffer C, Tjitrowirjo T, Winter C, Chen Y, Olsson E, Wang K, Törngren T, **Sjöström M**, Grabau D, Bendahl PO, Rydén L, Niméus E, Saal LH, Borg Å, Gruvberger-Saal SK.*
Oncotarget **2015** Nov 10;6(35):37169-84.

Cancer associated proteins in blood plasma: Determining normal variation.
*Stenemo M, Teleman J, **Sjöström M**, Grubb G, Malmström E, Malmström J, Niméus E.*
Proteomics. **2016** Jul;16(13):1928-37.

TOP2A and EZH2 provide early detection of an aggressive prostate cancer subgroup. *Labbé DP, Sweeney CJ, Brown M, Galbo P, Rosario S, Wadosky KM, Ku SY, **Sjöström M**, Alshalalfa M, Erho N, Davicioni E, Karnes RJ, Schaeffer EM, Jenkins RB, Den RB, Ross AE, Bowden M, Huang Y, Gray KP, Feng FY, Spratt DE, Goodrich DW, Eng KH, Ellis L.*
Clin Cancer Res. **2017** 23(22); 1-12

Thesis at a glance

Study	Question	Patients and methods	Figure	Results
I	Is GPR30 prognostic and/or treatment predictive for tamoxifen?	Tumors from 742 patients from the randomized trial Stockholm-3 (2-5y of tamoxifen or no systemic adjuvant treatment) were scored for GPR30 expression and plasma membrane expression by immunohistochemistry (IHC).		GPR30 was not treatment predictive for adjuvant tamoxifen, but expression in the plasma membrane was associated with a worse prognosis.
II	Is GPR30 prognostic for breast cancer outcome? Is GPR30 proliferative or antiproliferative, and what are the basic signaling mechanisms?	Tumors from two patient cohorts (N=273 and N=237) were scored for GPR30 expression with IHC. Cell lines were analyzed with GPR30 transfection/knockdown, viability assays, western blotting, and flow cytometry.		GPR30 is prognostic for breast cancer outcome and appears to be constitutively active and pro-apoptotic in its signaling.
III	Is breast cancer subtype treatment predictive for adjuvant radiotherapy after breast-conserving surgery?	Tumors from 1003 patients in the SweBCG91-RT randomized trial (adjuvant radiotherapy vs no radiotherapy) were collected and stained/analyzed for ER, PR, HER2 and Ki67 for surrogate subtyping.		Subtype is not predictive of adjuvant radiotherapy, but HER2+ tumors appear to have a lower effect. A presumed low-risk group of patients have an excellent effect of radiotherapy.
IV	Can we use gene expression analysis to predict ipsilateral breast tumor recurrence and response to radiotherapy after breast-conserving surgery?	Tumors from 336 patients were collected and analyzed with gene expression analysis. Top discriminating genes and genes from the litterature were combined to a targeted panel. Single Sample Predictors were created based on machine learning algorithms.		The Single Sample Predictors were prognostic for ipsilateral breast tumor recurrence and showed promise in radiotherapy treatment stratification.
V	Can we combine different types of mass spectrometry-based proteomics techniques for biomarker discovery and validation in breast cancer?	Tumors from 80 patients with or without distant recurrence were collected. The N-glycosylated proteins were analyzed with discovery shotgun mass spectrometry. Top proteins were chosen for a targeted mass spectrometry technique (SRM), and all samples were re-run with SRM.		The workflow was established and applied to the tumors. The top five potential new biomarkers were further validated by gene expression analysis.

Introduction and background

Breast cancer epidemiology

Breast cancer is the most common cancer among women worldwide with almost 1.7 million estimated new cases in 2012, and it is the leading cause of cancer death with over 500,000 estimated annual deaths globally.[1] In Sweden, 9,382 malignant breast tumors in 7,929 women were diagnosed in 2015, of which 7,368 were diagnosed with breast cancer for the first time, representing approximately 30% of all cancers among women.[2] The survival chances after a breast cancer diagnosis has steadily improved, and the 10-year relative survival is over 80% in Sweden. Lung cancer has surpassed breast cancer as the leading cause of cancer death in both Sweden and other high income countries.[1,3] However, the incidence of breast cancer is increasing, which is likely in part related to changes in life style patterns such as longer life expectancy, fewer child births, later birth of first child, hormonal replacement therapy, less physical activity, overweight and higher alcohol consumption.[4] Despite the increased survival, the total number of breast cancer deaths and overall mortality has only seen a slight decrease during the last years.[2] The decrease in mortality is generally attributed to better adjuvant treatments, and to a lesser extent earlier detection of treatable breast cancer with mammography screening programs.[5-8] However, due to the small decrease in mortality compared to the increase in survival, concerns have also been raised regarding detection and over-diagnosis of indolent cancers that were previously not detected, and did not affect mortality.[8,9]

Breast cancer biology

In general, cancer transformation of normal breast cells is believed to start in the terminal duct lobular units.[10] By acquiring mutations in the DNA, the cells follow the hallmarks of cancer described by Hanahan and Weinberg, such as sustained proliferative signaling, evading growth suppressors, inducing angiogenesis and inflammation, activating invasion and metastasis, avoiding immune destruction and genome instability and mutation.[11,12] Most breast cancers are sporadic, meaning that

there is no known genetic predisposition and the process appear to be stochastic. Around 5-10% of patients have a known genetic background, most notably due to germline mutations in the tumor suppressor genes *BRCA1* and *BRCA2* that are part of the DNA-repair system.[13] Other genes with germline variants of more moderate penetrance have also been described, such as *PALB2, CHEK2, CDH1* and *ATM*, and testing is being evaluated before introduction in clinical routine.[14]

Increasingly deeper characterizations of somatic alterations in breast tumors have both improved the understanding of tumor biology, and provided insights in personalized treatment strategies.[15] Large efforts have catalogued the changes in breast tumors, and we are on the verge of understanding the carcinogenesis in individual tumors.[16-18]

Breast cancer subtypes

For decades it has been known that not all breast cancers are the same, and that the prognosis and treatment response varies with the pathological characterization of the tumor. Traditionally, and most important for treatment decision purposes, the tumors have been characterized based on the expression of the hormone receptors estrogen receptor alpha (ER) and the progesterone receptor (PR), which determine the response to endocrine therapy. The proliferative profile (e.g. measured with the proliferation marker Ki67) and the aggressiveness based on a morphological examination summarized as the histologic grade, have been estimated to determine the overall risk, and the associated benefit of chemotherapy. Further, assessment of the amplification and/or overexpression of the human epidermal growth factor receptor 2 (HER2) have been made, as HER2 overexpression and/or amplification (HER2+) increases the proliferation and aggressiveness of the tumor, and determines the response to anti-HER2 treatment. Once a marker of poor prognosis, HER2+ now indicates a treatment opportunity, and anti-HER2 directed therapy with trastuzumab, pertuzumab, lapatinib and the antibody-drug conjugate trastuzumab emtansine (TDM-1) has dramatically changed the prognosis for these patients.[19-21] Furthermore, *in vitro* data suggest that HER2 overexpression and/or amplification may contribute to radioresistance.[22-26]

Genomic profiling of breast cancer using transcriptomics revealed another level of the biologic diversity.[27,28] On the global gene expression level, it was first discovered that breast cancers cluster together in distinct groups, with implications for breast cancer prognosis and treatment. Originally, five subgroups of breast cancer (Luminal A, Luminal B, HER2-enriched, Basal-like and Normal-like) were described based on the global gene expression profiles, and the similarity to luminal and basal ductal cells.[19,29-31] The luminal-like tumors are characterized by expressing hormone receptors, while the basal-like subtype largely corresponds to a clinical

triple-negative tumor (lacking expression of ER, PR and HER2). The HER2-enriched subtype can be either luminal or non-luminal, and is characterized by increased HER2 signaling. The separation of Luminal A from Luminal B has implications for risk of recurrence and time to recurrence, with Luminal A having a better prognosis and later recurrences. The Luminal A from B separation is used to guide the use of adjuvant chemotherapy, but mainly captures a difference in baseline risk, which alters the absolute value of added treatment. Other subtypes, especially further divisions of the basal subtype, have been characterized and although the biologic difference is interesting, the implication for patient care is limited so far.[32]

Refinements of the clustering techniques has led to the use of a 50-gene subset (widely referred to as PAM50).[33] PAM50 classification is used in a commercially available and FDA-approved assay, that also calculates a risk of recurrence score, and do not contain the Normal-like subtype.[34,35] Several other commercial gene expression tests have been created for prognostication and to guide selection of adjuvant therapy.[31,36-40] However, the widespread use, particularly in Sweden, is awaiting definitive evidence of patient benefit relative to the cost, and many argue that the same information can be achieved by traditional pathology information. Indeed, cost-effective tests based on traditional pathology and immunohistochemistry based markers, with an associated surrogate subtyping, may with sufficient accuracy capture the intrinsic molecular subtypes, and are currently used in Sweden and elsewhere, although the performance and complementary information to gene expression is being evaluated.[29,41-43]

Ongoing clinical trials that evaluate the clinical utility of the gene expression tests include the MINDACT, RxPONDER and TAILORx trials.[39,44-48] First analyses have been presented, suggesting that that a low-risk group can be spared adjuvant chemotherapy treatment with remained safety, and that gene expression-based decision tools detects a larger low-risk group than traditional clinicopathologic variables, with the potential decrease of over-treatment.[36,49,50] However, the final long-term results are still to be presented.

Estrogen signaling in breast cancer and endocrine treatment

Estrogens are a class of steroid hormones, with 17-β estradiol being the physiologically most abundant, with important effects in breast cancer. The estrogen signaling is mainly mediated through the ER, which is activated upon ligand binding. Ligand binding causes dimerization, translocation to the nucleus, and activation of transcription that leads to proliferation and cell differentiation. Blocking of this signaling has long been used for treatment of breast cancer, but not all breast cancers are dependent on estrogen signaling for proliferative stimuli, and the tumor may not express ER at all. In addition, the tumor can lose the expression

of ER during tumor progression.[51,52] Intact estrogen signaling and expression of ER have large effects on the global transcriptome activity, reflected in the luminal vs non-luminal molecular subtypes, and ER positive (ER+) and ER negative (ER-) tumors may be regarded as different diseases.[28,53] It has long been suggested that the progesterone receptor (PR) is also required for intact estrogen signaling, and recently the molecular interplay between the receptors was shown, with implications for treatment efficacy.[54,55]

Possible ways of targeting the estrogen signaling is through selective estrogen receptor modulators (SERMs) such as tamoxifen, selective estrogen receptor down regulators (SERDs) such as fulvestrant, aromatase inhibitors (AIs) such as examestane, anastrozole or letrozole, oophorectomy or chemical castration with gonadotropin-releasing hormone (GnRH) analogues. However, not all ER+ tumors are sensitive to endocrine therapy, and they may present with intrinsic or *de novo* resistance, or acquire resistance during treatment. Several mechanisms of resistance have been described, with or without loss of ER expression, such as mutations in the *ESR1* gene, in the ligand binding domain or activating mutations, aberrations in co-factors or transcription factors, activating phosphorylation of the receptor, epigenetic changes, and cross talk with other signaling patwhays.[56,57] Many attempts have been made to find treatment-predictive markers and classifiers for endocrine treatment besides ER, including at the protein level, but no test is yet in clinical use.[58-63]

There are other estrogen receptors, most notably the estrogen receptor beta (ER beta), the splice variant of ER alpha, ER alpha-36, and the G protein-coupled receptor 30 (GPR30, also known as the G protein-coupled estrogen receptor, GPER), that may have different roles in breast cancer, and possibly in endocrine resistance.[64,65] The expression of ER beta in breast cancer has been suggested to affect prognosis and treatment response, but its role is less well understood, and has been widely disputed.[66,67] In addition, the research has been hampered by poorly validated antibodies.[68]

GPR30 is a putative estrogen receptor that may mediate non-genomic estrogen effects.[69-72] The receptor has attracted wide attention in cancer biology and is proposed to be a driver of cancer, mediate treatment resistance and to be a possible treatment target.[73-76] However, studies on the receptor are not conclusive in terms of basic cellular function such as subcellular localization, signaling mechanism, and the identity of estrogen as the cognate ligand.[77-85] Further, the receptor has been proposed to be both proliferative and antiproliferative.[86-91] The association with other clinicopathologic variables in breast cancer and prognostic implications are unclear.[92-94] Interestingly, this receptor has been proposed to be related to tamoxifen resistance, as several investigators have reported that tamoxifen, or one of its active metabolites 4-OH-tamoxifen, activates proliferation and migration through

GPR30.[86,95,96] In addition, studies demonstrated tamoxifen resistant MCF-7 cells to be more sensitive to 17-β-estradiol, and a proposed GPR30 agonist (G1). The same authors associated GPR30 expression with a worse prognosis only in tamoxifen treated patients, and showed GPR30 expression to increases with tamoxifen treatment *in vivo*.[87,97] Further, GPR30 activation by tamoxifen to up-regulate aromatase expression has been suggested to contribute to tamoxifen resistance, and GPR30 may be important for the interplay between cancer cells and cancer-associated fibroblasts contributing to drug resistance.[98-100] Interestingly, prolonged treatment with tamoxifen was also reported to up-regulate GPR30 specifically at the cell surface.[87,101]

Breast cancer recurrence

Breast cancer is a disease of the breast, but may spread or recur in the same breast, or at other sites. Hereafter, the term recurrence is used for any presentation of a breast cancer that is (thought to be) related to a first primary tumor, regardless of the timing of events. The site of recurrence, which is directly related to the chance of curing the disease, can be divided into ipsilateral breast tumor recurrence (IBTR, also commonly referred to as local recurrence), regional recurrence and distant recurrence (often referred to as metastasis).

Ipsilateral breast tumor recurrences (IBTRs) are recurrences of the tumor after surgery, still localized to the same breast. IBTRs are treated with repeated surgery (most often mastectomy), and possibly adjuvant treatment, but the recurrence may be of a more aggressive phenotype than the primary tumor. Indeed, a meta-analysis from the Early Breast Cancer Trialists' Collaborative Group (EBCTCG) on adjuvant radiotherapy, suggest that the prevention of four local recurrences can prevent one breast cancer death, emphasizing the importance of avoiding local recurrences.[102] In addition, the local recurrence may be a new primary tumor, not clonally related to the first tumor, with unknown treatment implications.

Regional recurrence is defined as a recurrence in the nearby (regional) lymph nodes, such as in the axilla, in the infraclavicular or supraclavicular fossae, or parasternal lymph nodes. A regional recurrence is still curable, but is one of the strongest risk factors for distant recurrence and breast cancer death.[103]

Distant recurrence to sites in the body beyond the regional lymph nodes is life-threatening for the patient. The process of cancer cells escaping the tumor and the breast stroma and environment, and being able to establish themselves at distant sites, is under intense research. Particularly, the time-point when a breast cancer acquires this phenotype has been widely discussed during the last century. From a macroscopic view, models have been proposed ranging from a fully *per continuitatem* spread, eventually reaching the bloodstream, i.e. the metastatic

potential is a late feature during tumor development, to a view that breast cancer is a systemic disease from the onset. In terms of cancer lineage, this is represented by either a linear or a parallel progression. This progression may also occur during the development within the primary tumor, contributing to intra tumor heterogeneity.[104] Modern techniques (such as next generation sequencing, NGS, and mass spectrometry-based proteomics) have revealed the evolution from primary tumor to distant recurrence by showing a constant clonal selection and dynamic evolution. The degree of similarity of the recurrences to the primary tumor varies, with different studies suggesting that many mutations occur late in the history of a cancer, or that many of the mutations are shared between primary tumor and distant recurrence, suggesting early events.[105] On a broader genomic scale, the similarity between the primary tumor and distant recurrence may be high, and larger rearrangements have been shown to be surprisingly similar between primary tumor and distant recurrence, suggesting early catastrophic events.[106,107] On the protein level, changes in protein abundance from primary tumor to recurrence have been described, which may be important for the metastatic phenotype.[52] There may also be a time-dependent effect on the degree of similarity between primary tumor and recurrence, i.e. the primary tumor is a good surrogate for choosing adjuvant treatment, but later distant recurrences have had time to evolve and acquire more mutations, and should be sampled for choice of treatment.[108] Importantly, the genomic changes that occur in the recurrences may be targeted with treatments, and large clinical programs exploiting this are ongoing.[109,110]

Breast cancer treatment

Modern treatment of breast cancer started in the late 19th century with the ideas of William Halsted and Willy Meyer that the tumor needed to be removed with its roots, leading to the radical mastectomy (lat. *radix* = root). The mutilating surgery of the radical mastectomy led to extreme morbidity.[111] Much has happened since, and the increased understanding of breast cancer biology has changed the "one size fits all strategy" to a personalized approach. With improved survival, more focus is on avoiding over-treatment, and improving the quality of life after breast cancer treatment.[112]

Risk assessment of the primary tumor

To determine the risk of recurrence and death from breast cancer, which also aids in the decision to add adjuvant treatments, factors of the tumor and patient are assessed. Traditionally, breast cancer has been classified and staged according to

the TNM classification, which stages the tumor for tumor size (T), if the tumor has spread to the lymph nodes (N) or to distant sites in the body (M). The TNM scores are combined to stages between 0-IV.[113] Further, the histologic type, as well as the histological grade of the tumor, is determined based on the tissue differentiation, exemplified by tubular formation, nuclear morphology and number of mitoses, according to Elston and Ellis.[114] In Sweden, the presence of the receptors ER, PR, HER2 and the proliferation marker Ki67, are routinely assessed with both prognostic and treatment-predictive implications.[115] Another important factor for an unfavorable prognosis is young age.[116] In parts of the world, the risk assessment is further guided by genomic tests, most importantly based on the assessment of gene expression for selected genes, mainly focusing on proliferation, cell cycle control, mitosis and receptor tyrosine kinase signaling.[117]

Currently, besides the use of ER, PR and HER2, the American Society of Clinical Oncology recommends the use of OncotypeDX (Genomic Health, Redwood City, CA), EndoPredict (Sividon Diagnostics, Köln, Germany), PAM50-ROR/Prosigna (Nanostring Technologies, Seattle, WA), Breast Cancer Index (Biotheranostics, San Diego, CA), and urokinase plasminogen activator and plasminogen activator inhibitor type 1, as biomarkers in primary breast cancer.[115] Further, the MammaPrint assay (Agendia, Irvine, CA) was recently also recommended for use in patients with high clinical risk, to avoid over-treatment with adjuvant chemotherapy.[31] Many of the gene expression tests capture the same underlying biology, but the exact set of genes included in each test show remarkably little overlap. To unify the concepts in the assays, genes from several assays were recently combined to a 95-gene panel with improved performance that may identify gene modules for targeted treatment.[117] On the other hand, a large proportion of gene expression values are correlated with each other, and with proliferation. The selection of genes at random have shown that sufficiently large gene sets (>100 genes) produce prognostic signatures, questioning the rationale for considering genes important for breast cancer biology by their association with outcome.[118]

Using traditional clinicopathologic factors, increased risk of IBTR has been associated with young age, multifocal tumors, large tumor size, and non-clear surgical margins.[119,120] Breast cancer subtypes have been associated with risk of IBTR with the same patterns as seen for distant recurrence, i.e. basal/triple-negative tumors having more early recurrences, while luminal tumors have fewer early recurrences, but more late recurrences.[121-125] The risk of IBTR for HER2+ tumors depends on anti-HER2 treatment.[126] Also, commercially available gene expression tests are associated with the risk of IBTR.[127,128]

Primary breast cancer treatment

The treatment of primary breast cancer, here defined as a cancer that has not spread beyond the regional lymph nodes to distant sites, and is thus considered curable, has undergone dramatic shifts in strategy since Halsted and Meyer.[111] First, the mastectomy was extended, and later modified to a less extensive procedure with less morbidity, and hereafter mastectomy will refer to the modern modified radical mastectomy. Although some evidence emerged that reducing the surgery did not alter the prognosis, the idea that an extensive operation was needed to cure breast cancer was largely dominant during the first half of the 20th century.[111]

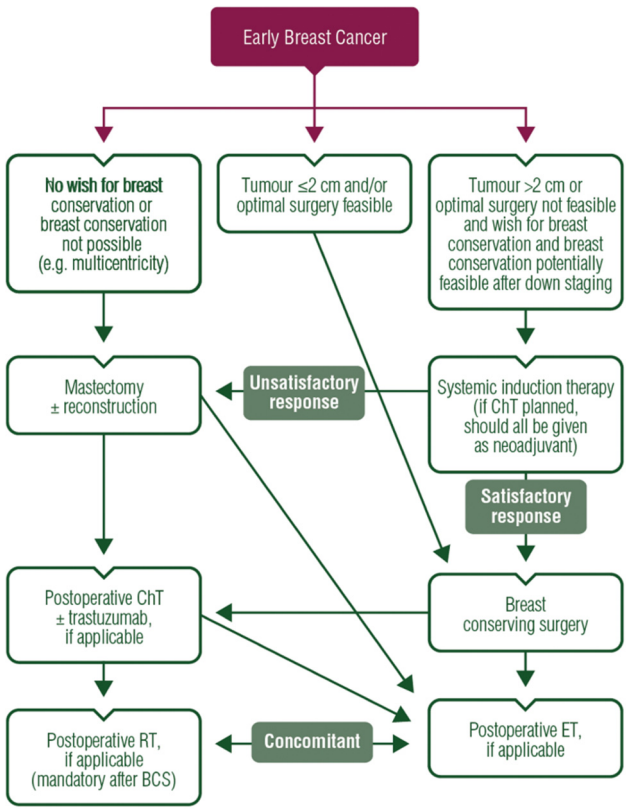


Figure 1. Treatment algorithm for primary (early) breast cancer by the European Society for Medical Oncolgy (ESMO).[129] ChT – chemotherapy, ET – endocrine therapy, RT – radiotherapy, BCS – breast-conserving surgery. Reprinted with permission from ESMO and Oxford University Press.

However, the traditional model of breast cancer as a disease with continuous spread was challenged with the growing evidence that breast cancer can be a systemic disease at time of diagnosis.[130] Indeed, occult disease may be present in the body at other sites without manifest disease, and tumor cells are commonly found in the

blood and in the bone marrow of primary breast cancer patients.[131] To treat also this occult disease, adjuvant therapy with radiotherapy, endocrine therapy, chemotherapy and targeted therapy is given to stop the breast cancer from recurring after surgery (endocrine therapy, such as tamoxifen, may be considered to be "targeted" to some extent, by I here reserve the term for more modern drugs designed for a specific target in the tumor). The advancement of adjuvant treatment allowed less extensive surgery to be developed, with the aim to minimize morbidity. The techniques of breast-conserving surgery were applied, and it has been shown in large trials that breast-conserving surgery with the addition of adjuvant RT (known as breast conservation therapy, BCT) is a safe alternative to mastectomy, or even superior.[132-135] Today, the consensus is that an invasive tumor should be removed macroscopically ("no tumor on ink"), including surrounding ductal carcinoma in situ (DCIS) if present, provided that appropriate adjuvant therapy is offered (Figure 1 and 2).[19]

Surgery

Breast-conserving surgery is recommended for the majority of patients where optimal surgery is feasible (e.g. excluding very large tumors and usually multicentric tumors), if the patient is willing to undergo the treatment, regardless if the tumor is multifocal (or multicentric per se). For patients undergoing mastectomy, immediate reconstruction should be offered if no oncologic contraindications are present, such as inflammatory breast cancer, but the availability varies between centers. Sentinel node biopsy is recommended for patients with clinically negative axilla. If no tumor cells, isolated tumor cells or micro-metastases are present, no further axillary surgery is needed, provided that the patient is offered adjuvant treatment according to established guidelines.[136] With macro-metastasis, the traditional recommendation was axillary clearance, but trials have shown that this may not be necessary.[19,129,137-140] For more conclusive results, the axillary clearance with macro-metastasis in one to two of the sentinel nodes is under further investigation.[141]

Adjuvant radiotherapy

Adjuvant RT to the breast is currently recommended for patients undergoing breast-conserving surgery, with a boost to patients with aggressive tumors (e.g. in Sweden to younger patients). Adjuvant RT to the chest wall internationally is recommended for patients undergoing mastectomy with four or more lymph nodes and/or large tumors (pT3). Nodal irradiation is recommended with one to three positive lymph nodes and other high-risk features, as well as for patients with four or more positive lymph nodes.[19] The value of administering RT to regional lymph nodes when micro-metastasis is present is unclear and under current investigation, but not generally recommended in Sweden.[129,142]

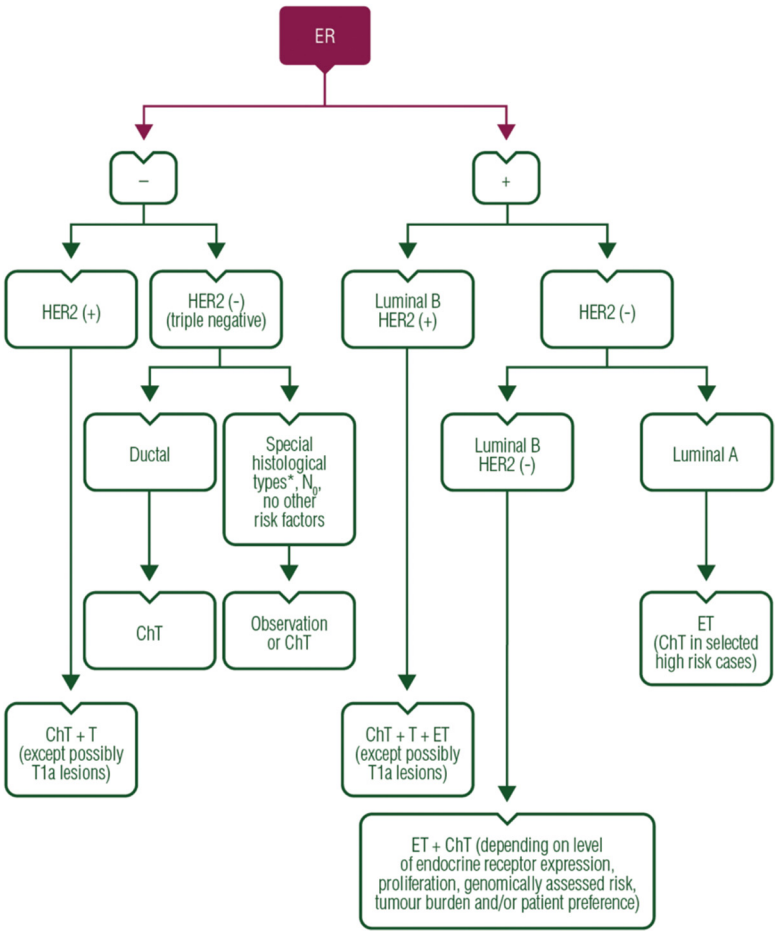


Figure 2. Treatment algorithm for adjuvant treatment after surgery for primary breast cancer by the European Society for Medical Oncology (ESMO).[129] ER – estrogen receptor, HER2 – human epidermal growth factor receptor 2, N_0 – lymph node negative, ET – endocrine therapy, ChT – chemotherapy, T – trastuzumab. *For For special histological types, ESMO recommends following the St Gallen 2013 recommendations that propose ET for endocrine responsive histologies (cribriform, tubular and mucinous), ChT for high-risk endocrine nonresponsive (medullary, metaplastic) and no systemic therapy for low-risk endocrine nonresponsive (secretory juvenile, adenoid cystic and apocrine).[29,129] Reprinted with persmission from ESMO and Oxford Universtiy Press

The addition of RT after breast-conserving surgery reduces primarily IBTRs, and has been shown to be a safe and effective treatment. A large meta-analysis by the EBCTCG have shown that RT also lowers breast cancer mortality, with a ratio that one breast cancer death is avoided at year 15 for every four IBTRs prevented at year 10.[102] However, most patients will not suffer a recurrence without RT, and some patients will suffer a relapse even with RT. The meta-analysis suggests that as much as 80% of patients may be over-treated or under-treated. Thus, there is a considerable interest in personalizing RT after surgery, and it would be very beneficial both to identify the patients that may be spared treatment, and the patients

with tumors that are radioresistant and will suffer a relapse even after postoperative RT. Indeed, several de-escalation trials are ongoing to investigate the safety of sparing a low-risk group of patients RT (e.g. the LUMINA study, NCT02653755, and the PRIMETIME study).[19,143] These initiatives are largely based on the assumption that the effect of RT is uniform across subgroups, and that the background risk determines the absolute benefit of the treatment. The goal is thus to find a subgroup with so low background risk that the absolute benefit of RT does not warrant the side-effects and cost, as opposed to assess the biological response in the tumor.

Several attempts for more personalized approaches have been made to individualize RT, but no study has been able to find a group of tumors without effect of RT, based on traditional biological markers.[102,144,145] The search for treatment-predictive markers has also been made on a molecular level, with most attempts being made at the gene expression level.[146-155] Promising results have been presented, but no marker or genetic profile is yet in clinical use.[156]

Adjuvant endocrine treatment

Endocrine therapy directed to interrupt the estrogen signaling is generally effective against tumors expressing ER, although there may be alterations of function not related to ER expression. Moreover, the exact definition of ER+ tumors is disputed, e.g. the cut-point of percentage of cells stained to consider a tumor ER+ varies.[157,158] Endocrine treatment reduces the rates of recurrence and breast cancer mortality.[158,159] For premenopausal women, tamoxifen is recommended, although ovarian functional suppression can be added for young women and women with high-risk tumors, in selected cases.[19] For post-menopausal patients, aromatase inhibitors appear to be superior to tamoxifen, but are associated with more adverse side effects.[19,158] The length of treatment is traditionally five years, but improved outcome has been demonstrated by prolonging the treatment, especially if patients received tamoxifen during the first five years, and for women with high-risk disease.[19,160,161] Recently, the long-term outcome after stopping endocrine therapy after five years was presented, and the risk of distant recurrence was high in several groups, further strengthening the rationale to continue endocrine treatment.[103] In Sweden, patients with N+ tumors are recommended extended adjuvant endocrine treatment up to 10 years.[142,162]

Adjuvant Chemotherapy

Adjuvant chemotherapy with anthracycline- and taxane-based regimens is generally recommended for patients with intermediate or high-risk breast cancer. The risk may be considered intermediate or high in triple-negative and HER2+ disease, and ER+ disease with risk factors such as positive lymph nodes, high proliferation or high genetic score on gene expression-based test.[19] Of note is that the biomarkers based

on traditional pathology and immunohistochemistry, as well as more modern molecular techniques, are more indicative of baseline risk than being truly treatment predictive. They are used to guide the use of chemotherapy by stratifying patients where the absolute benefit of treatment is greater than the side effects.[19,159,163]

Adjuvant targeted therapy

Anti-HER2 therapy has dramatically changed the outcome for patients with HER2+ disease. One year of adjuvant trastuzumab treatment is recommended for HER2+ disease, regardless if neoadjuvant dual HER2 blockade with trastuzumab and pertuzumab was given.[19]

Neoadjuvant treatment

Neoadjuvant treatment, i.e. giving the treatment before surgery, has the purpose of down-staging the tumor to make it operable, or operable with less extensive surgery, and has the advantage of early assessment of treatment response at the time of surgery. Neoadjuvant chemotherapy is recommended for stage II or III disease that is HER2+ or triple-negative, especially when down-staging for reduced surgery or radiotherapy might be possible. Dual HER2 blockade with trastuzumab and pertuzumab is recommended for HER2+ disease.[19]

Other adjuvant treatments

Postmenopausal patients are recommended adjuvant bisphosphonate treatment, as it reduces the rate of bone recurrence, and prolong survival.[19]

Advanced breast cancer treatment

The goal of the treatment of advanced breast cancer, i.e. breast cancer that have spread to distant sites, is to stop the cancer cells from growing and prolong the time before progression. Importantly, the treatment also aims to provide better quality of life, and treat pain and other problems related to the cancer. All of the treatment modalities used for early breast cancer treatment may be used in the metastatic setting. The evidence for one treatment over the other is however weak, although single drug regimens and lower doses may be used for reduced toxicity and improved quality of life.[20,164,165] Generally, endocrine treatment is given first-line to ER+ disease, HER2+ directed treatment in combination with chemotherapy for HER2+ disease (with the addition of TDM-1 at second-line, and possibly lapatinib third-line), and at progression chemotherapy is used sequentially until further progression. Of particular note is that the tumor characteristics of the primary tumor may change, and molecular analysis of the recurrence may alter the treatment options. It is therefore recommended to sample the recurrence if possible.[51,52] Interestingly, circulating tumor cells or circulating tumor DNA may be sampled in

the future as surrogates for the metastatic disease and used to guide treatment. Indeed, they show high discrepancy regarding HER2 status.[166-168] Several new treatment options are being introduced in clinical practice or are awaiting clinical approval, such as the CDK4/6 inhibitors for the treatment of ER+ breast cancer, PARP-inhibitors for *BRCA1/2* mutated tumors, mTOR-inhibitors and immunomodulating therapy.[169-171]

Conclusion of introduction

In conclusion, I believe certain areas of breast cancer research are critical for the improved understanding of the disease and improved patient care. One such area is systems biology, meaning that we must understand what happens in the cancer cells at a comprehensive level.[172] The individual technologies have so far been mainly implemented one at a time, providing an incomplete understanding on the different levels of biology. During the last couple of years, there have been efforts to combine data on the genomic, transcriptomic and proteomic levels to achieve a more complete picture of the cancer biology. Currently, researchers are working hard to understand how to combine the techniques. Much of the work and analysis made with high-throughput techniques rely heavily on knowledge of biological function, which is often limited and incomplete. Many times, the mechanistic details on how the molecules function cannot be achieved with current technologies, and the details must be elucidated with more time-consuming, traditional techniques such as *in vitro* or *in vivo* model systems. In the future, functional proteogenomics may facilitate this step, but it is still in need of orthogonal validation (independent with another technique).[173-175] I have deliberately aimed to use several different high-throughput techniques in my thesis work, as I believe a broad understanding of the techniques will be most valuable in future attempts to combine them, as well as to follow up the possible mechanistic implications.

A second prioritized area of research is personalized treatment of breast cancer, as highlighted by the cancer moonshot initiative.[176,177] The heterogeneity between tumors is large, and different types of breast cancers may be viewed as completely different diseases. It follows from the increasing understanding of tumor biology that every tumor has its own aberrations driving the carcinogenesis, and cancer transformation or progression, and must be treated accordingly. Ultimately, every breast cancer may be regarded as a rare disease on its own.[178] The improved outcome after a breast cancer diagnosis, and a survival approaching 90%, also means that a large number of persons will live long after a breast cancer diagnosis. This fact stresses the importance of reducing and eliminating the treatment side-effects. Thus, de-escalation when possible, and abandoning treatments that are not effective, has

become a major goal of breast cancer research today.[19] In the studies presented in this thesis work, we work towards individualizing the care of primary breast cancer to avoid both over-treatment and under-treatment. We investigate the possibility to tailor adjuvant endocrine therapy and radiotherapy by using a novel estrogen receptor (GPR30), evaluating its role in response and resistance, and also investigate the role of the breast cancer subtypes in response to radiotherapy. Additionally, we advance the high-throughput search for protein biomarkers for risk of metastatic disease, for possible use to guide adjuvant therapy.

A third area of prioritized research is completely novel treatments, and to overcome treatment resistance. Ultimately, personalized medicine in the sense that we better adopt the treatments that are already at hands, may only achieve better outcome to a certain extent. The prognosis of advanced breast cancer remains dismal as, eventually, the tumor becomes resistant, and no current treatment is longer effective. To overcome resistance, the main strategy has been to use information of the primary tumor, and theoretically predict response based on prior evidence and/or biologic rationale. However, ways of empirically testing treatments on tumors derived from patients are being investigated, and although not clinically feasible for all patients, it may find resistance mechanisms.[179] To truly transform the treatment of cancer, the hope is that the increasingly powerful techniques to analyze tumors will speed up discoveries, and provide an understanding of treatment resistance as well as completely new treatment targets. We explore the role of the novel estrogen receptor GPR30 in endocrine resistance, and its basic functional role in the cell, for a future use as a treatment target. Further, by using mass spectrometry to explore the proteins that may be responsible for distant recurrence, we move closer to potential treatment targets than is possible with genomics or transcriptomics.

Aims

The general aim of the work in this thesis was to:

- Use high-throughput methods to increase the knowledge of primary breast cancer biology, and develop tools and strategies for personalized medicine.

More specifically we aimed to:

- Investigate the role of the putative estrogen receptor GPR30 in breast cancer biology and progression, and its role as a potential biomarker for prognosis and treatment prediction for adjuvant endocrine therapy.
- Investigate if breast cancers of different subtypes respond differently to adjuvant radiotherapy after breast-conserving surgery, if the subtypes can be used for treatment stratification, and if a low-risk group of breast cancer can be spared adjuvant radiotherapy with remained safety.
- Develop a gene expression-based method to assess radiosensitivity that can be used for prognostication of ipsilateral breast tumor recurrence, and for treatment stratification of adjuvant radiotherapy after breast-conserving surgery.
- Develop a method for protein biomarker discovery, prioritization and validation, using different complementary mass spectrometry methods, and apply it to find biomarkers for risk of metastatic breast cancer.

Patients and methods

Patients

In this thesis work, we have studied tumors from six different patient cohorts (Table 1). All cohorts consist of patients with primary breast cancer, but differ in their type (RCT, cohort or case-control), menopausal status, nodal status, adjuvant treatment, primary endpoint and follow-up time. Of note is that endpoint distant disease-free survival (DDFS) follows an old definition, and only includes distant recurrence or breast cancer death (and not contralateral breast cancer or death from any cause).[180,181] Further, endpoint ipsilateral breast tumor recurrence (IBTR) is IBTR as first event.

Table 1. Patient characteristics of the studies included in the thesis.

Study	No	Type	Menop.	N	RT	ET	CT	Endpoint	FU
I	912	RCT	Post	N0	+/-	2-5y vs 0	0%	DDFS	17.0y
IIa	273	Cohort	Pre,Post	N0,N+	+/-	2y	0%	DDFS	6.1y
IIb	237	Cohort	Pre	N0	+/-	4%	10%	DDFS	10.8y
III	1003	RCT	Pre,Post	N0	+/-	7%	2%	IBTR	15.2y
IV	336	Case-control	Pre,Post	N0,N+	+/-	49%	21%	IBTR	12.6y
V	80	Case-control	Pre,Post	N0,N+	+/-	2y	0%	DDFS	6.4y

Study IIa – study II cohort I, study IIb – study II cohort II, No - number of patients before exclusion due to technical reasons, RCT- randomized controlled trial, Menop.- menopausal status, N – positive lymph nodes, RT – adjuvant radiotherapy, ET – adjuvant endocrine therapy, CT – adjuvant chemotherapy, FU – median follow-up time for patients alive and free from event. Few or no patients were treated with anti-HER2 therapy in the adjuvant setting.

Study I

Study I is based on the low-risk part of the Stockholm-3 (STO-3) randomized controlled clinical trial, which randomized post-menopausal patients with low-risk breast cancer (defined as tumor size ≤30mm and N0) to 2-5 years of tamoxifen treatment or no systemic adjuvant treatment.[182,183] 1,780 patients were included in the original trial and tumor tissue was collected from 912 patients for the construction of tissue microarrays (TMAs). Of the 912 samples, we were able to

stain and score 742 tumors for GPR30. DDFS was the primary endpoint, and median follow-up time was 17.0 years for patients free of event.

Study II

In study II, we used two retrospective breast cancer cohorts based in part on clinical trials.[184,185] Patient cohort I in study II (study IIa) consisted of 273 pre- and post-menopausal stage II breast cancer patients operated with modified mastectomy or breast-conserving surgery, with axillary dissection and adjuvant radiotherapy (RT) as indicated. All patients were treated with two years of adjuvant tamoxifen, irrespective of ER status. Endpoint used was DDFS, and median follow-up time was 6.1 years for patients alive and free from event.

Patient cohort II in study II (study IIb) consisted of 237 pre-menopausal women with node-negative breast cancer operated with modified mastectomy or breast-conserving surgery with adjuvant RT as indicated. A majority of patients received no adjuvant systemic treatment (4% endocrine treatment and 10% chemotherapy). Endpoint was DDFS, and median follow-up time was 10.8 years for patients alive and free from event. 29 patients were excluded because of no tumor material left, only normal tissue or cancer *in situ* present, or technical problems with staining and scoring.

Study III

Study III is based on the SweBCG91-RT trial which is a randomized controlled trial of breast-conserving surgery +/- adjuvant whole breast RT.[144,186] It consists of 1,178 patients with N0 stage I and II breast cancer with very little use of systemic adjuvant treatment (8%). Primary endpoint was ipsilateral breast tumor recurrence (IBTR) within 10 years, and secondary endpoints were any recurrence at 10 years, breast cancer death and death from any cause. Follow-up time was 15.2, 15.2, 20.0 and 21.2 years, for patients alive and free from event, for the respective endpoints. We were able to collect tissue from 1,003 patients, and successfully stained and scored markers on TMAs necessary for subtyping for 958 patients.

Study IV

Study IV is based on a retrospective collection of 336 fresh frozen tumors from three biobank centers in Sweden, divided into a training cohort (N=172) and a validation cohort (N=164). Tumors were selected and sampled based on development of an IBTR (cases) or being recurrence free (controls) for at least the same time as a matched case. Matching was done for RT, ER status and follow-up time. Patients were pre- or post-menopausal, and operated with breast-conserving surgery with or without adjuvant RT. Axillary dissection was performed as indicated (22% were node-positive), and adjuvant systemic treatment was administered according to regional guidelines at the time. Primary endpoint was IBTR and median follow-up

time for patients free of event (controls) were 13.2 years for the training cohort, and 12.6 years for the validation cohort, respectively.

Study V

Study V is based on retrospective collection of 80 breast tumor samples from patients participating in previous clinical trials.[184,185] The tumors were sampled in a semi case-control fashion so that the patients were either suffering distant recurrence (N=41), or were recurrence-free for at least 6.4 years (N=39). All patients were diagnosed with stage II breast cancer and treated with surgery, RT and adjuvant tamoxifen for two years.

Methods

Cell lines and *in vitro* experiments

Two breast cancer tumor-derived cell lines natively expressing GPR30 were used in study II, MCF-7 and T47-D, which are both ER+ cell lines and belonging to the Luminal B subtype, based on gene expression studies.[187] In addition, we used two recombinant cell lines without or with stable expression of human GPR30 tagged at the N-terminus with the FLAG epitope. These were HEK293 cells, a well proven model system for G protein-coupled receptors, and HeLa cells, used for receptor antibody validation.

Western blotting was performed in study II to assess GPR30 expression and signaling protein responses. Proteins analyzed were native and recombinant FLAG-tagged GPR30, signaling proteins in proliferation (ERK1/2 phosphorylation, p53), and signaling proteins in apoptosis and cellular stress (ubiquitin, PARP, caspase-3, cytochrome C).

Flow cytometry was used in study II for assessing plasma membrane expression of GPR30 and cell cycle analysis. Confocal immunofluorescence microscopy was used to assess cellular localization and expression of GPR30 and cytokeratin 8.

In study II, cell viability was assessed by visual inspection of cell morphology, by Hoechst 3342 staining and phase-contrast microscopy to visualize nuclear fragmentation, and by the MMT assay and spectrophotometry to monitor the conversion of MTT to formazan in living cells.

Constitutive GPR30 activity was studied, in the absence of any added receptor stimulus, by assaying signaling proteins in 1) HEK293 cells stably expressing GPR30 (HEK-R) as compared to naïve HEK293 cells (HEK), and 2) naïve MCF-7

cells as compared to MCF-7 cells in which GPR30 had been transiently knocked down with GPR30-specific shRNA.

Immunohistochemistry and tissue microarrays

Immunohistochemistry (IHC) was used to evaluate the protein expression on tissue sections with protein specific antibodies, and conjugated reporter systems for visualization. IHC is based on a two-step process of protein targeting and visualization. A pre-treatment step is used for antigen retrieval for formalin-fixed paraffin-embedded (FFPE) tissue. A primary antibody, targeting the protein of interest, is added and bound to proteins on the tissue slide. After washing, a secondary antibody with species specific targeting of the primary antibody, and conjugated to a reporter system, or biotinylated for further amplification with other reporter systems, such as streptavidin-HRP, is added. Finally, a chromogenic substrate is added for visualization. The major reason for the two-step process is the signal amplification and the high cost to produce antibody-reporter conjugate for specific antibodies. IHC gives an opportunity to study protein expression *in situ*, with the possibility to evaluate only the cancer cells (or stromal cells), and subcellular localization of staining, such as plasma membrane or nuclear expression. Potential problems with IHC may be associated with the fixation and pre-treatment of the tissue, protein degradation due to sample age or incorrect tissue handling, nonspecific staining and, importantly, antibody specificity. Problems may also arise in the scoring and analysis of the data, as discussed later.[188] Important to note is that IHC is semi-quantitative due to the amplification process, and attempts to use staining intensity for quantitative measurements of protein expression must be interpreted carefully.

We used IHC in study I, II and III. The GPR30 antibody was validated in-house with wild-type HeLa cells (not expressing GPR30), and cells transfected with GPR30. The antibodies used in study III are used in clinical routine and have been validated extensively by others.

To increase the throughput of tumor evaluation of protein expression by IHC, we used TMAs.[189] Instead of analyzing whole tissue sections for each tumor, several representative core punches (usually 0.6 or 1.0 mm in diameter) are taken and mounted in a recipient paraffin block. Up to 200 cores can be mounted in one block, which is then sectioned and stained, allowing an investigator to evaluate 200 cores on one microscopy slide. Concerns have been raised that the small tumor area analyzed may not be representative of the tumor, especially considering the heterogeneity in breast cancer.[190] However, good correlation between whole tissue sections and TMAs have been presented for common clinically used IHC

markers.[191] The evaluation of TMAs has been further facilitated by the use of digital pathology, making it even more high-throughput.

We use IHC and TMAs to evaluate GPR30 on patient cohorts in study I and II, and to evaluate the factors for subtyping in study III. At least two investigators scored the tumors, and two or three cores were analyzed for most tumors. Depending on the factor studied, the method to combine cores and cut-offs for grouping was different (for details, see respective study and the discussion section). To evaluate other pathologic factors such as histological grade, whole tissue sections were used.

Gene expression analysis

The central dogma of molecular biology states that DNA is transcribed to RNA, which is translated to proteins.[192] This process is tightly regulated at many steps, and the concept of gene expression analysis is based on the idea that the activity and expression of a gene, meaning the amount of RNA that is present at any moment, carries information important for the function of the cell.[193] Usually that is interpreted as more protein being translated, resulting in higher activity of the protein, although the correlation between RNA expression and protein amount is sometimes weak, and RNA may have other regulatory effects.[194,195] Traditionally, the measurement of mRNA molecules has attracted most attention, but other types of RNA species, such as microRNAs and long non-coding RNAs, are getting more attention for roles in cellular and tumor biology.[196,197]

The analysis of gene expression in tumor biology started to grow when oligonucleotide or cDNA arrays were created, which made the analysis of thousands of genes possible for hundreds of samples. Oligonucleotide probes were "spotted", or synthesized *in situ*, to a solid surface and organized in tiny arrays – microarrays. RNA from samples are extracted, reverse transcribed to cDNA, amplified and labeled by either Cy5 (red) or Cy3 (green). The experimental and a reference sample, with different color labeling, are hybridized to the microarray with the most abundant sample binding more probes. A scanner registers intensity for each dye, and the ratio between the experimental and reference sample is interpreted as a relative measurement of RNA amount, and thus gene expression and activity. Reports soon followed that the transcriptional patterns provided crucial information on the active state of the tumor, and clustering was used to divide breast cancers into distinct subtypes, with clinical implications.[27,28] The oligonucleotide technique was refined, and highly stable commercial gene expression microarrays were created and widely used. Some of the refinements included single channel arrays without the need for a reference sample, and bead arrays.[198-200]

However, gene expression microarrays are limited by several shortcomings. First, they require relatively high-quality RNA that is typically only achieved by RNA

extraction from fresh frozen tissues. They are thus hard to implement in a clinical setting, although microarrays have recently been adapted to also being able to handle lower quality RNA, e.g. from FFPE samples, which is achieved either by a RNA pre-processing step involving amplification, or more robust microarray techniques.[201] Secondly, traditional microarrays are restricted to analyze gene products represented by the probes on the array, meaning that alternative splice products, as well as mutations, are not detected. Thirdly, the sensitivity may be low, and limit of detection high, for traditional microarrays. Another aspect, although not limited to microarrays, is that the extraction of RNA from tumor tissue usually includes both cancer cells and stromal cells. This makes it hard, or impossible, to determine which part of the tumor (e.g. tumor cells, fibroblasts or immune cells) contribute to the observed gene expression results.

Other genes expression analysis techniques have been developed with advantages that overcome some of the shortcomings of microarrays. Targeted techniques have higher sensitivity and can be tailored to find almost any gene product of interest. Typical targeted techniques include quantitative polymerase chain reaction (qPCR), used e.g. in the OncotypeDX test, and which has been modified for the analysis of very small amounts of RNA (digital droplet PCR).[202] Another targeted technique is the Nanostring nCounter platform.[203]

RNA sequencing (RNA-seq) has been developed and is widely used in cancer research today. It has the advantage of the possibility of detecting novel gene transcripts but still have disadvantages, such as high cost, requirement of high-quality RNA, and more complicated data-analysis.[204]

In study IV, gene expression analysis was performed with the transcriptome wide Illumina HT12 v4 microarray (Illumina, San Diego, CA), and the targeted Nanostring nCounter platform (Nanostring Technologies, Seattle, WA), with a custom design assay.

Illumina HT12 v4 microarray

Illumina HT12 v4 microarray is a bead array oligonucleotide microarray. Nucleotide sequence probes are attached to small beads, which are randomly distributed over an array, and an individual key is needed for each array to decipher the positions. It is a single-channel technique which involves cDNA production, amplification, *in vitro* transcription to cRNA, and labelling with biotin. Each chip has 12 arrays for the analysis of 12 samples. The array includes 47,231 target probes that mainly targets coding genes, and further includes negative and positive controls. The microarray has been widely used and is considered robust and well annotated. Typical pre-processing includes background correction, positive offset for avoiding negative values, log2 transformation, quantile normalization and filtering of probes for quality, intensity and variance.[200,205]

Nanostring nCounter platform

The Nanostring nCounter technique uses a target probe and a reporter probe, with the aim to identify and count individual molecules. Initially it was created only for mRNA molecules, but the technique has been extended to analyze also other molecules, including proteins.[206] The reporter probe and capture probe have 50 nucleotides each that are adjacent and target a specific 100 nucleotide target sequence.[203] The reporter probe have an additional specific sequence of fluorescent molecules that identify the probe. The capture probe has a biotin molecule attached that is used for in solution capture and later attachment to a cartridge. Excess probes are washed away, and a scanner is used to detect the reporter probes with an individual key to decode the scan. The probes can be designed to detect any specific nucleotide sequence, and up to 800 probes can be analyzed simultaneously in a run. Typically, house-keeping genes are included for normalization purposes. Pre-processing is more custom than for a standard microarray, but typically includes normalization based on house-keeping genes, and filtering of probes and samples based on quality.[207] The assay benefits from high sensitivity and specificity, but is limited to the subset of targets included in the assay.

Mass spectrometry-based proteomics

Protein analysis and proteomics

Proteomics is the study of the proteome, meaning the entire set of proteins in a given system. The proteome is vastly more complex than the genome or transcriptome, as every gene can give rise to many protein isoforms, and proteins may be subject to post translational modifications governing their function (e.g. phosphorylation and glycosylation), creating a large number of proteoforms. In addition, the temporal control of protein abundance is important, and the cellular response can be much faster at the protein level than the genomic or transcriptomic response, e.g. by activating phosphorylation. As proteins are the major functional molecules in the cell, additional functional understanding of cancer biology could be obtained by analyzing proteins in addition to DNA and mRNA sequences. Although major biological concepts seem to be similar across techniques, the correlation between mRNA expression and protein abundance is sometimes weak.[208] Ultimately, proteomics may provide better biomarkers and novel treatment targets.[209-211] Indeed, characterizations of tumors at the protein level partly recapitulates the genomic and transcriptomic findings, but also provide additional information.[17,212,213]

Traditionally, protein analyses have been performed with low-throughput methods such as antibody-based techniques, e.g. enzyme-linked immunosorbent assay (ELISA), western blotting (WB) or IHC, alone or in combination with gel electrophoresis-based protein separation. Besides low sample throughput, the

analysis with antibody-based methods is dependent on the development and validation of antibodies, which may be time-consuming and costly. However, with the improvement of chromatographic separation and mass spectrometry instrumentation, the field of proteomics has grown exponentially, becoming a high-throughput technology able to analyze thousands of proteins out of a biological sample in a few hours.

Mass spectrometry principles

Mass spectrometry identifies molecules based on an accurate determination of their mass to charge ratio (m/z). The basic principle is to ionize the molecules and accelerate the ions within an electric field. The mass analyzer uses the properties of the charged molecule in a magnetic field to separate the ions, exploiting the principle that the charged particles will deflect in the magnetic field proportionally to their mass. Finally, a detector records the ion. Several types of mass analyzers exist, such as quadrupoles (Q), Linear ion traps (LIT), time-of-flight (TOF) analyzers, Fourier transform ion cyclotron resonance (FTICR) analyzers and orbital traps (Orbitrap), as well as different detectors such as electron multipliers and detectors based on FTICR (Figure 3). In study V, we use LIT combined with FTICR analyzers for a discovery analysis, and quadrupoles for targeted proteomics.

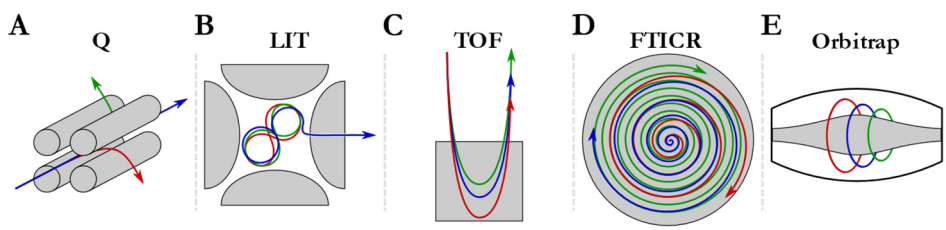


Figure 3. Different mass analyzers. The colored arrows represent molecules with different m/z, and the grey are electromagnets. **A)** In a quadrupole mass analyzer (Q), four magnetic rods act as a filters and only lets a molecule with a certain m/z through, deflecting the others. **B)** In a linear ion trap (LIT) the ions are "trapped" in an electromagnetic field based on their m/z. **C)** A time-of-flight mass analyzer uses the principle that a charged particle deflects proportional to the mass in an electromagnetic field. **D)** Fourier transform ion cyclotron resonance (FTICR) mass analyzers analyze the movement within an electromagnetic field. **E)** An Orbitrap is a barrel-like structure and the ions oscillate around the barrel, with the oscillations dependent on the m/z. Illustration by Kristoffer Sjöholm, reprinted with permission from *A holistic approach to host-pathogen interactions – detecting the large to unravel the small.*[214]

Several steps must be performed before proteins of a sample can be run on a mass spectrometer. Typically, the samples are prepared either chemically or mechanically to extract the proteins, cleaned for contamination, denatured and digested with an enzyme. The digestion of peptides before analysis is called bottom-up proteomics, as opposed to top-down approaches, where the intact proteins are separated before analysis. Digestion is usually done by trypsin, as this produces the most proteotypic

peptides (i.e. specific to a protein). Trypsin cuts proteins into peptides of 6-25 amino acids, split at the C-terminal side of lysine and arginine, except (maybe) when the next amino acid is proline.[215] If the digestion is made in solution, the peptides need to be reduced and alkylated.[216,217]

Further, the online separation of peptides by liquid chromatography (LC) is usually applied to separate the peptides based on chemical properties, such as hydrophobicity, before injection into the mass spectrometer.[210]

The mass of an intact peptide is generally not sufficient to identify the exact amino acid sequence. The solution is to apply two mass spectrometers that work in tandem (MS/MS – tandem mass spectrometry), either in space, or time (Figure 4). The first MS (MS1) analyses the intact peptides (precursor ions) which are then fragmented, typically in a collision cell with an inert gas using collision induced dissociation (CID), and the second MS (MS2) analyzes the fragment ions.

The specific analysis of proteomics data consists of several parts that can be divided into identifying the peptide, combining peptides to proteins, and quantification. The ways to perform these steps are dependent on the mass spectrometry setup.[218]

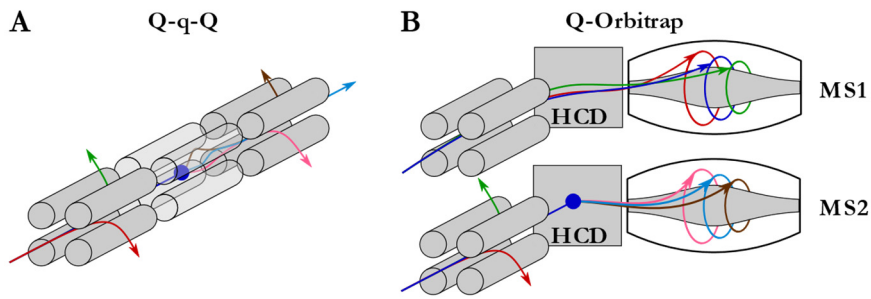


Figure 4. Tandem mass spectrometry using two mass analyzers that work together. **A)** In targeted mass spectrometry-based proteomics, a common approach is to use a triple quadrupole (Q-q-Q) set-up, using two quadrupoles as filters, and the third quadrupole as a collision cell for fragmentation. **B)** In discovery mass spectrometry proteomics, a common set-up is to use a quadrupole to filter peptides for analysis with an orbitrap. In the MS1 scan, several precursor ions are let through and the orbitrap analyzes the intact peptides. For fragmentation analysis (MS2 scan), the quadrupole filters a specific peptide for fragmentation with high-energy collision induced dissociation (HCD), and the fragments are analyzed by the orbitrap. Illustration by Kristoffer Sjöholm, reprinted with permission from *A holistic approach to host-pathogen interactions – detecting the large to unravel the small*.[214]

Discovery mass spectrometry proteomics

Shotgun, or sometimes "discovery", proteomics refers to using a bottom-up LC-MS/MS configuration operating in data-dependent acquisition (DDA) mode. In DDA mode, the mass spectrometer first performs a MS1 run of intact peptides,

selects a number of peptides based on the MS intensities (dependent on the data) to fragment, and then performs sequential runs to fragment and analyze the peptides with MS2. The precursor ions (intact peptides) are then placed in a dynamic exclusion list, and other precursor ions are selected for fragmentation and MS2 analysis.

The subsequent identification of peptides can be made by interpreting the fragmentation (MS2) spectra as amino acid sequences (*de novo* sequencing). However, this is generally too time consuming and difficult for practical purposes in clinical research, as some amino acids and amino acid pairs have very similar or identical mass, and fragmentation does not occur at every peptide bond or cannot be observed in the mass spectrometer. A faster and usually more feasible way is to use a reference database of with *in silico* tryptic peptides, create theoretical mass spectra, and compare with the experimental spectra using specialized software.

Error estimates of peptide identifications are usually made by a search against a false decoy database, such as the reference database in reverse, or random sequences, generating a false discovery rate (FDR) estimate. Important to note is that the probability of finding a false hit increase with the size of the reference database – there are more possible false hits. This becomes very important when we think of analyzing cancer genomes including somatic changes: if all possible or known mutations are included in the database, the search space will be large and the probability of a false hit high. To overcome this, a reference database from the specific DNA sequence of the tumor would be valuable, and indeed, this area of combining genomics and proteomics – proteogenomics – promises to give new insights in cancer biology.[219,220]

After peptide identification, the next step is to match peptide to proteins. The principle is the same in that the most likely peptide sequence, given from the mass spectra searches, is searched against a protein database. This is also associated with uncertainty, and error rates on the peptide level are amplified on the protein level.

The results of LC-MS/MS operated in DDA mode are inherently random to some extent due to the selection of precursor ions (intact peptides) for fragmentation, meaning that two runs may not be identical, and DDA produces large numbers of missing data. Further, the quantification based on the precursor may be unreliable as the precursor will be excluded from fragmentation for a period of time. To overcome this problem, targeted approaches have been created.

Targeted mass spectrometry proteomics

A common targeted approach is selected reaction monitoring (SRM), or multiple reaction monitoring (MRM), which is based on a triple quadrupole mass spectrometry set-up, with two quadrupoles acting as filters, and a third quadrupole as a collision cell resulting in a triple quadrupole set up, Q-q-Q (Figure 5).[221] It is

operated so that the first quadruple is a static filter for a specific peptide, the second quadrupole fragments the peptide, and the third is a static filter measuring a specific fragment ion while. The final detector records amount of fragment ions that are let through the filters, giving a relative estimate of the abundance. The combination of precursor ion and fragment ion is called a transition, which in combination with dilution time from LC results in very high peptide specificity. Typically, several transitions are used per peptide to create an assay for reliable measurements. The pre-requisite to use SRM is that the properties of the peptide is known, and large efforts have been made to create libraries of assays.[222] The increased quantitative performance and dynamic range of SRM, compared to shotgun LC-MS/MS, promises to greatly facilitate the clinical use of mass spectrometry.[223]

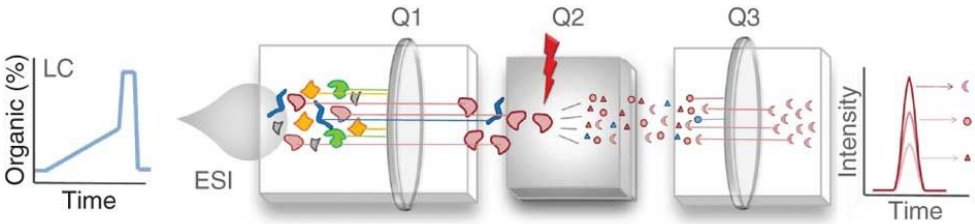


Figure 5. Overview of the selected reaction monitoring (SRM) method of targeted proteomics. First, the peptides are separated by liquid chromatography (LC), then ionized by electrospray ionization (ESI). In the first quadrupole (Q1), a specific intact peptde is selected for fragmentation in the second quadrupole (Q2). The third quadrupole (Q3) is a filter for selected fragment ions. Finally, a detector is registering the amount of fragment peptides, which is used for quantification. Reprinted from Picotti et al in *Nature methods*[224] by permission from Macmillan Publishers Ltd, copyright (2012).

Data-independent analysis

Recently, a data-independent acquisition (DIA) mode has been set up, which aims to combine the identification advantages of shotgun/discovery mass spectrometry with the quantification advantages of targeted mass spectrometry.[221] In DIA-MS, all precursor peptides are fragmented and recorded, creating a highly populated dataset (Figure 6). With advanced data analysis, the peptides can be identified, and abundance measurements achieved retrospectively.[225] Still being optimized and relying on libraries of peptide properties created with DDA approaches, and better data analysis tools, it could provide a middle ground between shotgun and targeted mass spectrometry. One implementation of the concept, termed Sequential Windowed Acquisition of all THeoretical fragmentation spectra (SWATH) uses the first mass analyzer not set to select a specific precursor ion, but instead an entire mass window that is typically 2-25 m/z wide. All precursor ions in that window will be fragmented and mass spectra of all fragment ions generated. Given that we have previous knowledge of the properties of the spectra, we can identify and quantify all peptides in a retrospective/post-acquisition analysis of the data. The technology is promising, and intensive research is ongoing to optimize the methods.[225,226]

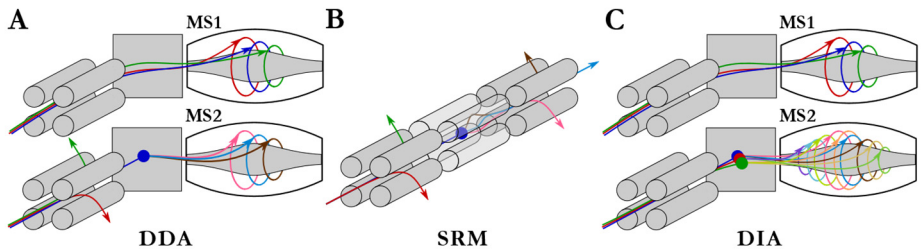


Figure 6. Comparison of different mass spectrometry set-ups. **A)** In data-dependent acquisition (DDA), precursor ions are selected for fragmentation. **B)** In selected reaction monitoring (SRM), precursor ions and fragment ions are selected by quadrupoles that act as filters. **C)** In data-independent acquisition (DIA), the precursor ions in a selected mass window are fragmented, and all the fragment ions are analyzed. Illustration by Kristoffer Sjöholm, reprinted with permission from *A holistic approach to host-pathogen interactions – detecting the large to unravel the small.*[214]

Quantification

After peptide and protein identification, protein abundance is calculated. For LC-MS/MS operated in DDA mode, a simple way is to use the number of times mass spectra from a peptide is appearing (spectral count). More common is to use the peak area under the chromatogram/intensity of the MS1 peak, commonly referred to as label-free quantification (LFQ).[221] LFQ can timely quantify many peptides, but the intensity of the peak is dependent on the chemical and physical properties of the peptide, meaning that the peak intensity is hard to interpret in terms of absolute abundance.[227]

The quantitative performance of shotgun mass spectrometry can be enhanced by the use of labeling methodologies, such as stable in culture labeling of amino acids (SILAC), or methods of chemical labeling, such as isobaric tag for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT). Peptides from different samples are tagged with heavy or light isotopes, and can then be compared relative to each for a reliable relative quantification (Figure 7).[218,228]

In SRM, the goal of the analysis is not to identify the molecule, as the properties are already known, and the machine is set-up to specifically analyze the peptide of interest.[221] The intensities of the fragments detected with the final detector are compared across samples as a relative measurement of abundance, and automated software have been created for this, as well as for manually evaluating the profiles.[229-233] LFQ can be used with SRM, but the performance may be improved by adding labeled standards.[228]

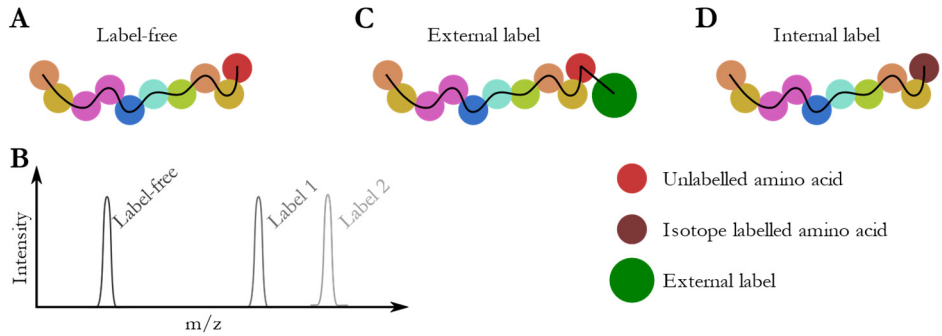


Figure 7. Labeling of amino acids for quantification. The principle is to tag peptides from different samples with heavy or light isotopes to make them separable in the mass spectrum created by injecting the samples togeher **(B).** Illustration by Kristoffer Sjöholm, reprinted with permission from *A holistic approach to host-pathogen interactions – detecting the large to unravel the small.*[214]

The N-glycosylated subproteome

To achieve even greater depth of analysis than can be achieved separating the peptides with LC alone, further fractionation of the samples can be made. Oher ways of decreasing the complexity are to remove proteins that are not relevant for analysis, such as the depletion of plasma proteins, or to focus the analysis on a selected subproteome, which also excludes irrelevant proteins such as albumin or immunoglobulins.[234-236]

The N-glycosylated subproteome constitutes an interesting subproteome in cancer research since it contains proteins at the cell surface important for cell-cell communication, interaction with the intracellular matrix, receptors and other signaling molecules, and are commonly found among clinically used biomarkers. Further, they are commonly secreted to the bloodstream which may make the clinical detection easier.[237,238] Therefore, in study V, we chose to analyze this subproteome.

Glycosylated proteins are post-translationally modified by attaching a carbohydrate group, either to an asparagine which takes place in the endoplasmic reticulum (N-linked glycosylation) or to the oxygen atom of a serine or threonine performed in the Golgi apparatus (O-linked glycosylation). N-glycosylation is performed to an asparagine with the surrounding motif N-X/P-ST, where X/P means any amino acid except proline, and ST means serine or threonine. In study V, we enriched for N-glycosylated peptides using hydrazide chemistry. Briefly, the enrichment is made on peptides cleaved by trypsin. The carbohydrate *cis*-diol group is oxidized to aldehyde which then can form a covalent hydrazone bond with hydrazide groups immobilized at a solid support (beads). The bound peptides are washed to remove non-bound material. Finally, the peptides are released by cleavage of the carbohydrate residue with PNGase F. This cleavage also results in the deamidation

of the asparagine, producing an aspartic acid, which is used for further validation of the specific N-glycosite.[239,240] Of note is that in study V, the carbamidomethylated cysteine is denoted as B, and the deamidated asparagine is denoted as an aspartic acid (D). Also, trypsination may cleave the peptide in the N-glycosylation motif (if the motif is N-RK-ST, RK meaning arginine or lysine), meaning that the N-glycosylation sequence motif can be found at the peptide end, and thus only have a deamidated asparagine (aspartic acid - D) and an arginine (R), or lysine (K), at the C-terminal end, but a serine (S) or threonine (T) at the other side of the cleavage, which is the case for some of the peptides in study V.

Statistical analysis

Statistical significance and two group comparisons

The definition of a p-value is the probability of getting the observed result, or a more extreme result, if the null hypothesis is true, with the null hypothesis being that there is no difference between the groups tested. The p-value is traditionally used to reject the null hypothesis if a certain significance level is reached, typically <0.05. However, this has nothing to do with "clinical significance", meaning the effect size and implication for patients, and thus a p-value should always be interpreted in combination with effect size.[241] Important to note is that for a given effect size, a larger sample size generates a lower p-value, and thus with enough sample size small differences become "significant". This is important to note in relation e.g. to the *Kaplan-Meier plotter* results in study IV.[242] Conversely, for a given effect size, a small sample size produces a high p-value and it is thus important not to interpret a non-significant p-value in an underpowered study as evidence that there is no difference.

For standard two-group comparisons, the t-test and Mann-Whitney test (also known as Wilcoxon's test) can be used as the parametric and non-parametric variants, respectively, and both can also be used with modifications for paired data. Similar tests for associations with categorical variables include Fisher's exact test and the Chi-square test which both calculates the probability to get the observed or more extreme results in a cross-tabulation. The Chi-square test is a parametric test that is built on an approximation to the chi-square distribution, while Fisher's exact test analyzes the exact probabilities making it better suited for small groups.[243,244] The Chi-square test can be used to test for trend, which equals a linear regression with test for zero slope.

Survival analysis

Survival analysis deals with time-to event data, which is commonly used in cancer research. Problems arise on how to handle observations that do not start and end at the same time, and observations that may end for other reasons than the intended endpoint, e.g. due to end of follow-up.[245] Initially this was made by creating life tables but a more direct method was applied by Kaplan and Meier, that censors observations and calculates the risk for the remaining population.[245-247] To test if the observed differences between groups are likely to happen by chance, a log-rank test is usually used. The log-rank test can be extended to test several groups, or to test for trend.

The most commonly used model to compare the risk between groups, as separated by a biomarker, is the proportional hazards regression model proposed by Sir David Cox.[248] The model calculates the hazard ratio between groups, and is assuming the hazard to be proportional over time, which may be checked with Schoenfeld's test, or by graphical examination of the model or residuals.[249] If the hazards appear to be non-proportional, a proportional hazards regression model may still be fitted, but the interpretation would only allow to consider the hazard ratio as the mean over the time period studied. The Cox proportional hazards model can be used to compare several groups, and can either handle factor variables describing the hazard ratios between pairs of groups, or treat the variable as a linear/continuous score. The hazard ratio is then interpreted as the hazard ratio of one unit in the predictor. The univariable model can further be extended by adding more covariates into a multivariable model. The estimated HRs then represent adjusted effects, i.e. the hazard ratio of the variable of interest, given that all other variables are held constant.

Competing risks

When working with the Kaplan-Meier method, the assumption is that the censoring is non-informative, i.e. that the process of censoring is independent of the process of events. In many situations this assumption is not valid, e.g. when death is not the endpoint. Patients can then end the study when an event (e.g. death) other than the endpoint occur, and the event obviously hinders or alters the probability of the endpoint of interest (you cannot get a cancer recurrence if you are dead). This is known as competing risk events.[250-254] This is a common problem when working with recurrence endpoints in cancer, and is even more so when working with "early" recurrences such as local recurrences.[255] If standard Kaplan-Meier analysis is used, the estimates of risk would be biased upward, since patients that can no longer have the event of interest are treated as censored (i.e. assuming they are still under risk of the event).[256] Instead, the cumulative incidence should be computed to account

for the competing risk, e.g. by treating the competing events as another event, instead of censoring.[252,257,258]

For modeling hazards, there are mainly two alternatives: either using a cause specific Cox proportional hazards regression model, or to model the hazard of the so-called sub distribution as described by Fine and Gray.[259] The cause-specific Cox model treats the censored/competing risk events as loss to follow-up, and thus the hazard ratio should be interpreted as in a world where only the event of interest exists. This can be good for studying the overall treatment effects and for biological interpretation, but will not give the "true" hazard ratios as expected in a natural cohort, i.e. a cohort where the event is sometimes excluded due to a competing event. The Fine and Gray model, on the other hand, corresponds directly to the cumulative incidence estimate of risk, i.e. the model takes the competing events into account. Accordingly, the Fine and Gray model can only be generalized to cohorts with similar sub distributions, and where the hazard has an unintuitive interpretation, as representing the world where persons that encounter the competing events are regarded to remain under risk for the endpoint of interest for an infinite time.

Analogously, there are different approaches to test for significance, where the cause specific log-rank test is analogous to the cause specific Cox, while Gray's test is analogous to modeling the sub distributions hazard.[260] To calculate the change in hazard, cause specific Cox proportional hazards regression may be better to study biology or treatment effects, while the Fine and Gray sub distributions hazard regression gives a better estimate of the hazard in a clinical context, but is dependent on the cohort composition.[261-263] Calculating with competing risks is an important part of study III.

Interaction test

When analyzing treatment-predictive biomarkers, the goal is to find differences in the effect of a treatment on prognosis, in different groups separated by the biomarker. Doing the analysis stratified for grouping by the biomarker of interest gives descriptive information, but does not allow to directly assess the difference of effect. Importantly, it is not correct to draw the conclusion that the treatment effect differs between groups because there is a statistically significant effect in one group, but not the other. Instead, the difference in effect can be tested by creating a model with both the treatment and the biomarker as variables, and then an interaction term for the product of the treatment and biomarker. The interaction variable allows the treatment effect to be different between the biomarker stratified groups, in contrast to the standard multivariable Cox proportional hazards regression model, where the effect is estimated with the other variables kept fixed, i.e. assumed to be constant over different subgroups. The p-value of the interaction term is usually interpreted as the statistical evidence against the null hypothesis that the treatment effect is

equal between the groups, but it is important to consider that large sample sizes are needed for appropriate power to perform interaction testing.[264-267] Differences in treatment effects between groups are tested by interaction variables in study I, II, III and IV.

Discriminating performance

To assess the discriminating performance with binary outcomes, especially important for diagnostic tests and classification problems with two distinct groups, a receiver operating characteristics (ROC) analysis and the associated C-index is commonly used. The ROC analysis is made by plotting the sensitivity against 1-specificity for all possible cut-points. The C-index is defined as the probability that the test correctly orders the event times of a randomly selected pair of observations, and is a generalization to the area under the curve (AUC) in the ROC analysis. In both cases, 1.0 means perfect discriminating power and 0.5 is the performance by pure chance.[268] Sometimes it is not relevant to know the overall performance, since a cut-off with say very low sensitivity would never be used, and the performance for high sensitivity values would be more interesting. In these cases the partial AUC can be used.[269] However, the discriminating performance is often not enough to fully evaluate the tests and for survival data, the AUC at a specific time point may be of interest. Also, the clinician using the test may be interested in what the results mean for the patient taking the test, and the negative predictive value and the positive predictive value are then used. They are calculated to be the probabilities of being truly negative when the test is negative, and the probability of being truly positive when the test is positive. Both are dependent on the prevalence, or population probability, of being negative or positive.[270,271]

High-dimensional data analysis and bioinformatics

With the advancement of high-throughput methods, the field of bioinformatics, which lies in the intersection of computer science, programming, big data, machine learning and biology, has grown exponentially. An important characteristic of omics data is that it contains many variables, or features, per observation, and sometimes many thousands of features. Special consideration must be applied to prepare the raw data as produced by the machines, referred to as pre-processing.[272,273]

Pre-processing

Pre-processing of data refers to the process of taking the raw measurements from a high-throughput machine and make it ready for downstream or "high-level"

analysis. It includes quality control of the data, filtering of low quality measurements and samples, normalization of the data with or without background correction, assessment of batch effects and possibly correction, scaling of the data, and transformation of values.

Normalization and filtering

Normalization is the attempt to make the data comparable from one analysis to another, either between samples or across cohorts.[274,275] The reason for the need of normalization is that measurement values can systematically vary both due to technical/machine related sources and sample related sources, e.g. change in intensity over time or sample input amount. These systematic differences can make the analysis impossible without correction. Many normalization techniques exist and are under constant development. They can broadly be divided into techniques that shift the overall expression to be similar across samples under the assumption that most genes or peptides does not change between samples, such as quantile normalization, normalization to a common (geometric) mean or total sum of measurements, and methods that use internal controls, either spiked in references or endogenous house-keeping genes or peptides. Single sample methods, such as the SCAN-algorithm, have also been developed with the aim to improve personalized medicine workflows.[276] Advantages and disadvantages exist for all techniques. For the Illumina HT12 microarray analysis in study IV, we used quantile normalization, which has been proven to be robust for this platform.[200] For the Nanostring nCounter data in study IV, we normalized based on positive control probes for each sample (to correct for technical variation in counts over time) and for housekeeping genes (to correct for differences in sample input). For the LC-SRM data in study V we did not normalize since the input was highly standardized, and we did not want to normalize based on peptides that were selected to be different across samples.

In omics methods using image analysis, i.e. scanning machines that measure the light intensity as relative to the amount of the molecule present, and most mass spectrometry techniques, the total measurement will consist of both background intensity and the intensity from the molecule of interest. The intensity is relative in its nature and cannot be interpreted as an absolute value in the absence of explicit standards. Improvements in the analysis have been seen by subtracting the background in microarray experiments. To avoid negative values and/or missing data an offset can be applied. In the analysis of the Illumina data in study IV, we subtracted the background signal based on the negative control probes, and added an offset of 16.[200]

From the analysis of the global transcriptome or proteome it follows that a large proportion of measured molecules will not be present in the sample, and the resulting measurements for these molecules will be equivalent to the background signal. This is expected since not all genes are expressed at a certain timepoint in a

cell. However, from an analysis point of view, it is hard to tell which molecules are truly not present, and which are not measured correctly by the machine due to technical errors. The Illumina HT12 analysis outputs a p-value that is describing the probability that the transcript is expressed above background. A common pre-processing approach is to exclude all measurement above a certain p-value threshold, but that results in missing data points. Another way is to exclude a probe entirely that is expressed with high confidence in e.g. less than 80% of samples. The fundamental problem is that we do not know if the molecule is truly absent in the sample or not correctly measured. It may well be expected that, considering the heterogeneity of breast cancers, only a minority of cancers express a certain gene with important biological function. To account for this, we applied two strategies to filter probes in the Illumina data. First, we included probes with a detection p-value of < 0.01 in $\geq 80\%$ of samples, and second we included probes with detection p-value of < 0.01 in $<80\%$ samples, but with a high intensity in the measurements with high confidence.

A similar filtering procedure was made for the LC-SRM data in study V, where mProphet was used. The program filters for high confidence measurements, and we show in the supplemental data that this is highly equivalent to an intensity filter.

In an additional attempt to alleviate the problem of technical issues, an annotation package for Illumina HT12 v4 has been created, with additional information on which probes are theoretically well designed for a specific transcript.[205,277]

From my experience, the filtering based on intensity and/or confidence score of a measurement roughly filters out 50% of probes in a typical microarray experiment. To improve performance of downstream analyses, an unbiased further reduction of features can be made based on the variance of the probes, and in the Illumina HT12 data, we filtered down to the 5,000 most varying probes.

In many high-throughput analyses, the same gene or protein is measured several times by several probes or peptides. For downstream analysis and biological interpretation, a single value per gene/protein is often desired, and the combination can be made on the mean, the sum, picking the highest varying probe or other more advanced methods. For microarrays, typically the mean is better for long transcript platforms, while the most varying is better for short transcript platforms.[33] For the Illumina HT12 data, we did the analysis per probe as we wanted to retain the possibility to use the probe information for the design of the targeted assay. For the LC-MS/MS experiments we used the sum of peptide abundances as the protein abundance, and for the LC-SRM we analyzed individual peptides.

Batch effect correction

So far, the pre-processing has attempted to correct for general technical differences or individual sample differences. However, the processing of biological samples can

be very sensitive for external factors, especially the handling of RNA. Factors such as storage, extraction method, time of extraction, shipping etc. can influence the RNA on a global level. Techniques have been developed to adjust for these differences, here called "batch effects", once the important batches are known. The advantage of batch effect correction is that several batches can be studied together with many statistical analysis techniques, but the downside is that the analysis becomes cohort dependent, and there is a possibility that the correction alters the data in undesirable ways. Especially, batch effect correction with the specification of the phenotype of interest is known to produce artefacts with false positive findings.[278,279] In the Illumina HT12 data we used batch effect correction, without specifying the variable of interest, for study center and hybridization plate. For the Nanostring nCounter data, we did not use batch effect correction as we wanted to make the analysis as platform independent as possible.

Scaling of the data

The absolute expression of different genes or proteins can differ greatly, which can impact the downstream analyses by e.g. assigning greater weight to highly expressed genes or proteins. This may be undesirable, and the data may be transformed in a way so that the individual features are weighted equally. The most common transformation is to median center and scale the data by the standard deviation, also known as Z-transformation. Other methods are to median center and scale to a common min and max (used in the radiosensitivity score in study IV), or to use gene ranks instead of expression measurement (used in radiosensitivity index in study IV). Finally, before downstream analysis, values are typically log transformed to alter the usually highly skewed distribution of values, to make them proportional instead of additive, and to make the variance less dependent on absolute values.

High-level analyses

After pre-processing is completed, the "high-level" analysis starts, which is where biological or clinical conclusions are to be drawn from the data. Many concepts are similar between platforms, and high-level analyses can broadly be divided into unsupervised analysis and supervised analysis. Unsupervised analysis is characterized by the goal to find overall patterns in the data that explain the biology without pre-specifying the phenotype to characterize. The typical example is to find new subgroups of a cancer based on the data. In contrast, supervised analysis starts with a pre-defined phenotype of interest, and we typically want to find differences between phenotypes. Possible goals are to find differentially regulated pathways, or classify future samples as high or low risk of recurrence based on data with known outcomes.

Unsupervised analysis

A general feature of omics data is the high number of features (genes, peptides) per sample, and a typical start of an analysis is to understand the data by reducing the number of dimensions to examine the overall structure of the data. Common methods include multi-dimensional scaling (MDS), principle component analysis (PCA) and Sammon maps.

Principle component analysis (PCA)

Principle component analysis performs dimension reduction by finding components that maximizes the variation. The first, or leading, component (PC1) is calculated as the linear combination of all variables (loading vectors) that maximizes the overall variation. The second principal component (PC2) is calculated so that the remaining variance is maximized in a vector orthogonal to PC1. The following components are calculated in the same way.[272] PCA analysis can be used for various further analysis including regression on the components. We used PCA in study IV 1) for visualization, 2) to check for batch effects and 3) to compare the contribution by the classical clinicopathologic variables to the variation in the data.

Clustering

Clustering is broadly defined as a collection of methods to find groups of associated observations in the data. Commonly used methods include k-means clustering and hierarchical clustering. K-means clustering aims to partition the data into k pre-defined groups, while hierarchical clustering produces a dendrogram joining the two most similar observations at a time, and the number of groups can be decided at any height of the dendrogram. In study IV, we used hierarchical clustering to monitor batch effects and investigate gene clusters selected for the targeted assays.

Hierarchical clustering first calculates all the distances between the observations based on a defined distance metric, often defaulting to Euclidean distance or Pearson correlation. Then a summary metric, or linkage, for the two joined observations, which now forms an own cluster or group, is calculated. The most common types of linkage methods are *complete*, *average*, *single* and *centroid*. Next, the second most similar sample or cluster is joined.[272] Important to note is that the vertical height of the joining node in the dendrogram is describing how similar the observations are, and not the horizontal ordering, which is usually random within a cluster.

A typical hierarchical clustering on high-throughput data is made both for samples, to find related samples, and for genes/proteins to find related clusters/networks/pathways of with similar biological functions. Commonly associated with hierarchical clustering are heatmaps as a graphical representation of the data producing the clustering. Usually expression values are artificially colored

to highlight up and down-regulation of genes. Clustering of samples based on mRNA expression data has been extensively used in breast cancer, resulting in the intrinsic subtypes.[27,28] The clustering was later refined using a subset of 50 genes selected by supervised prediction analysis of microarrays (PAM).[33] To overcome the problems with this type of clustering, namely platform compatibility issues, and cohort composition, single sample predictors have been proposed, and subtyping is constantly developing with better algorithms.[280]

Supervised analysis

Supervised analysis deals with prediction in the form of regression analysis and classification, and is here described together with differential expression analysis. The goal of the analysis is 1) to find biological differences between known phenotypes, 2) to use that information to create models describing the biology, and 3) to make predictions of unknown samples.

One might speculate that increasing the number of variables to analyze by adding more genes or proteins will automatically increase the performance of the models and make the predictions stronger. Unfortunately, this is not the case and the opposite may be true, sometimes referred to as the curse of dimensionality. The reason is that many of the measured variables are not in a meaningful way correlated with the outcome of interest, introducing noise in the data. If not handled carefully, the powerful machine learning techniques available may dramatically overfit the data with consequently deteriorated performance.[272]

Supervised analysis involving classification problems must further deal with the bias-variance trade-off when selecting and training models. Put simply, the modern techniques are so powerful that they easily can find patterns in the data that separate the outcome of interest in a training dataset. However, this relies heavily on the individual observations, which introduces variability, meaning that the model will look very different depending on the training dataset. To reduce the variance, the model can be biased towards a hypothetic relationship among predictors and outcome variable, such as linear relationship, or more complex relationships. The trained model will not follow the training data as closely, with reduced variance and increased performance until a point where the model becomes too inflexible to accurately describe the true relationship. The trade-off between bias and variance is an important task in creating a good model.[272,281]

Differential expression

The first aim of many gene expression and proteomics studies is to find which gene expression or protein abundances that differ between two phenotypes. The most basic approach is to use traditional statistics to calculate a statistical confidence that the gene is indeed differentially expressed e.g. by a t-test or a Mann-Whitney test. As discussed elsewhere, this is of little importance without the effect size and an

accompanying calculation of fold-change between the two different phenotypes, which can be presented in a volcano plot. For purposes of ranking the genes, this may be sufficient, and the resulting gene list may be used for further analysis of the important features by e.g. pathway analysis. If the goal is to more conclusively determine that these genes are differently regulated, more advanced models such as linear modelling for microarrays (limma), or significance analysis of microarrays (SAM), or methods that include multiple correction testing, false-discovery rate calculations or permutation calculations, should be used.

Regression and classification

The next type of analysis is usually the prediction of new samples based on the knowledge of previous samples, today often performed through different machine learning algorithms. In this context, we have mainly used linear regression (study I, II, and IV), random forest modeling (study IV), support vector machines (IV) and k-top scoring pairs (study IV).

Standard linear regression consists of fitting a presumed near linear relationship between a single predictor variable and the outcome variable, using least squares methodology. One can easily incorporate more predictor variables in a multiple linear regression, but with more predictor variables, the risk of overfitting increases, and with the number of predictors being the number of observations -1, there is always a perfect fit. This becomes important in the high-dimensional setting. The linear model has been further extended and many other techniques may be seen as a generalization of the linear model.[272]

Decision trees are methods of segmenting predictors into regions, which can be easily visualized in a tree-like structure. However, they do not have the same performance as more advanced models, but their performance can be increased by combining many trees into a single model to yield a consensus prediction. Random forest is a tree-based method that utilizes this; the average of many decision trees reduces variance and is thus more robust than one tree alone. The method uses a way of decorrelating individual trees by allowing only a random subset of the variables in each node split, usually the square root of the total number of variables (but may be optimized using cross validation), and creates a tree of a subset of the observations, around 2/3 of the data. The performance can then be tested with the remaining 1/3 of samples, called the out-of-bag samples, which resembles cross validation. Also, the importance of each variable in the model can be calculated, and used for recursive feature elimination and feature selection, as exemplified in study IV. Random forest models are usually using hundreds to thousands of trees, and the average of the trees is used for the final model, making them much harder to interpret than a single decision tree, but usually with higher performance.[272]

Support vector machines is another machine learning method that we evaluated in study IV, but did not choose to use as random forest models performed as good, and were easier to implement. Support vector machines use a hyperplane that maximally separates the observations in different groups from each other, and further uses support vectors (vectors that are created by the closest observations to the separating hyperplane), to calculate the hyperplane. In the case that a hyperplane cannot perfectly separate the observations, or is not desirable as it may introduce extreme variance in the model, a soft margin that allows observations to be on the "wrong" side of the separating hyperplane is used, and the width of the soft margin is used as a tuning parameter (C or "slack" parameter). In addition, the support vector machine can be extended to use non-linear kernels to find even more complex patterns and decision boundaries.[272]

Top scoring pairs (TSP) is a method using simple, rule based, classifiers in the form of gene A is higher than gene B. The process of selecting a k number of top scoring pairs is called k-top scoring pairs (kTSP). The advantage of a kTSP model is that it is only dependent on the relative values within a sample, and should thus in theory be independent of the platform used, normalization and scaling, provided that the relative intensities within a sample do not shift.[282] In addition, the simplicity of the model may make it easier to understand in terms of underlying biology. The value of a kTSP approach has also been demonstrated for feature selection.[283] In study IV, we used kTSP and data from a targeted gene expression assay to train single sample predictors for risk of breast cancer IBTR after breast conserving surgery, with and without adjuvant RT.

For both random forest and support vector machine models, the tuning parameters (number of variables to input for random forest, and C-value specifying the slack or the width of the soft margin, for support vector machine) must be specified. This is a potential source of information leakage, and proper use of cross-validation must be used.

Pathway analysis

Pathway analysis aims to go beyond the findings of individual genes and proteins that differ between phenotypes, and interpret the findings in a larger biological context. Pathways in this sense is very loosely defined as a set of genes that are related in a certain biological function, and a more appropriate nomenclature may be gene sets. There are several ways to analyze genes together, of which clustering of genes has already been discussed. Other common methods include gene set enrichment analysis, overrepresentation test of gene lists, and network based analysis. Important to mention is that these methods rely heavily on previous gene annotation, and the results cannot be expected to be better than the prior biological

knowledge. However, attempts to collect and curate such gene sets are ongoing and promises to increase the accuracy of pathway analysis.[284]

In addition to finding the biological context, the use of pathways or biological concepts to analyze tumors has indeed been shown to be more robust than individual genes. There is a small overlap between prognostic genes in public datasets, but large collections of genes are correlated, and many genes capture the same biology, even to the extent that randomly selected genes perform as good as pre-selected genes, provided the number of genes is sufficiently large.[285]

Gene set enrichment methods, such as the gene set enrichment analysis (GSEA), use ranked gene lists of difference between two phenotypes, and calculate an enrichment score based on how many genes are ranked in the top or bottom of the list. The exact calculation of the enrichment scores may be different, and simpler methods have proved more successful.[286] Over-representation tests also use gene sets, and analyze the top differing genes to assess if certain gene sets have more genes present than would be expected by chance. This is commonly used together with gene ontology (GO) annotations, such as implemented in the PANTHER tool.[287] Network analysis focuses on physical or chemical interactions between molecules, known for biological experiments or *in silico* predictions, and searches for such patterns in the data.[288]

We have used the PANTHER tool in study V, and in study IV we use both clustering of genes and the correlation of these clusters with previously described breast cancer gene clusters, and calculate proliferation scores and immune scores by manually creating gene lists.[289,290]

Important considerations for high-dimensional data

Usage of high-dimensional data techniques demands certain considerations to be able to draw correct conclusions. First, proper validation cannot be overstated, and ideally in independent and prospective data. This is often not possible, and validation of models under development can be done by dividing the data into a training and a validation part, with no access to the validation data before a final model is locked. Further, independent retrospective cohorts, and already published data, can be used. Before validation in independent data, cross-validation is used to optimize model parameters, and to estimate the performance in future cohorts. Cross-validation is performed by dividing the data into k-number of folds, usually five or ten folds, meaning equally sized splits of the data. A model is then trained in k-1 folds and tested in the left-out fold. The procedure is repeated for all folds and the performance is averaged. A special case of k-fold cross-validation is with k set as the number of observations, meaning that the model is trained on all except one observation, also known as leave one out cross validation (LOOCV). Cross-

validation gives an opportunity to estimate model performance and optimize the model without information leakage before further testing in a locked dataset. Correct use of cross validation should in theory produce an overestimate of the model error, since not all data is used to train the model and the model should improve with more training observations. However, empiric studies have shown that cross validation often over-estimate the performance, which is probably due to improper use, such as training several models and use the cross-validated performance of the best one, or other types of information leakage.[291]

In study IV, we use random forest and recursive feature selection and must thus use a double-loop of cross validation, both for the selection of features, and for the performance of the model itself.

Another consequence of high dimensions is that the search space for important features increases the risk of chance findings. This becomes problematic e.g. when trying to search for peptides in databases combining different species, or in databases incorporating all possible cancerogenic mutations. Three general ways of dealing with this dimensionality problem can be used.[272] The first alternative is to use only a subset of the predictors. This can be achieved by prior knowledge of important variables, e.g. variables that have been previously reported as important in a certain setting. It can also be achieved by testing how variables perform in a known dataset, and the iterative approach of removing features from a model is known as recursive feature elimination. We use recursive feature elimination in study IV to select genes for further development into a targeted assay, ranking of features based on simple t-tests in study V, and we also use prior knowledge in study IV and V. When the data is used to select for variables, this is of course a situation of information leakage and must be properly validated, otherwise chance findings are highly likely. This has been demonstrated in a colon cancer dataset where perfect discriminating performance was found after randomly permutating the data.[292]

The second method is directed at altering the coefficients in the models to reduce variance, which is achieved by shrinking the coefficients towards zero. Lasso regression is a method that removes non-contributing predictors completely, while ridge regression keeps all predictors. The elastic net method represents a middle ground between lasso and ridge regression.

The third way is to use dimension reduction, such as PCA or MDS, and use the reduced variables for modeling.

Summary of results

This thesis consists of three main parts. Study I and II investigate the role of the putative estrogen receptor GPR30 in response to adjuvant endocrine therapy, and the signaling mechanisms of this receptor. Study III and IV consider the possibility to personalize adjuvant radiotherapy (RT) treatment after breast-conserving surgery (BCS), either by breast cancer subtype or gene expression-based classifiers. Study V explores the use of mass spectrometry-based proteomics to find biomarkers of increased risk of metastatic breast cancer after endocrine treatment failure.

GPR30 and endocrine therapy (study I and II)

GPR30 as a prognostic and treatment-predictive biomarker

In study I, we investigated GPR30 as a prognostic biomarker in early breast cancer, and as a predictive biomarker for adjuvant endocrine therapy. The study is based on tissue from the Stockholm-3 (STO-3) trial, which was a randomized controlled trial of low-risk breast cancer patients randomized to 2-5 years of tamoxifen treatment or no systemic adjuvant treatment. The original study showed a benefit of adding tamoxifen after surgery, and that the response was dependent on the expression of the estrogen receptor alpha (ER).[183] The low-risk part of the STO-3 study consisted of 1780 patients, of which 912 were sampled for tissue, and tissue microarrays (TMAs) were constructed. On the TMAs, we were able to stain and score 742 tumors for GPR30 with immunohistochemistry (IHC), both as overall staining in the cancer cells, and as plasma membrane staining (Figure 8).

First, we found that GPR30 was not prognostic for distant disease-free survival (DDFS) when analyzed as a total score (p=0.38). However, when analyzed as expression localized to the plasma membrane (PM), it was associated with a shorter DDFS (HR 1.8 95%CI 1.2–2.5, p = 0.002) in all patients. Stratifying for ER status, the results remained significant only in the ER+ tumors (HR 2.1 95%CI 1.4–3.1, p<0.001) but not in the ER- tumors (HR 1.1 95%CI 0.55–2.2, p = 0.79), although an interaction test was not significant (p=0.13), calling for caution on the subgroup analyses. In a multivariable analysis, GPR30 PM expression remained a significant prognostic factor. To further investigate the prognostic potential of GPR30, we

investigated a pre-defined group of patients with a supposedly excellent response to endocrine therapy, the ER+ and progesterone receptor (PR) positive patients, and GPR30 could still separate this group, with the GPR30 PM negative tumors having a 20-year DDFS of 91% (95%CI 84-95) (Figure 9).

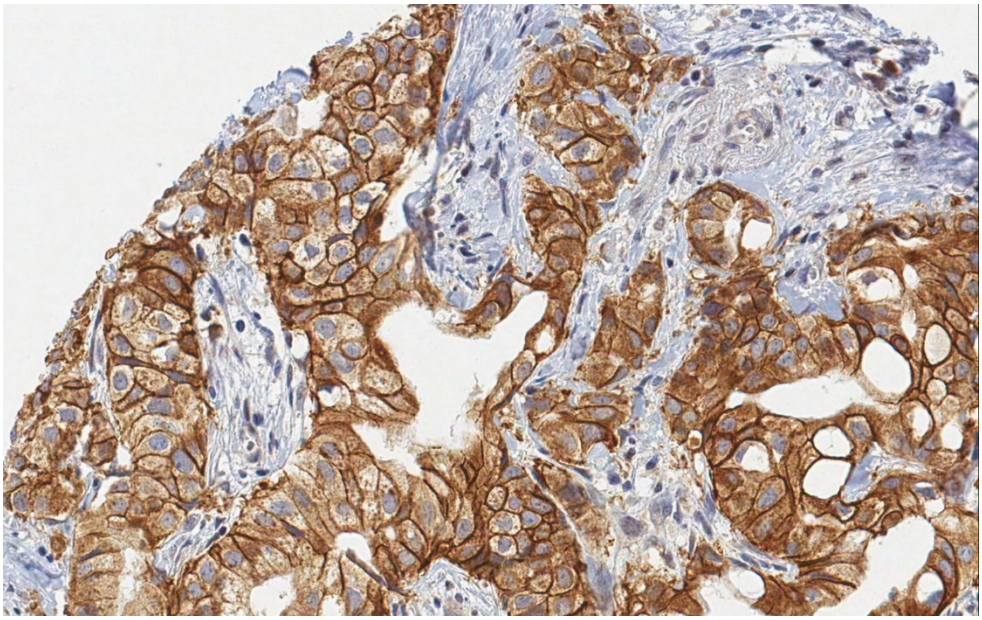


Figure 8. GPR30 staining of breast cancer cells on tissue microarray with immunohistochemistry. Strong staining in the plasma membrane is seen.

Second, we sought to evaluate if GPR30 expression, either total score or expression in the plasma membrane, was treatment-predictive of adjuvant tamoxifen treatment. However, neither of the staining scores could show any treatment prediction information in the adjuvant setting.

In study II, we investigated the mechanistic role of GPR30 in terms of its signaling, and its prognostic effect in two different breast cancer cohorts. The first cohort consisted of 273 pre- and postmenopausal patients with stage II breast cancer treated with surgery and RT as indicated, and two years of tamoxifen treatment, but no other systemic treatment. Cohort II consisted of 237 premenopausal patients with lymph node-negative breast cancer treated with surgery and RT as indicated, and little adjuvant systemic treatment (13%).

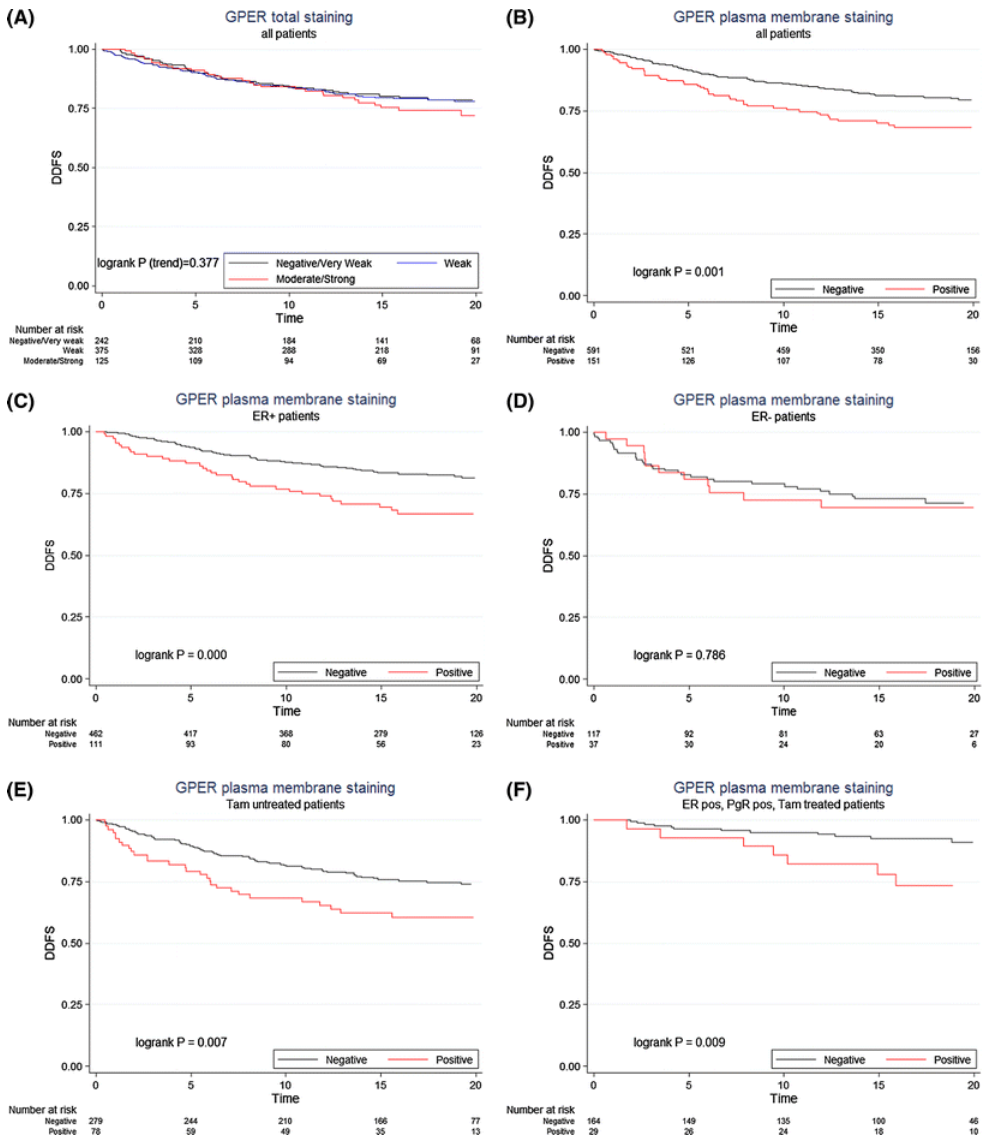


Figure 9. GPR30 (GPER) as a prognostic biomarker in the STO-3 study. Overall staining was not prognostic **(A)**. However, GPR30 in the plasma membrane was a biomarker for worse prognosis **(B)**. The results were similar among ER+ patients **(C)** and tamoxifen untreated patients **(E)**, but not for ER- patients **(D)**. Among the ER+, PR+ and tamoxifen treated patients, GPR30 could still separate patients with a long-term better vs worse prognosis **(F)**. First published in Sjöström, M., Hartman, L., Grabau, D. et al. Breast Cancer Res Treat (2014) 145: 61. Reprinted with permission.

GPR30 was evaluated with IHC on TMAs for the two cohorts. In cohort I, GPR30 expression was associated with a longer DDFS (HR 0.75 95% CI 0.60-0.94, p=0.014), and the results remained significant in the ER+ subgroup (HR 0.66 95%CI 0.49–0.88, p=0.007), but not in the ER- subgroup (HR 1.06 95%CI 0.76–1.48, p=0.7), and an interaction test showed moderate evidence for interaction between ER status and GPR30 (p=0.06). In, cohort II, we found a trend towards association with a longer DDFS (HR 0.75 95%CI 0.55-1.04, p=0.08) and again, this was the pattern in ER+ disease (HR 0.62 95%CI 0.39–0.99, p=0.05) but not in ER- disease (HR 0.97 95%CI 0.63–1.49; p=0.89), although an interaction test was not significant (p=0.16).

Apoptotic GPR30 signaling in breast cancer

As previous reports have suggested GPR30 to be pro-apoptotic, and because the prognostic effect of GPR30 appeared to be independent of tamoxifen treatment, we investigated the constitutive pro-apoptotic signaling of GPR30. Such signaling was investigated in HEK293 cells (HEK), a model system cell lacking GPR30, before and after stable transfection of GPR30. Cells expressing GPR30 (HEK-R) showed a 35% decrease in cell viability, as compared to HEK. Moreover, the HEK-R cells also had an increased release of mitochondrial cytochrome C and PARP cleavage, both markers of apoptosis, and increased ERK1/2 phosphorylation. The pro-apoptotic signaling was increased with addition of the proteasome inhibitor epoxomicin, which blocks GPR30 degradation, and thus increases the levels of GPR30. Pro-apoptotic signaling was further confirmed with visual inspection of fragmented nuclei and condensed chromatin using phase-contrast and fluorescence microscopy.

The pro-apoptotic signaling was further tested in the breast cancer cell lines MCF-7 and T47-D. As MCF-7 cells express GPR30, we changed strategy and used transient knockdown of GPR30 to study the effect of receptor expression. Doing so, we found decreased p53 expression and cytochrome C release, but no change in PARP cleavage or ERK1/2 phosphorylation. We also tested if this was affected by a GPR30 agonist (G1), and one of the active metabolites of tamoxifen (4-OH-tamoxifen). Generally, we found markers of apoptosis to be up-regulated when treating MCF7 cells with G1 or 4-OH-tamoxifen, and that the effect was weaker in MCF7 with transient knockdown of GPR30, and in T47D cells, which express less GPR30. The results should be interpreted with caution, keeping in mind that both G1 and 4-OH-tamoxifen have effects that are not mediated through GPR30. Nevertheless, the results indicate that GPR30 effects can be pharmacologically modified.

Association of GPR30 with clinical factors and possible mechanistic change

In both study I and II, the association of total GPR30 expression with other clinicopathologic variables was tested. In cohort I in study II, an association of total GPR30 was found for ER and PR, but no other variable, while in cohort II the same association was found for ER and PR, but also with age and a negative association with tumor size, histologic grade and Ki67 expression.

In study I, we also tested the association of total GPR30 and clinicopathologic variables. Again, total GPR30 was associated with ER and PR, but interestingly the GPR30 expression in the PM was negatively associated with ER and PR, and positively correlated with histological grade. This led us to investigate in more detail how the clincopathologic variables vary with different levels of GPR30 expression. Indeed, a biphasic correlation was found, meaning that the most extreme values had the same association. We also redid the analysis for cohorts I and II in study II, and the results were strikingly similar, showing that both the GPR30 negative tumors, and the tumors most highly expressing GPR30, were associated with a more aggressive phenotype (ER-, PR- and histologic grade 3, Figure 10).

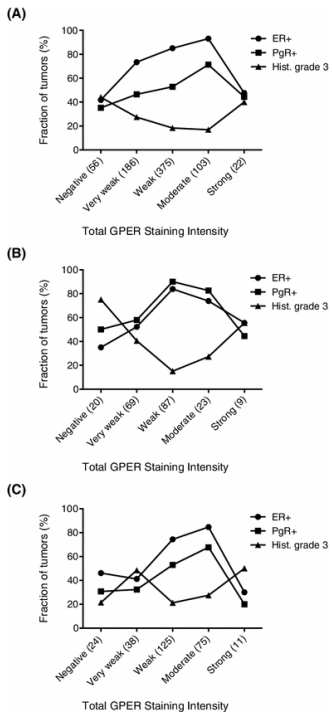


Figure 10. Biphasic relationship of GPR30 staining intensity with estrogen receptor alpha expression (ER+), progesterone receptor expression (PgR+) and histological grade 3 in three cohorts: **A)** study I, **B)** study II cohort II and **C)** study II cohort I. First published in Sjöström, M., Hartman, L., Grabau, D. et al. Breast Cancer Res Treat (2014) 145: 61. Reprinted with permission.

This may well explain why the literature has been so inconsistent regarding GPR30 as a prognostic marker for better or worse prognosis. From our data, the role of GPR30 appears to be highly dependent on how the data is split in groups.

Taken together, several observations in study I and study II would be consistent with a hypothesis that GPR30 is expressed in normal cells and decreases in expression during cancer transformation and acquisition of more non-differentiated phenotype in most cells. This is also the case for e.g. ER.[293] However, in a minority of tumors, an activating event, may it be a mutation, amplification or post translational modification, alters the function of the receptor, makes it shift signaling, and localize to the plasma membrane. This is also supported by the high staining of normal cells and cancer *in situ*, which generally are highly expressing GPR30, with a marked decrease when the cancer become invasive (Figure 11). Second, this would explain the biphasic relationship with other clinicopathologic variables, and markers of an aggressive phenotype. Third, the fact that the expression of GPR30 is only prognostic in the STO-3 trial when expressed in the plasma membrane, and confers a worse prognosis, suggests that the receptor shifts function, localizes to the plasma membrane, and potentially drives the cancer. Although speculative, this hypothesis merits future investigations.

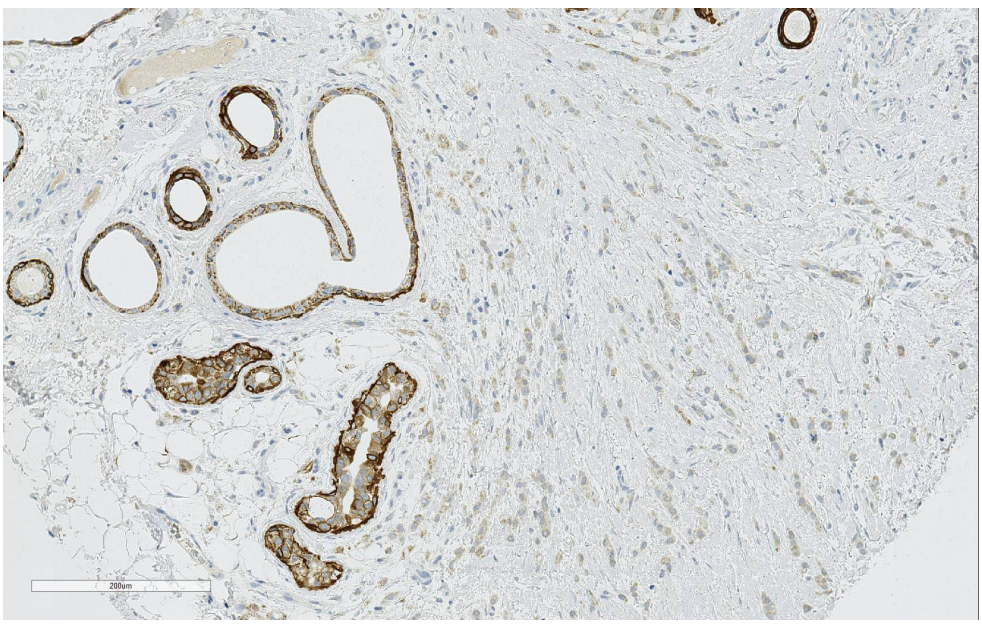


Figure 11. GPR30 staining in cancer *in situ* and in surrounding infiltrating tumor cells. A clear decrease in GPR30 expression is seen in the cancer cells, compared to cancer *in situ*.

Personalized radiotherapy (study III and IV)

In study III, we investigated if the breast cancer subtypes respond differently to adjuvant RT after BCS, and could thus be used as a treatment-predictive tool. We further evaluated the effect of RT in the absence of systemic adjuvant treatment in a presumed low-risk group of breast cancer, similar to the groups that are being investigated in de-escalation trials. In study IV, we continued the work to find a predictor for response to adjuvant RT on the gene expression level, with the aim to create prognostic signatures for ipsilateral breast tumor recurrence (IBTR), and investigate the biologic details behind gene expression predictors for radiosensitivity.

Breast cancer subtype and the response to adjuvant radiotherapy after BCS

Study III is based on the SweBCG91-RT randomized trial, which is a trial of BCS +/- RT. Patients with lymph node negative stage I-II breast cancer were included. Important to note is that systemic adjuvant treatment was given according to regional guidelines at the time, meaning that the patients in SweBCG91-RT were largely systemically untreated (92%). From the 1,178 patients in the trial, we could retrieve tissue and construct TMAs from 1,003 tumors. Tumors were stained for ER, PR, HER2 and Ki67, and surrogate subtyping was performed according to the St. Gallen 2013 guidelines, with the modification that luminal HER2+ and non-luminal HER2+ tumors were grouped together, due to group sizes (Figure 12). The effect of RT on the 10-year cumulative incidence of IBTR as first recurrence was studied as the primary endpoint with a competing risk analysis. In addition, all recurrences, breast cancer death and death from any cause were studied.

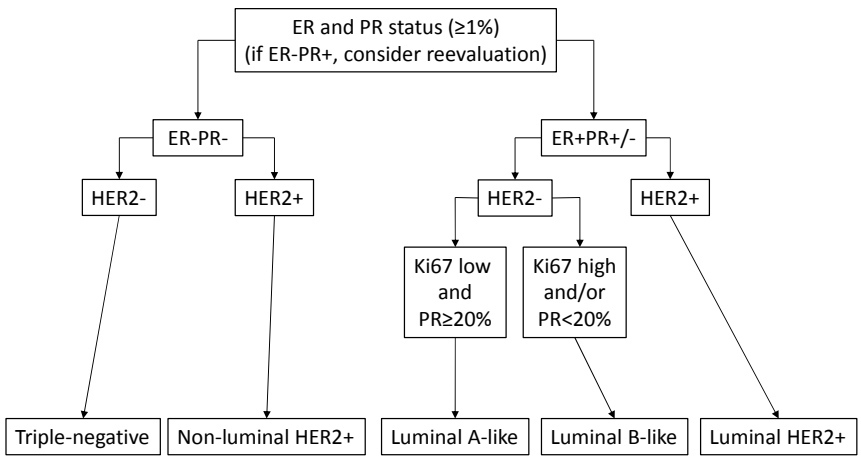


Figure 12. Algorithm for surrogate breast cancer subtyping used in study III. ER – estrogen receptor, PR – progesterone receptor, HER2 – human epidermal growth factor receptor 2.

RT reduced the cumulative incidence of IBTR at 10 years for all subtypes except the HER2+ groups (Luminal A–like tumors: 19% vs 9%, HR 0.46, p = 0.001; Luminal B–like tumors: 24% vs 8%, HR 0.33, p=0.001; triple-negative tumors: 21% vs 6%, HR 0.25, p= 0.08; HER2+ 15% vs 19%, HR 1.29, p = 0.6), but an interaction test did not support an overall difference in effect (p=0.21). The results were similar when analyzing all recurrences, except that Luminal B-like tumors had less effect of RT (Luminal A–like 26% vs 14%, HR 0.50, p=0.001; Luminal B–like tumors 29% vs 23%, HR 0.76, p=0.3; triple-negative tumors 38% vs 15%, HR 0.35, p = 0.03; HER2+ 30% v 30%, HR 1.0, p = 1.0). These results are in line with the idea that Luminal B tumors have a higher risk of distant metastasis than Luminal A tumors, and that RT is less effective to prevent distant metastasis than IBTR. For the endpoints breast cancer death (BCD) and death from any cause, no subtype had a significant effect from RT (except the triple-negative that were borderline significant for endpoint BCD), which is consistent with the entire study, as well as the literature, where a large meta-analysis was needed to obtain the required power to demonstrate benefit from RT on survival.[102]

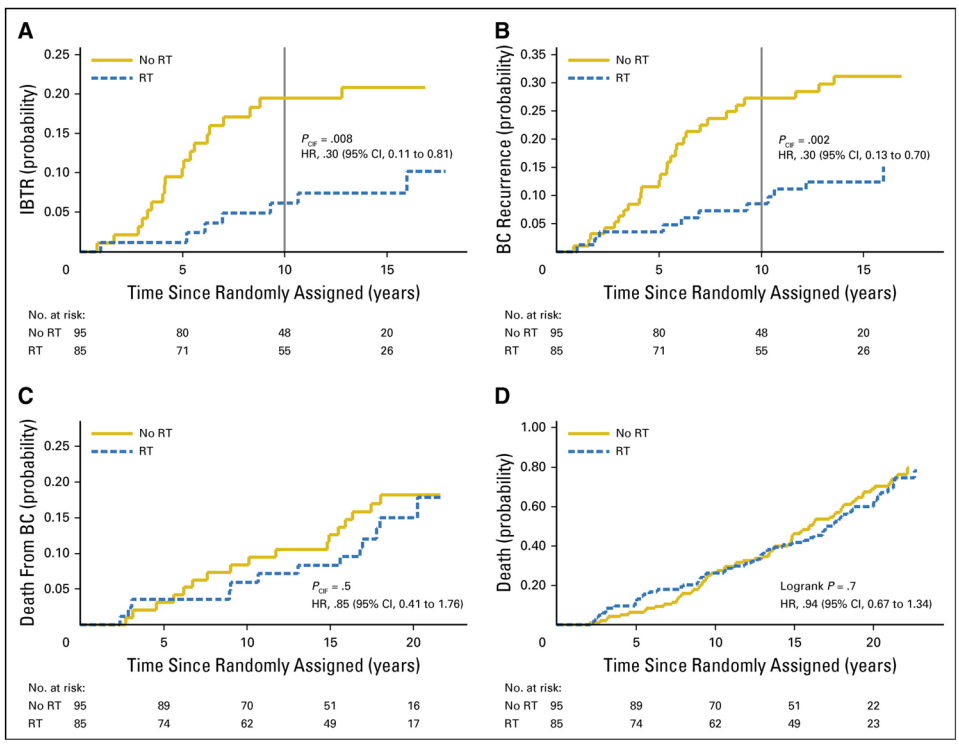


Figure 13. Effect of adjuvant whole-breast radiotherapy (RT) for a clinical low-risk (N0, age ≥ 65 years) luminal A–like group on cumulative incidence of **(A)** ipsilateral breast tumor recurrence (IBTR), **(B)** any breast cancer (BC) recurrence, **(C)** breast cancer death, and **(D)** any death. CIF, cumulative incidence function; HR, hazard ratio. First published in: Martin Sjöström et al. JCO 2017, 35, 3222-3229.Copyright © 2017 American Society of Clinical Oncology. Reprinted with permission.

Effect of adjuvant radiotherapy in a presumed low-risk group of breast cancer

The effect of RT was tested in a pre-defined low-risk group of breast cancer (lymph node-negative, > 65 years of age and Luminal A-like). The effect of RT on cumulative incidence of IBTR as first event at 10 years was excellent (20% vs 6%, HR, 0.30, p = 0.008) but no effect was seen on BCD or death from any cause (Figure 13).

Development of a targeted breast cancer gene expression radiosensitivity assay

In study IV, we aimed to develop a gene expression assay with classifiers prognostic for IBTR after BCS, and treatment-predictive for adjuvant RT (Figure 14). We collected fresh frozen breast cancer tissue samples from three health care regions in Sweden. The samples were collected in a case-control fashion with cases being defined as patients later suffering from an IBTR, and controls as being recurrence-free for at least the same time as the matched case. The patients were matched for ER status and RT treatment, and the analyses were performed stratified for ER status and RT. The aims were to identify a low-risk group of patients that do not need any addition of RT, and a group that is radioresistant and would need additional treatment. These aims can be thought of as creating a "radiation omission" and a "radiation intensification" signature. We hypothesized that the biology governing these signatures would not be the same, and therefore decided to use the RT- tumors as the group in which to develop the "radiation omission" signature, and the RT+ tumors as the group in which to develop the "radiation intensification" signature. This resulted in the analysis of four separate groups: ER+RT+, ER+RT-, ER-RT+ and ER-RT-.

In all, we were able to collect 336 fresh frozen tumors. Total RNA was extracted, and we noted that samples from one of the centers had significantly lower RNA quality than the others. Since targeted, or more robust assays, are required for the analysis of lower quality RNA, we decided to use the higher quality RNA samples as a discovery cohort (N=172), transfer the results from the discovery cohort to a targeted assay, and use the lower quality RNA samples as a validation cohort (N=164) with the targeted assay.

The first step was to select the panel of genes to use for a targeted assay. To that end, the discovery cohort was analyzed with the Illumina HT12 v4 whole transcriptome microarray. A random forest model with cross-validation and recursive feature elimination was used for each of the four groups to rank the top discriminating genes for IBTR. For the groups were discriminating potential was found (ER+RT+, ER+RT-, ER-RT-), the top 50 genes, respectively, were selected for the targeted assay.

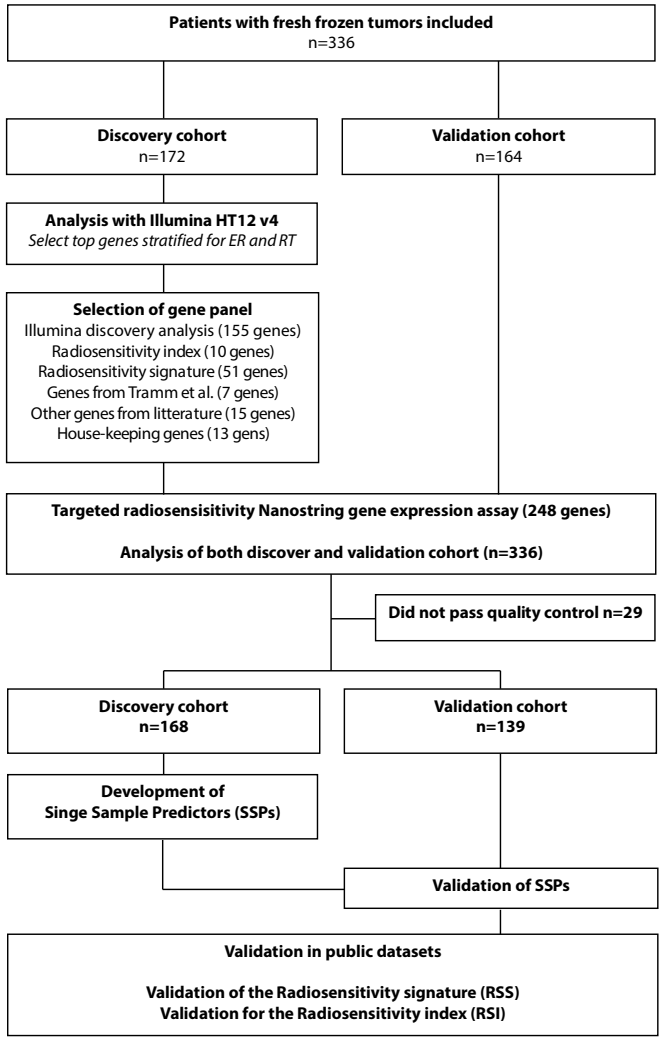


Figure 14. Overview of study IV. Patients were included as cases and controls, and divided into a discovery cohort and a validation cohort. The discovery cohort was analyzed with a whole transcriptome analysis (Illumina HT12 v4), and genes were selected for a targeted panel. All samples were rerun using the targeted assay. Singe Sample Predictors (SSPs) for risk of IBTR, and treatment stratification, were developed in the discovery cohort and tested in the validation cohort. Finally, the SSPs were validated in public datasets, and previously published signatures were tested in our data.

To the genes selected from the discovery cohort analysis we added genes from three previously described radiosensitivity signatures,[147,154,294] genes from the literature reported to be associated with radiosensitivity or biology relevant to radiobiology such as DNA repair, apoptosis and hypoxia, and house-keeping genes for normalization purposes. The final panel included 248 genes.

The next step was to select a platform for the targeted assay. For clinical use, the ability to analyze samples that are formalin-fixed and paraffin-embedded (FFPE) is important. We therefore chose the Nanostring nCounter platform, which is already FDA approved for use with clinical FFPE samples. All samples from the discovery cohort and the validation cohort were analyzed using the custom designed nCounter assay.

Single Sample Predictors for predicting ipsilateral breast tumor recurrence

With the data from the targeted assay, we created signatures for classifying a sample as a case (high risk of having an IBTR) and control (low risk of having an IBTR) for each of the four groups. Data was normalized for technical and sample factors, and probes and samples were filtered for quality. As expected when analyzing data from several centers with different handling and storage (e.g. embedded in OCT embedding media), and even more with one center having samples of lower quality, batch effects were observed. One strategy to overcome this problem is to correct for batch effects (as was done for the Illumina HT12 data), but that risks making any final predictor/classifier cohort specific, and requires re-training or re-calibration for each new dataset. Instead, we chose a Single Sample Predictors (SSP) approach, using the k-top scoring pairs (kTSP) algorithm, a method that in theory should be cohort independent and only analyze the relative expression values within a tumor.[282,283] Four classifiers, one for each group, were trained in the data from the discovery cohort using the targeted assay, and the results were validated in the corresponding group in the validation data.

The SSPs were prognostic for IBTR in the ER+RT+ group (p=0.005, AUC 0.67 95%CI 0.56-0.78), the ER+RT- group (p=0.015, AUC 0.89 95%CI 0.73-1) and in the ER-RT+ group (p<0.001, AUC 0.78 95%CI 0.58-1). We could not test the predictor in the ER-RT- group, as there were too few samples remaining for meaningful analysis (N=3).

In addition, we tested the SSPs in two publicly available breast cancer datasets: the data by Servant et al., and the data by van de Vijver et al.[150,295] In the Servant dataset, all patients were treated with RT, while a majority was treated in the van de Vijver dataset, and we therefore only could test the SSPs trained in the RT+ data. The ER+RT+ SSP was prognostic both in the Servant dataset and the van de Vijver dataset (p=0.006, AUC 0.62 95%CI 0.54-0.69 and p=0.001, AUC 0.69 95%CI 0.63-0.76, respectively). The ER+RT- SSP was prognostic for IBTR in the Servant data (p=0.006, AUC 0.74 95%CI 0.61-0.86) but not in the van de Vijver data (p=0.85 AUC 0.48 95%CI 0.33-0.62).

Combination of Single Sample Predictors for treatment prediction

So far, the SSPs were validated to be prognostic for IBTR in the respective groups. However, this requires that we know if the patient was (or will be) treated with RT

to apply the correct model. The goal is instead to be able to stratify the patients before a treatment decision. To that end, we hypothesized that the SSPs developed in the RT- tumors could be used to determine if the patient needs addition of RT, which would be equivalent to a "radiation omission" signature. If predicted low risk, the patients would be stratified to the "No RT" group. If the patient was predicted to be at high risk of IBTR, the SSP developed in the RT+ patients was used, which we hypothesized could determine if the patient is at high risk of IBTR when given RT. If predicted low-risk by the second classifier, the patient would be stratified in the "Give RT" group, while if at high risk of IBTR after RT, the patient would be assigned to the "More treatment" group (Figure 15).

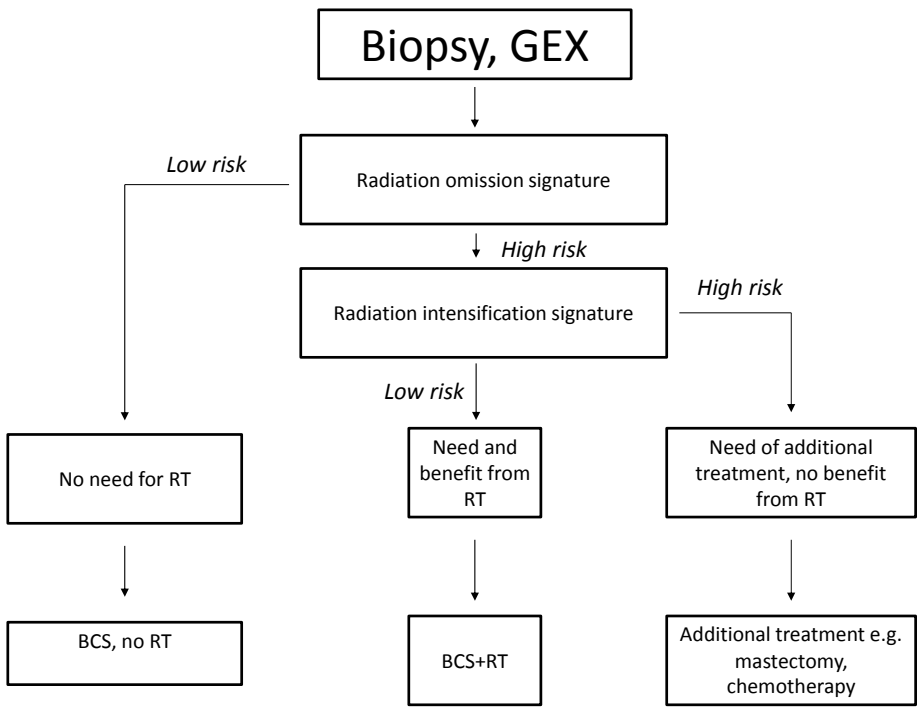


Figure 15. Proposed combination of Single Sample Predictors for radiation omission and radiation intensification to stratify patients for adjuvant radiotherapy. GEX – gene expression profiling, BCS – breast-conserving surgery, RT- adjuvant radiotherapy.

We applied this sequential algorithm separately for the ER+ and ER- tumors, and assessed the effect of giving RT in the predicted groups. For the ER+ tumors, the "No RT group" had no benefit of RT (p=0.4), but did not have a low risk of developing IBTR (25% at 10 years). Notably, these numbers are hard to evaluate as this study is enriched for patients suffering IBTR, and not representative of a consecutive breast cancer cohort. The group predicted as "Give RT" on the other

hand had an excellent effect of RT (p<0.001). The group predicted as "More treatment" did not have an effect of RT (p=0.4), and a substantially higher risk of IBTR than the "No RT" group (55% at 10 years). Again, these number should be interpreted in the context of the current study only (Figure 16).

For ER- tumors, only two were RT untreated, and we could thus only investigate the prognostic effect in the whole group. The "More treatment" group had a significantly higher rate of IBTR than the patients predicted in the "No RT" and "Give RT" groups (p<0.001) (Figure 16).

Although the analysis in study IV is exploratory, the conceptual idea show promise in stratifying patients for RT treatment.

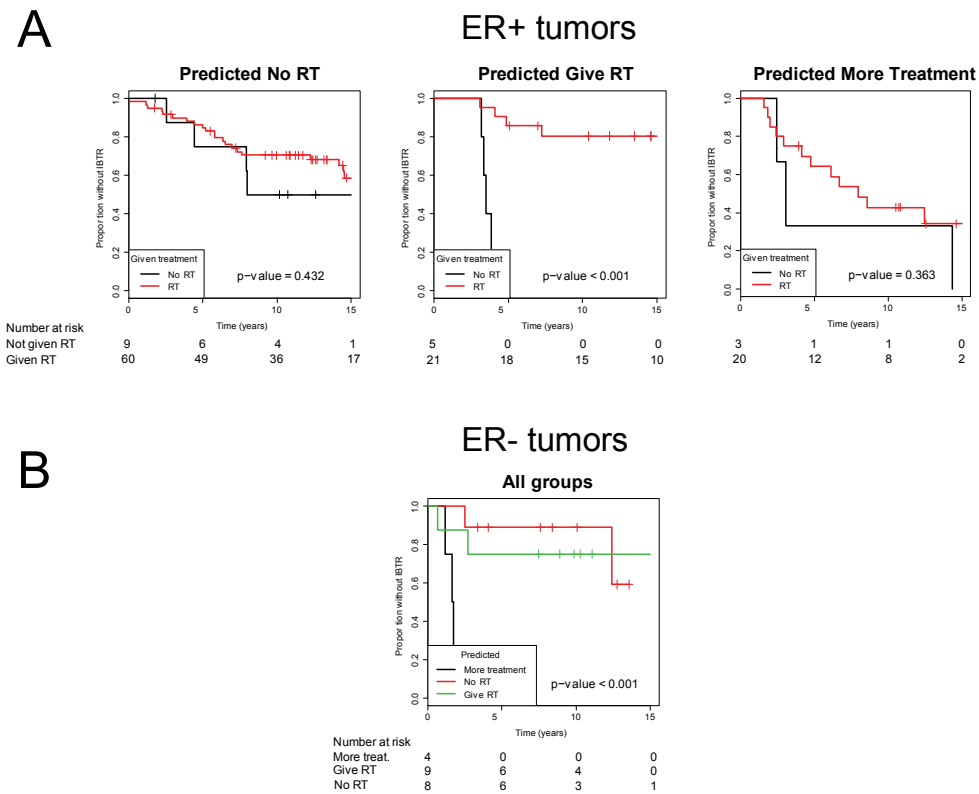


Figure 16. Effect of radiotherapy in the groups as predicted by combining Single Sample Predictors for radiation omission and radiation intensification in ER+ tumors **(A)**. For ER- tumors, only two were radiotherapy untreated and we thus only assessed the prognostic potential **(B).**

The underlying biology behind radiosensitivity predictors

The SSPs were developed in different groups of breast cancer, and the aim was to answer different biological questions about the tumors. In addition, we saw a clear difference in performance for different subgroups when testing the two public profiles (the radiosensitivity signature (RSS) and the radiosensitivity index (RSI)), which has also been shown previously by the authors.[296] To this end, we correlated the scores from the profiles with proliferation and overall immune response, two of the most important biological determinants of breast cancer outcome. The SSPs developed in ER+ tumors showed a correlation with proliferation, while SSPs developed in ER- tumors were correlated with immune response. The RSS was correlated with proliferation, and negatively with immune response. RSI on the other hand, was negatively correlated with both immune response and proliferation. Interestingly, the original authors of RSI recently showed the same correlation with immune activation.[296]

Proliferation is the major determinant of prognosis in ER+ disease and immune response is a major prognostic factor in ER- disease.[297] Thus, considering how the signatures correlate with biology, the subgroup specific performance seems logical. The RSS, correlated with proliferation, performed best in ER+ disease, while the RSI, correlated with immune response, performed better in ER- disease. In fact, the tumors that score high in RSI, and thus have a worse prognosis and/or are radioresistant, are slowly proliferating and have a low immune response. Indeed, the ER+ untreated tumors that are predicted to be at high risk, have the better prognosis.

Taken together, the underlying biology that the signatures capture may well explain why they perform differently in ER+ and ER- disease, and this should be considered when planning future studies. Further, it strengthens the rationale to create subgroup-specific classifiers and predictors in breast cancer.

Protein biomarkers for distant recurrence (study V)

In study V, we explored a method of combining two conceptually different mass spectrometry techniques to analyze proteins, namely shotgun LC-MS/MS and targeted LC-SRM (Figure 17). The method was applied to the N-glycosylated subproteome of a set of 80 primary tumors from patients with primary breast cancer that underwent surgery 1983-1991. The patients were treated with RT as indicated, and treated with two years of adjuvant tamoxifen.

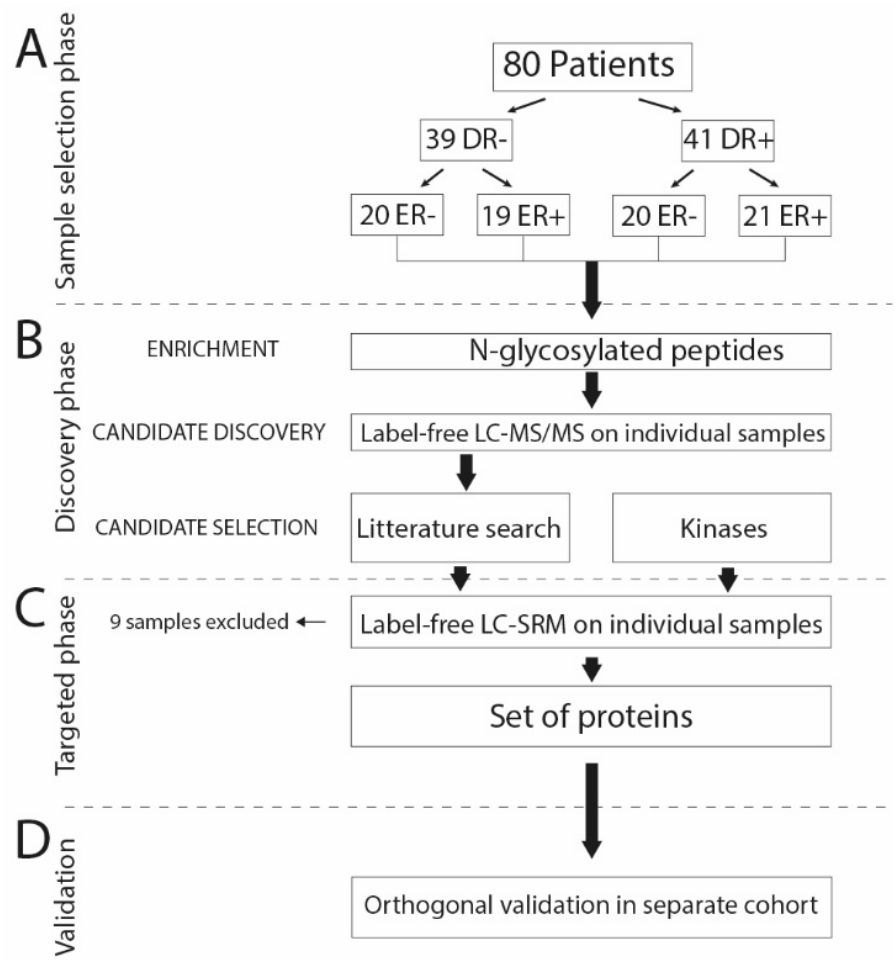


Figure 17. Overview of study V. DR – distant recurrence, ER – estrogen receptor, LC – liquid chromatography, MS/MS – tandem mass spectrometry, SRM – selected reaction monitoring. First published in Sjöström et al. J. Proteome Res., 2015, 14 (7), pp 2807–2818. Reprinted with permission.

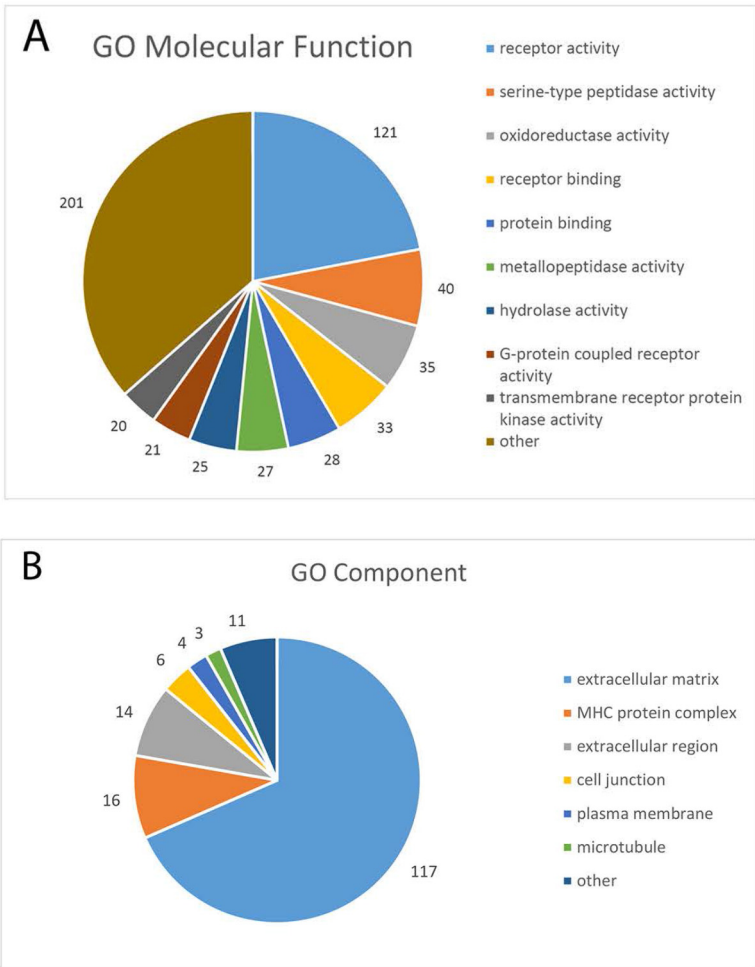


Figure 18. A gene ontology analysis using the online tool PANTHER, for the N-glycosylated proteins identified in the discovery phase of study V. Results are shown for molecular function **(A)** and cellular component **(B).** First published in Sjöström et al. J. Proteome Res., 2015, 14 (7), pp 2807–2818. Reprinted with permission.

In a first phase of the study, we did a label-free shotgun LC-MS/MS analysis of the samples. The major reason for this analysis was to identify proteins present in the samples, and their peptide composition and properties to develop the targeted assay. The LC-MS/MS shotgun approach is a good method of identifying peptides and proteins present in a sample, but it lacks in quantitative accuracy. With stringent filtering criteria, we could identify 1,515 N-glycosylated peptides from 778 proteins. To confirm the enrichment for N-glycosylated peptides of relevant biology, a search with the PANTHER gene ontology tool was made and the molecular function and cellular components of the proteins were assessed.[287] The

proteins showed the expected enrichment for receptors, signaling molecules and components of the cell surface and extracellular matrix (Figure 18).

In the second phase, we developed a 92-protein-plex LC-SRM assay. The candidate proteins selected for the targeted assay were based on several criteria. First, we did an overview differential expression analysis on the LC-MS/MS, although the results should be interpreted with statistical caution, as the quantitative accuracy is suboptimal, and the data suffers from a high fraction of missing data. Second, we added kinases, which are important signaling molecules that can drive cancer progression.[298] Finally, SRM assays were already developed for several important proteins in breast cancer biology, and we included those in the assay. In total, the panel consisted of 155 proteotypic peptides from 92 proteins.

The samples were then re-run using the targeted LC-SRM assay. To assess the clinical accuracy and validity, we specifically assessed the HER2 protein, from which two peptides were chosen for the targeted LC-SRM. At the time of surgery HER2 evaluation was not routine, and we therefore only had clinical data for 28 of the patients. Of these 28 patients, we detected at least one HER2 peptide in 18 tumors, and the correlation for the measurements of the HER2 peptides with clinical status was perfect, except for one case which had a very high abundance measured for both HER2 peptides, but was clinically negative (Figure 19). Not detecting a HER2 peptide in a sample could be because the peptide was below the level of detection, and all samples where we did not detect a HER2 peptide, and had clinical data, were clinically negative.

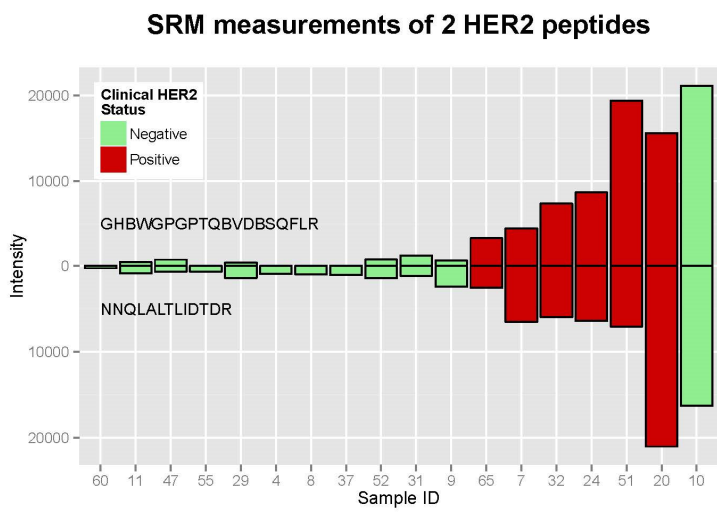


Figure 19. Selected reaction monitoring (SRM) measurements of two HER2 peptides compared to clinical status, as reported from the pathology laboratory. First published in Sjöström et al. J. Proteome Res., 2015, 14 (7), pp 2807–2818. Reprinted with permission.

We further assessed the differential regulation between patients with or without distant recurrence, stratified for ER status, with a Wilcoxon rank-sum test. At the 0.05 significance level, 10 peptides from 10 proteins were significantly regulated and interestingly, only 1 was significant in both ER+ and ER- tumors. Moreover, 15 peptides from 11 proteins were significantly regulated between ER+ and ER- tumors, strengthening the rationale to analyze them as separate groups. A literature search showed that most of the discovered proteins were known to be important in cancer biology, but some of the proteins were also previously not reported in cancer biology.

Finally, we performed orthogonal validation of the candidate proteins in publicly available gene expression data using the online tool Kaplan-Meier plotter.[299] Good quality probes were available for analysis for 9 out of 10 proteins, and 5 were significantly regulated also at the gene expression level. The patterns with ER status specificity was largely reproduced for these proteins at the gene expression level.

Discussion

In the following discussion I will comment on the main results and some of the methodological considerations that have been highlighted during our work towards this thesis. The main results concern the use of endocrine therapy and radiotherapy (RT) in an adjuvant setting, and possible ways to personalize the treatment. Further, the work goes into detail regarding the role and functioning of the putative estrogen receptor GPR30 in breast cancer. From a methodological perspective, I have focused much on the data analysis, and many aspects are important for the interpretation of the work, and worth discussing, such as patient and sample selection, subgroup analysis, cut-point definitions, study endpoints and multiple hypothesis testing. Finally, study IV and V aim to select features and creating targeted assays for clinical use, and I discuss demands for such assays.

Personalized endocrine therapy or radiotherapy

Endocrine therapy and RT are widely used adjuvant treatments for primary breast cancer. Both are proven to reduce the risk of recurrence, and risk of breast cancer death. In the attempts to personalize treatment, the trend is towards de-escalation of RT (e.g. by avoiding RT to a low-risk group of patients), and sometimes escalation of endocrine therapy, such as prolonging the treatment to ten years (although other personalized approaches are also used, such as tamoxifen and GnRH analogue in younger women, and switching AI to tamoxifen in older women). In addition, endocrine therapy is attracting increased attention for a role in prevention of breast cancer.[19,300] This is interesting, since although RT may appear like a more invasive treatment than endocrine medications, the side-effects of endocrine therapy may be severe and followed by a low adherence to the treatment, and many women may prefer a treatment spanning weeks instead of years.[301,302] Based on the results we present in study III, this is worth discussing. We report that a proposed low-risk group of breast cancer patients, similar to the groups that are part of the de-escalating trials, have a very good effect of adjuvant RT for local control, and we propose that the low-risk group instead could be spared endocrine therapy in selected cases. It must be stressed that the SweBCG91-RT study, which our study is based on, differs from modern treatment in that other systemic treatments were

used sparsely, and it is important to remember that endocrine therapy exerts its effect systemically throughout the body. Although RT prevents breast cancer deaths in meta-analyses, we observed no effect on breast cancer death, nor overall survival, in our study. A pre-requisite for only administering RT is that the risk for distant recurrence is sufficiently low.

Our results are in line with previous studies that show an effect of adding RT to endocrine treatment, and that the effect on reduction of ipsilateral breast tumor recurrence (IBTR) is similar between endocrine therapy and RT, or that RT is superior.[303-308] In our study, we have a unique opportunity to study the effect of adjuvant RT as a single adjuvant treatment, and the results suggest it is highly effective for local control. Further, the rate of IBTR in this supposedly low-risk group is high without RT. Interestingly, some studies on the other hand indicate that the risk is sufficiently low in the untreated patients to avoid RT, but importantly those were treated with endocrine therapy and still have limited follow-up.[309] The follow-up time becomes even more important in the low-risk group, since they tend to have a longer interval to recurrence.

Taken together, I believe that the omission of endocrine therapy in favor of RT is not to be recommended based on the limited documentation of effect of RT on distant events. However, in cases where side-effects, compliance, age and patient preferences are against the use of endocrine therapy, RT should be considered. Indeed, it is worth noting that RT spans 3 to 5 weeks instead of 5 to 10 years, and that endocrine therapy is accompanied with side-effects and an adherence rate of 50% to 80%.[301,302] These options should be discussed with the patients, and in the end, the patient should decide which treatment effects are worth the side-effects. If RT is to be avoided, the patient should be informed that the risk of an IBTR may be substantial. This discussion highlights the need for better understanding of which patients will respond to radiotherapy, and ultimately true treatment-predictive biomarkers.

A low-risk group that may be spared treatment?

Attempts to further stratify patients for endocrine therapy have shown that high-risk and low-risk groups within ER+ patients have an effect of tamoxifen, regardless of risk score by transcriptomic tests.[310] In study I and II, we also show that GPR30 is not a treatment-predictive biomarker for endocrine therapy. This raises the idea that if differences in treatment response cannot be detected, an alternative strategy may be to identify a group with so low baseline risk that that the absolute effect of treatment is negligible, although the relative reduction may still exist. This is the basis for the use of proliferation markers and gene expression tests for treatment stratification of chemotherapy. In study I, we examine the use of GPR30 to define such a low-risk group, with promising results in the tamoxifen-treated setting with possible implications for choice of chemotherapy or RT. Interestingly, the same

patient cohort was recently used to define such a low-risk group by gene expression signatures, with similar results to our GPR30 study, but importantly also the finding of a low-risk group among the systemically untreated patients. This may indicate that for some patients, surgery alone could be a treatment strategy.[311] However, in the already low-risk Stockholm-3 trial, the proportion of patients that is considered "ultra-low risk" is 15% by gene expression signatures, and 22% by the classical markers, tamoxifen treatment, and lack GPR30 plasma membrane staining, suggesting that further optimization of the classifiers is needed to better define the low-risk group. Conversely, in study III, the proposed low-risk group defined by clinicopathologic markers did not have a low risk of IBTR. It would be highly interesting to investigate if molecular high-throughput techniques could be used to define a low-risk group also among these patients. In summary, the identification of a group of patients that can be treated with surgery alone is very interesting, but so far only small subset analyses have been presented, and more studies are needed before we safely can recommend surgery alone.

Interesting to note is the overall high risk of recurrence, especially IBTR, reported in study III and in other studies from the same time period. Modern treatment outcomes tend to be better, and this may be because of both better treatments but also earlier detection. It will be interesting to follow the long-term results from current de-escalation trials, and if the promising early results will hold also long-term.

Gene expression-based predictors for personalized adjuvant radiotherapy

Several groups have attempted to create gene expression-based predictors for response to adjuvant RT after breast-conserving surgery (BCS). One of the first reports came from Kreike et al. who used a small number of patients (19 with later IBTR vs 31 without later IBTR) and found no evidence of overall differences between tumors that developed, or did not develop, IBTR.[153] However, the same group simultaneously published that refinement of already established gene expression signatures were prognostic of IBTR.[151] Niméus et al. published that a support vector machine classifier on the global gene expression profile could differentiate between ER+ tumors with or without later IBTR in LOOCV, but the top discriminating genes could not be validated in follow-up studies.[146] The group of Kreike et al. continued the work and presented a cross platform validated signature for IBTR mainly based on cell proliferation, but again, the sample size was relatively small (56 vs 109 samples).[152] The authors tried to validate the signature in a larger cohort (343 samples), but were not able to reproduce the performance in this cohort and concluded that there are no significant gene expression patterns for the risk of IBTR.[150] In separate attempts, Eschrich et al. validated the previously created radiosensitivity index (RSI), a radiosensitivity classifier based on the relative ranks of ten genes, also in breast cancer.[147] The

performance was later confirmed only in ER- tumors.[148] The group has since that combined the RSI with the linear quadratic model to a genomic-adjusted radiation dose (GARD), and also showed the signature to be correlated with immune response.[149,296] Speers et al. have presented another model, based on a 51-gene random forest model, that appears prognostic for IBTR in RT treated patients.[154] Finally, Tramm et al. have also proposed a four gene classifier in the post-mastectomy setting, with promising performance, but so far without further validation.[294]

Taken together, promising results have been presented but definitive validations have been elusive. In study IV, we tried to address some of the problems remaining before introducing gene expression-based classifiers in the clinic, such as platform and pre-processing independence, handling of lower quality samples, biological basis for the classifiers, determining the performance in subgroups of breast cancer, and further validation of previously presented signatures. However, neither our newly proposed signatures, nor any of the previously published signatures, reach the needed accuracy for changing practice of administering RT. Part of the problem is the low power associated with the low number of samples and events we and others use, and further the non-randomized nature of the cohorts. In practice, the lack of high-quality samples and long-term follow-up is a major obstacle, and the search for high-quality materials is ongoing, and we are awaiting the definitive test of the signatures in large cohorts.

Eventually, it may turn out that the biological drive for developing local recurrence is too small compared to the randomness of the recurrence process or response to RT. In combination with the effectiveness, and relatively mild side-effects with modern RT, perhaps we must conclude that some patients will be over-treated and instead use the baseline risk in the discussion with the patient, and make the patient involved in the decision to receive therapy. Here, gene expression tests may prove very valuable to determine the baseline risk.

GPR30 in breast cancer biology

We investigated the role of GPR30 in endocrine resistance of adjuvant treatment for early breast cancer and found no support for a treatment-predictive role. However, the adjuvant setting may be different from the metastatic setting, and the *in vitro* experiments have shown that GPR30 appears to be up-regulated during prolonged treatment with tamoxifen.[87] This may be more similar to the metastatic setting where tumors acquire treatment resistance during treatment, and investigations specifically in that setting would be very interesting. A study in a consecutive sampling of the

same tumor during tumor progression, and preferably under endocrine treatment, would be an ideal study setting.

A very interesting hypothesis derived from our data, and with some evidence in the literature, is the possibility that GPR30 is generally downregulated during breast cancer progression, but changed in a minority of tumors with shifted signaling and increased localization to the plasma membrane (Figure 20). Recently, more evidence for the first part was presented as hyper-methylation and down-regulation of GPR30 occurs with breast cancer progression.[312] Also, prolonged treatment with tamoxifen has been reported to up-regulate GPR30 at the cell surface.[87,101] Further studies of GPR30 in breast cancer is warranted, and in parallel, the basic mechanism of receptor function must be elucidated. A possible mutation or amplification could be searched for in the growing amount of publicly available data.

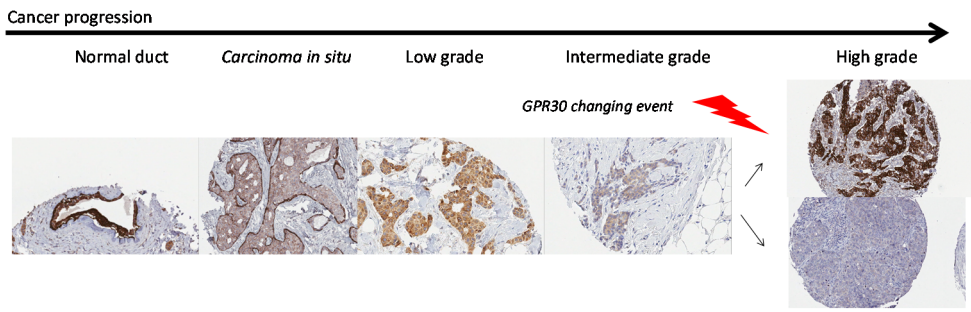


Figure 20. Proposed model of GPR30 down-regulation during cancer progression, until a minority of the tumors experience a changing event that alters GPR30 functioning, resulting in overexpression and translocation to plasma membrane.

Although we present evidence that it may be possible to affect the receptor pharmacologically, the receptor appears to be mainly constitutively active, meaning that it functions without a ligand, and thus regulated by the amount of receptor expressed. Following the success of anti-HER2 treatment, the possibility of GPR30 over-expression at the plasma membrane is very exciting. Indeed, a surface receptor may be targeted specifically by treatments without targeting downstream signaling, as antigen-dependent cell-mediated cytotoxicity (ADCC) plays a crucial role in the effectiveness in many targeted therapies using antibodies. I acknowledge that this is highly speculative, but at the very least, our observations should urge future efforts on GPR30 investigations in breast cancer to analyze more levels of GPR30 expression, and subcellular localization, and not simply dichotomize the measurements.

Statistics and data analysis

> All models are wrong, but some are useful.

Attributed to George E. P. Box, who also wrote the following when describing the famous statistician RA Fisher:

> Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary, following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist, so overelaboration and overparameterization is often the mark of mediocrity.[313]

In the era of big data, machine learning and bioinformatics, I think one sometimes should take a step back and look at what we are trying to achieve. The powerful tools at hand, and relative ease of use of advanced modeling, enables researchers without mathematical training, myself included, to explore these techniques. Careful consideration, as well as close collaboration with statisticians, is crucial for avoiding pitfalls. Correct statistical methodology used to analyze badly designed experiments may be worse than using the wrong statistical methods on well-designed experiments. With that said, here are some considerations regarding the data analysis used in my thesis work.

Patients and sample selection

The five studies in this thesis involve analyses of tumors from six different patient cohorts, and the analysis of experimental cell lines. The patient cohorts represent randomized clinical trials (study I and III), well-defined retrospective cohorts (study II) and semi-matched case-control cohorts (study IV and V). Each cohort type has strengths and weaknesses.

Randomized controlled trials (RCTs) have many advantages, the major advantage for our research being that it provides the best opportunity to search for treatment-predictive biomarkers. The purpose of the randomization is that the groups compared should only differ in the (treatment) variable of interest, and randomization takes care of both known and unknown bias.[314] Otherwise, treatment may be influenced by the factors we want to study. For instance, in study IV, RT is not improving the rate of IBTRs across patients from all centers. This is very likely due to physicians deciding treatment based on the risk of the patients predicted by clinical factors, meaning that the patients not receiving treatment are the ones with the best outcome already before treatment. This makes it hard to truly investigate the predictive potential of biomarkers, and some methodologies are not feasible to apply. Due to the mandatory power calculations, RCTs also often have the

advantage of a large sample size and high-quality follow-up of patients. The potential disadvantage of using RCTs is that they are usually fixed in what other types of treatment are allowed. This means that once the treatment guidelines are updated, the results from analyzing RCTs may be hard to interpret in a more recent context, which is the case with the results from study III. For the low-risk trials used, the number of events is smaller, meaning that the power of the studies decreases. This problem if further enhanced by studying subgroups in RCTs that the study was not originally designed for, and consequently, not powered for. In addition, RCTs are generally more expensive to conduct.

Retrospective cohorts selected from larger RCTs, such as used in study II, have the advantage that they are usually well defined with high-quality follow-up, but lack the full potential of analyzing treatment-predictive variables.

The semi case-controlled sampling used in study IV and V (semi meaning that the samples are selected as cases and controls, but with little matching) has the advantage from an experimental point of view that more samples can be included, and that the biological differences are enhanced when there are more tumors with extreme biology (i.e. aggressive tumors with recurrences). The downside is that most classical statistical analyses become very hard to interpret, and one could even argue that they should not be performed. E.g. the use of the Kaplan-Meier method presented in study IV can only be interpreted as the performance of the classifiers in this very specific cohort, and cannot be generalized to a wider cohort of patients. As already mentioned, the oversampling of cases may also produce other unexpected results, such as that RT has no overall effect in study IV, which calls for caution when interpreting the treatment-predictive results.

Taken together, several considerations must be made when selecting a cohort to study. In general, large RCTs are to be preferred, but this is not universally true as even these cohorts may have too few events to find any the underlying biology governing recurrences. This is even more true when studying early breast cancer and low-risk patients which generally have a very good prognosis. The enrichment of tumors with extreme biology can be needed to discover biologically relevant features, and case-cohort or nested case-control studies may be interesting study designs for biology driven research.[315,316] There is also practical considerations on what tissue is available, which stresses the importance of clinical specimen collection and patient participation in ongoing clinical trials.

Besides selecting the cohort to study, there is also a difference in the type of tissue specimens to analyze. The quality of the biologic material is quite different between snap-frozen tumors and FFPE samples. Until very recently, the analysis of FFPE samples was not possible with high-throughput methods, but technological advancements have made this possible, and the ability to analyze clinical FFPE samples would greatly enhance the possibility to use the methods in the clinic. But

even so, my experience from study IV is that although it would be possible to analyze FFPE samples, the biologic depth is diminished with lower RNA quality, and fresh-frozen tissue may still represent the best tissue for investigating detailed biology.

Analysis of subgroups

With the effort to personalize breast cancer treatment, it inevitably follows that the patient groups that will receive the same type of treatment will be numerically smaller. Ultimately, each individual tumor may be analyzed for the specific cancer transforming events, and treated accordingly. This poses several research challenges. In our studies, we have sub-grouped the patients mainly by ER status (study I, II, IV and V), as this is the major determining factor in breast cancer with genome wide changes in gene expression patterns,[53] for intrinsic molecular subtype (study III), and for treatments given (study I and IV).

The most obvious implication for future studies is that much larger sample populations must be collected to retain the necessary statistical power to detect treatment differences. On the other hand, targeted treatment strategies are hopefully more effective, which to some extent counterbalances this problem. When re-analyzing retrospective data, which is the case in my studies, the risk of performing both type I and type II statistical errors increases, and subgroup analyses have been claimed to be surprisingly unreliable.[317] The risk of performing type I errors increase because of multiple hypothesis testing, and the risk of a false positive finding is especially high when embarking on studies without a pre-specified hypothesis. The risk of type II errors also increases because of the loss of statistical power, and it is very important to remember that the statistical hypothesis testing is made to reject the null hypothesis, not to reject the alternative hypothesis. Or alternatively stated, absence of evidence is not evidence of absence.[318-320]

Possible measures to overcome problems with subgroup analyses have been suggested.[317] First is to pre-specify subgroup analyses of clinical trials, based on prior evidence or sound rationale, before trial start. This would require appropriate power calculations, and could then make the results more reliable. Another strategy is to avoid direct subgroup analyses and instead interpret the overall treatment effect as a relative risk uniform across subgroups, and focus on the absolute risk when deciding to treat patients. Generally, patients with a high-risk tumor usually also have a greater absolute risk reduction with treatment. However, this option may be well suited for large trials with relatively unspecific treatments, but less suited for future targeted trials. A third way of overcoming problems with under-powered subgroup analysis is to use other endpoints with more events, such as recurrence-free survival instead of overall survival.

Since we are performing subgroup analyses, the results must be approached with statistical caution, and the appropriate validation of results must be stressed before final conclusions. However, whenever possible we have taken measures to counteract this problem. In study I and II, we used multiple cohorts for independent validation and we based subgrouping (for ER and tamoxifen treatment) on a biologic rationale that GPR30 would have different functions depending on the endocrine environment. In study III, despite being a retrospective study, the study design was pre-specified before any analysis started. In study IV, the entire analysis was separate for ER+ and ER- disease, meaning that we created different classifiers for ER+ and ER- disease, based on biological knowledge that these are very distinct diseases. This is somewhat different than testing the same classifier in multiple groups, and provide the opportunity to study biologic differences between the groups. However, the numbers in many of the analyzed subgroups are small, and appropriate caution should be applied.

Future analysis of breast cancer will require more and more subdivision of patients, and future targeted treatments may be used for a small fraction of patients with cancers harboring the druggable change. The ability to perform high-quality meta-analyses and pooling studies, as well as planning studies that include patients world-wide, will become increasingly important. From my point of view, pre-specified subgroup analysis could be a good option for targeted treatment, and there is also a great interest in even faster endpoints, such as pathologic complete response with neoadjuvant treatment, although the association with survival is unclear. Eventually, designing future breast cancer studies may require redefining the RCT completely.[321]

Cut-point selection and grouping of data

The aim with personalized medicine is to divide the patients into smaller groups with different, more effective, treatments. A major task is to define those groups and choose the cut-point for division of groups. For established clinical markers this may be chosen with great confidence from existing data, but for experimental assays, this is not always straight forward. An important finding in study I and II is that the expression of GPR30 appear be non-linear with respect to its correlation with other clinicopathologic variables, and most important, outcome. In fact, the relationship is biphasic which means that the common way of dichotomizing the data in to two groups makes the results highly sensitive for the cut-point chosen. I realize that there is a balance between having statistical power to detect difference between groups and analyze relevant biology, but in this case, I think it calls for analyzing the entire data without grouping. Indeed, the conflicting results from earlier studies may well be related to the cut-points chosen. Interestingly, we found similar results for the stem cell biomarker ALDH1 in a study not included in this

thesis.[322] This calls for cautions interpretation of quantitative results from antibody based methods using experimental antibodies in general.

A similar problem is encountered in study III where we use the proliferation maker Ki67 to separate the Luminal A-like from the Luminal B-like tumors. Ki67 is notoriously known for its laboratory specific values and hard to define cut-point.[323,324] We used a method based on the Swedish guidelines at the time: that each laboratory should adjust the cut-point so that 1/3 of all tumors is scored Ki67 high. We adjusted this for the early nature of the SweBCG91-RT cohort and set the cut-point at 10% cells being Ki67 high, resulting in 27% of tumors being scored high. Arguably, the cut-point selection will produce different low-risk groups, but we also tested to include grade and size to define the low-risk group with only marginally altered results. Time will tell if gene expression-based predictions of low-risk tumors will perform better.

Endpoints

Endpoint selection is an important part of study design, and efforts are ongoing to standardize the definitions.[180,181] The distant disease-free survival (DDFS) endpoint provides a way of gaining power compared with overall survival, both because it produces events at an earlier point, and because cancer unrelated events do not dilute the signal (although the standardized definition efforts recommend to call this "interval" instead of "survival", this was not incorporated in our studies, and DDFS used in this thesis work excludes non-breast cancer death). DDFS is used in study I, II, and V. However, if there is any chance the factor studied (e.g. treatment) may have deleterious effect not related to the cancer, overall survival must be monitored.

In study III we used another type of primary endpoint: IBTR as first recurrence at 10 years, analyzed with a competing risk approach. This was the primary endpoint of the original study, and is further chosen for the best performance of detecting effects of adjuvant RT, which mainly reduces IBTRs. Further, IBTR after metastasis is not of equal clinical importance, and is not monitored as carefully. However, the endpoint IBTR as first event may be problematic for several reasons. First, IBTR may arguably be considered to depend less on biology and more on random factors and surgical performance than distant recurrence, as distant recurrence is a more definite proof of aggressive biology. Also, in the search for biomarkers associated with IBTR as a first event, this does not take later IBTR events into account. This results in an analysis trying to find a marker that is predicting IBTR as a first event, in contrast to samples with IBTR after distant recurrence. Thus, it detects genes that are mutually exclusive between those endpoints. As many of the traditional nongenomic markers are the same for IBTR and distant metastasis, as well as for commercially available gene expression tests, it seems logical not to conduct the experiment in this way.

In study IV on the other hand, we use IBTR or no IBTR as the endpoint, which is better suited to maximize the biological differences between the tumors, but is not representative for a clinical consecutive cohort of breast cancer patients. Study IV is also sampled as cases and controls, which enriches for patients suffering an IBTR, with the drawback that the assumption of many statistical methods does not hold in this setting.

Multiple hypothesis testing

Many problems with statistics comes from the question of what is "significant". I believe there is a lot of confusion here, and from my perspective, it really depends on the aim of the analysis. In study III, the aim was to draw generalizable conclusions, and thus the strict use of correct statistical methodology, predefined hypotheses, and to avoid multiple hypothesis testing, was important.

In study V in contrast, we tested the difference of abundance of proteins between groups. In my view, this was not with the aim to finally draw the conclusion that the proteins are differentially regulated, and to generalize the idea to other breast tumors. Because of the limited sample set, and exploratory nature of the analysis, the aim was to prioritize proteins for further investigation, and we therefore performed no multiple hypothesis correction. Ultimately, this should be viewed as a step in technological advancement towards a future assay that must be properly validated.

I believe that the ranking of variables by p-value for importance may be fine, but once you try to interpret the p-values to reject the null hypothesis, other considerations must be made. One way is to calculate the false-discovery rate as originally described by Benjamini and Hochberg, and the later described q-value. The idea is to control the total number of type I errors, which is less strict than controlling the risk of performing at least one type I error, the goal of e.g. Bonferroni correction. While the p-value describes the false-positive rate of all tested features, meaning the number of truly null features that are called significant, the q-value describes the false-discovery rate as the proportion of the features called significant that are in truly null. The latter may be a better choice when we do not know how many features to expect to be significant.[325-328] To illustrate the problem of not correcting for multiple hypothesis testing, a report using functional magnetic resonance imaging (MRI) was presented on the feelings by a dead Atlantic salmon when put in different socially demanding situations.[329]

Creating a targeted assay for clinical use

Much of our work has been focused on creating targeted assays for potential clinical use, and there are some points I believe need to be addressed for future success, and before eventual introduction into clinical practice.

Accuracy

The accuracy of the assay, and the proper reporting of assay development, is crucial for clinical utility, but several aspects of accuracy must be considered.[330-332] First, the technical accuracy of the measurements is important, often referred to as analytical validity.[115] The analytical validity of the methods in study IV and V is largely unknown, and the methods must be further tested before eventual clinical use.[333] To initiate the investigation of performance of label-free SRM under clinical conditions, we conducted an experiment (not part of this thesis) where we tested the normal variation, i.e. the variation not attributed to the phenotype studied, of other cancer associated proteins in blood plasma of healthy volunteers.[234] We could show that there was technical variation, but that the dominant source of variation was time, from day to day, and hour to hour. This is probably more extreme for blood biomarkers, but highlights the testing that needs to be done before clinical introduction. It also raises the question of the generalizability to other cohorts. I view the experiments in study V mainly as proof of concept work for developing the techniques, and ranking potential biomarkers for further investigation.

Further, the gene expression assay in study IV relies on a well-documented platform, but the individual probes used have only been validated *in silico*, and internally by the manufacturing company. The overall results seem very good, and the correlation with known clinical markers is excellent. However, experimental probes may suffer from high variance and design problems. Indeed, one probe was not working at all (*AKNA*). Further validation of the assay is motivated, especially with regards to the possible use in FFPE samples, for which it is not yet tested. We validated the overall signature in degraded fresh frozen samples that seem similar to FFPE samples, but the degradation mechanisms may be different, and a non-linear degradation could prove very difficult to handle for both the short probe based gene expression assay, and the relative top scoring pairs algorithm.[334]

Validation

As stated in methodological considerations, the use of high-dimensional methods brings the risk of interpreting noise for signal. Even though rigorous statistical plans are used, unconscious information leakage may be one of the reasons it has been hard to validate studies in other data. Further validation is therefore crucial to assess the clinical validity, meaning consistent ability to accurately divide patients into

meaningful groups.[115] For GPR30, we observe the same trends in three independent cohorts in study I and II. For the gene expression predictors in study IV, we validate them in two publicly available datasets, and for study V we perform orthogonal validation in publicly available gene expression data. Further, in study IV we build on previous knowledge and include previously described genes in our assay, and test these profiles in our data. I think this is a very important step, that we must follow up previous findings, while ultimately the proposed classifiers must be validated in prospective data.

Cost-benefit and clinical utility

Cost-benefit has become an increasingly important factor in modern health care, and novel molecular tests must prove a considerable cost-benefit improvement to be able to be introduced in clinical practice. It is not enough with perfect analytical and clinical validity, or interesting biology insights, if the patient's outcome is not changed in a meaningful way. I believe the key for cost-benefit is to target important clinical questions and from start aim at clinical utility, including the avoidance of costly treatment for patients that do not benefit from them.[115,335]

Sample throughput

Related to the cost-benefit is the sample throughput, and the analysis of samples in large clinical trials is emerging.[336] The targeted assay in study IV has proven scalability and the throughput is sufficient to meet the number of clinical breast cancer samples. The LC-SRM assay in study V on the other hand, needs improvements before true high-throughput capability. Part of the bottle-neck is the sample preparation that is time-consuming, especially if fractionation of samples is necessary for the targeted sensitivity. Also related to the separation of peptides is the nano-LC system that we used, which provide excellent research-grade peptide separation for clinical samples, but limits the throughput.

Ability to work under clinical conditions

The sample collection, storage and preparation are very important, and have large impact on the possible analysis, as we describe in study IV. The introduction of new sampling techniques or storage is hard to do in a clinical environment that is under considerable time and economic stress already. Also, we are far from, if ever, reaching a point where molecular diagnostics will fully overtake the pathologist manual inspection of the tumor. As analysis by the pathologist must be prioritized over newer high-throughput techniques, the possibility to analyze the same type of specimen, handled the way the pathologists do it for routine analysis, is a key for clinical introduction. The ability of the assay presented in study IV to handle degraded RNA is one step towards this goal.

The traditional way of handling clinically collected samples as FFPE tissue has made the use of modern technologies impossible until recently.[337] However, the data produced from FFPE samples still commonly contains more noise and fewer genes and proteins that can be measured with high confidence. For biology driven research, and finding the most relevant biological mechanism, samples with optimal handling will still have a place in the research setting.

Another important aspect of clinical utility is the possibility to take any new sample and classify it regardless of the characteristics of other samples. Most molecular high-dimensional techniques rely heavily on normalization, scaling and cohort composition of the samples. Even IHC techniques, such as Ki67 measurements, are not truly single sample as the cut-off to consider a sample high or low shift over time. Accordingly, the laboratory must calibrate itself. To truly be able to classify a single sample will make the step to the clinic easier. The top scoring pairs method we used in study IV is an important attempt towards this, and the same strategy has been adopted for several other breast cancer classifers.[280,338]

Strategy for selection of features for a targeted analysis

The definition of a targeted assay is that a small subset of all available features is measured. This can greatly improve the sensitivity, accuracy, throughput and cost of the assay. However, inherently it follows that information not conveyed by the targeted features are lost. The selection of features is therefore of high importance. Broadly, the selection process can be divided in features that have known biological associations with the outcome of interest, or methods that prove to work but where the function of the genes and proteins are largely unknown. We use both these strategies in study IV and V.

In study IV, we base the selection of the largest proportion of genes for the targeted assay on data from the full transcriptome analysis, without considering the biology conveyed by the selected genes. The reason for this is that we think that the genes and biology governing radioresistance in this setting are largely unknown. Indeed, when we analyzed which genes were chosen for the targeted assay, known biological functions such as proliferation and immune response were captured, but also genes without apparent known mechanism. We used the random forest model, as it can find combinations of genes with non-linear relationships with the outcome variable, but a comparison with ranking with an ordinary Mann-Whitney test showed a large gene overlap (data not shown). In my mind, it is important to use previous knowledge, and to that end we added previously reported signatures and genes reported to be associated with radioresistance. This combination of approaches allowed us to take advantage of both prior knowledge, and discover new biological concepts. In study V, we used a similar approach when we ranked genes from a global discovery analysis, and then added proteins based on prior biological knowledge from the literature.

As our biological knowledge increases, the use of that information to create targeted assays will also increase, and from a scientific standpoint, the use of existing biological knowledge to create targeted assays is very tempting. Especially in clinical trials, the targeted approach with hypothesis driven research may be valuable.[330] However, we must not let our lack of complete biological information hinder the development of useful clinical assays.

Conclusions

The specific conclusions from my thesis are:

- The putative estrogen receptor GPR30 appears to be constitutively apoptotic, and may undergo changes that alters the function of the receptor in breast cancer with potential use as a treatment target.
- GPR30 in the plasma membrane is a prognostic biomarker, but it is not treatment-predictive for adjuvant tamoxifen in primary breast cancer.
- Breast cancer subtype is not treatment-predictive for adjuvant whole breast radiotherapy, although the HER2+ tumors may respond less to radiotherapy without adjuvant anti-HER2 treatment.
- A presumed low-risk group of breast cancer patients, that are currently enrolled in clinical trials of de-escalating radiotherapy after breast-conserving surgery, has an excellent effect of radiotherapy and high risk of recurrence without radiotherapy, in the absence of systemic adjuvant treatment.
- A targeted 248-gene radiosensitivity gene expression panel with classifiers utilizing the top scoring pairs classifier algorithm can be used in a single sample manner, and is prognostic for the development of ipsilateral breast tumor recurrence.
- The use of two classifiers in a step-wise fashion on the 248-gene targeted panel that assess both radiation omission and radiation intensification may stratify patients for treatment with adjuvant radiotherapy after breast-conserving surgery.
- The combined use of shotgun LC-MS/MS and LC-SRM mass spectrometry is feasible for clinical samples, and show promise in detecting and prioritizing biomarkers during discovery phase research.

Future perspectives

The work presented in this thesis is a small step towards better personalized adjuvant treatment for primary breast cancer, and provides insights into the biology governing the response to therapy. Much work is still needed, and research is an ever-ongoing process. Here I present some thoughts on possible directions forward.

GPR30

To be able to utilize GPR30 in breast cancer, as a biomarker for prognosis, treatment prediction, or as a treatment target, I believe we first must understand the basic function of this receptor. Still, properties such as cellular localization, signaling mechanisms, trafficking, and even the true ligand are poorly understood and widely disputed. All these areas are under investigation, and it will be very interesting to follow the development. However, even without knowing the biological background, our data suggest that GPR30 changes the behavior in a minority of breast cancers, and that it may be targeted at the plasma membrane. One first step to investigate this hypothesis is to mine the public data repositories for any evidence that there are overexpression, amplification, mutations, or other events that may change the functioning of the receptor.

Interesting *in vitro* data suggest that GPR30 may play a role in acquired endocrine treatment resistance. Although we find no evidence for this in the adjuvant tamoxifen setting, other systems may be better to study. The effect of prolonged treatment would be very interesting to study, e.g. in the metastatic setting and after failure to prolonged endocrine treatment. Also, a way of analyzing the receptor at different stages during cancer progression could elucidate what happens during this course.

Will this receptor be used as a biomarker in the future? I doubt so. First, the scoring of the receptor is difficult and the finding of the receptor in the plasma membrane may be dependent on tissue handling and fixation. Second, although the prognostic value is significant in a multivariable analysis, the added value compared to classical markers, and especially modern gene expression tests, is only marginal. Third, without a known function or rationale, it will be very hard to motivate the introduction to clinical use.

However, as a single marker that appear to be over-expressed at the plasma membrane, the role as a treatment target, and potentially after endocrine treatment, remains very tempting, and would be very interesting to study further.

Individualized radiotherapy

Individualized radiotherapy, here meaning radiotherapy based on the characteristics of the tumor, has long been a goal of radiation oncology, but very hard to implement. In this thesis, we show that intrinsic subtype is not enough to personalize the adjuvant radiotherapy after breast-conserving surgery. We further show that on the gene expression level, there is a reproducible biological signal on who are at higher risk of recurrence with or without radiotherapy. However, the performance of the classifiers proposed here will not be enough to change clinical practice. The classifiers must be further tested and refined in larger studies, and ultimately in prospective cohorts. There are ongoing efforts to do this, but there is a long road ahead.

In the end, one must also ask whether the process to develop ipsilateral breast tumor recurrence is to some extent more random than the process of developing distant recurrence. If so, it may not be possible to determine exactly who will respond to radiotherapy or not based on the characteristics of the primary tumor. The usage of baseline risk to individualize radiotherapy is a way forward, such as performed in current de-escalation trials. Ultimately, we may have to live with some patients being over-treated, and the most important thing becomes to get the patient involved in the decision, and make an informed decision if the possible gains with treatment are worth the potential side-effects.

Future of proteomics and systems biology

Study V concerns the study of the proteome, and I have come in contact with proteomics in other studies during my work that is not included in this thesis.[52,234] The study of the proteome lags behind genomics and transcriptomics, and I sometimes feel like being in Plato's cave allegory. The allegory points out that we are only aware of what our senses tells us, but there is no reason, considering the limits of our senses, that our perceived world should be the complete and final reality. He describes it with a tale of a group of people that are trapped, since before they know, in a cave with no possibility to connect to the outside world. However, there is an opening of the cave, which they cannot look out from, but a fire outside casts shadows of objects and other humans on the cave walls. This is the outside world they know, and they interpret this world through the shadows. For them, this is the ultimate reality, and they have no reason to even question there should be another reality. For me, the study of gene expression and DNA sequences to gain insights into the functional biology of cancer is like studying the shadows. The point

here is that the study of proteins is one level closer to the functional phenotype of the cell, and should thus provide a better understanding of cancer biology.

Mass spectrometry-based proteomics has been associated with several problems and the development was halted with the improper discovery and use of unidentified MS peaks, that later turned out to be markers not associated with cancer.[339] The techniques have matured and are now providing considerably more reliable data. However, there is still much work to be done, especially in the field of combining non-targeted proteome-wide studies and targeted methods more applicable to clinical problems, which we advance in study V. An exciting development is the data-independent acquisition method that may combine the advantages of proteome wide coverage with the specificity of the targeted approach, and remained quantitative performance.

A further note is as proteomics becomes more reliable and more widely applied, there will be tremendous opportunities to combine data on several biological levels, and especially the connection of DNA changes that lead to functional protein changes will be exciting. So far, most cancer studies concern the dynamic regulation of abundances of molecules (genes, proteins) that still function as normal. However, it is expected that many changes important in cancer are changes at the structural level that results in altered function, i.e. DNA sequence changes that lead to altered protein structure, and ultimately function. Thus, the study of altered protein structure may provide new insights in cancer biology. This area has just recently attracted attention, and I believe much more is to come.[17,220,340-346]

The future oncologist

However risky it may be to foresee the future and make bold claims of what will happen during a lifetime, I will take this opportunity for a few remarks. I have started my clinical career in a time when the area of clinical medicine is on the verge of undergoing a change of magnitudes not previously seen. The rate at which we use the new powerful techniques to create biological data is simply breathtaking. During the coming years, the most important work will be to transform the amounts of data into real knowledge, and ultimately transform the knowledge into better tools to battle disease. To speak with Sir Winston Churchill: This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.

In terms of cancer care, every tumor will undergo sequencing and global analysis at several levels, including functional protein analysis. Eventually, the entire cancer biology dynamics of an individual tumor will be modeled, and the exact perturbations leading to the disease known. This will lead to targeted treatments, and although we may not cure all cancers, treatments will stop the tumors from growing until they are no longer a problem for the patient's life and health.

Ultimately, I believe that during my lifetime and work as a physician, cancer will no longer be a leading cause of death.

This will put much pressure on the future oncologist to be updated and master, or at least understand, many more areas than traditionally, such as data analysis.[273] Molecular tumor boards is one way forward, and the possibility to work more closely with tumor biologists and data scientists in the clinic to provide the best possible care is truly fascinating. However, to make these boards meaningful and discussions constructive, a deeper understanding of the methods used from the treating physician is required.

The nature of science is to take small steps forward to eventually make improvements, and it is a world-wide collaborative effort. The work I present in this thesis should be viewed in that light. The ultimate goal is to provide improved care for our patients, but to reach that goal, it is important to both deepen our understanding of biology, and how to use the novel techniques at our hands. I believe my work is a small step towards that goal, and I expect the work to continue.

Acknowledgements

The work in this thesis could not have been done without the support from many persons, both colleagues and friends. It is not possible to mention all here, and to all of you who have supported me – thank you!

To my main supervisor **Emma Niméus**. Thank you for bringing me onboard in 2011 with your great enthusiasm and energy. I remember the first time we met; you are probably the only person that can make the work of reading a couple of hundred patient charts, down in an archive basement, sound like the most exciting summer project. I did not understand all the different project ideas (and probably still do not), but you have taught me a way of seeing opportunities instead of obstacles. Your pragmatic view on research, and ability to always find good collaborations and projects, and just making things work, impresses me a lot. Thank you for letting me follow different research paths, always supporting me, and for making me test and learn so many different techniques. Thank you also for great opera and food suggestions! I could not have had a better research education.

To my co-supervisor, **Mårten Fernö**. Thank you for always believing in me, letting me follow my own ideas and interests, promoting me in collaborations, and pushing me just enough. You are a fantastic supervisor, with a way of changing the supervision in accordance with the demands of the moment. You have taught me so much on how to navigate in the quite weird field of science, and the value of collaborations. Thank you also for guiding me in San Antonio and Villa Aske, and showing me other aspects of this work, and how to balance life. Together with Emma, you have provided the best possible education.

To **Fredrik Leeb-Lundberg**, the person who first made me try research. Thank you for introducing me to the world of science, and for letting me try different things in the lab, but first and foremost, for sharing your genuine enthusiasm about basic science and the pure joy of discovering how nature works. Your happiness when showing how a GPCR may bite itself and self-activate is one of my most memorable scientific learning experiences. Thank you for your continued support and mentorship, it means a lot to discuss with someone outside the clinical world.

To **Per Malmström**, for teaching me how to think about clinical research and your way of always trying to achieve the highest possible standards, both in patient care and in research. I am sad I did not get the opportunity to have you as a clinical mentor.

To **Linda Hartman**, for constantly helping me with statistical issues, teaching me an approach to use statistics as a tool, not the goal, in research, and for stopping me when I move to fast or overinterpret things.

To all other co-authors, thank you for great collaborations! Especially, thank you to **Irma Fredriksson** and **Fredrik Wärnberg** for continued support and for providing a positive atmosphere, even in times of setbacks, and for hosting me in Uppsala. Also thank you to **Per Karlsson**, for great collaboration and combined ventures in Michigan and beyond.

To **Patrik Edén**, for showing me the joy of physics and mathematics.

To **Johan Staaf**, for bioinformatics expertise, and for careers discussions.

To **Lisa Rydén**, for always maintain highest scientific standards in the co-lead of the Fernö-Rydén group, and for your precise insights in the translational needs in breast cancer research.

To **Carina Forsare**, **Sara Baker** and **Kristina Lövgren** for your expert work in the lab (and with databases…). Without you, none of the work herein would have been possible.

To members of the Fernö-Rydén group: **Kristina Aaltonen, Maria Ekholm, Ulrik Narbe, Pär-Ola Bendahl, Looket Dighe, Anna Ehinger, Sara Jansson,** and **Charlotte Levin Tykjaer Jörgensen** for making it the best place for research studies!

To **Johan Malmström**, and the infection medicine proteomics groups. Thank you for letting me be a part of the lab and showing me some real science.

To **Susanne André** and **Björn Frostner**, for running all practical aspects of the department. I honestly do not know how I would have managed without you.

To **Ingrid Hedenfalk,** for scientific and careers discussions, and for showing me the cheesecake factory.

To **Tommaso De Marchi** and **Paul Pyl,** for joining the Breast Cancer Proteogenomics Group and bringing in your much-needed expertise.

To past and present heads of department: **Signe Borgquist, Mef Nilbert, Bo Baldetorp, Ingrid Wilson, Lars Ekblad and Karin Jirström,** for running the department in an excellent way and providing a good environment for doctoral studies.

To **Dick Killander**, for the continued support and interest in my doctoral studies, and for providing valuable insights in the general developments in oncology.

To my colleagues at the clinic: **Fredrika Killander, Jenny Jönsson, Laura Martin de la Fuente, Niklas Loman, Sara Alkner, Anna-Maria Larsson, Martin Nilsson, Tomas Breslin, Anna-Karin Falck, Marie Klintman, Ana Bosch Campos, Henrik Ekedahl, Karin Elebro** and many others. Thank you for being so awesome at what you do. Hopefully we work together in a not too far future.

To **Laura Chang** and **George Zhao,** for hosting me in Michigan and becoming my best friends for a month!

To my colleagues at the University, thank you for making it such a great place to study and work!

To **David Larsson** and **Max Hultgren**, for all the fun and memories throughout medical school, and for the continued friendship!

To **Anders Törnkvist** and **Benoit Almer**, the latin kings, for much fun with latin discussions.

To **Maria Simonsson**, for dinners, careers discussions and continued friendship.

To **Stefan Broselid**, for introducing me to science, and helping me with my first own steps.

To the **Swedish Breast Cancer Group**, **the South Swedish Breast Cancer Group** and the **West Sweden Breast Cancer Group** for collecting and managing tumor samples. Your work is the foundation for almost all studies in this thesis.

To all the **patients** that contributed tumor samples for these studies. Progress in cancer research will not be possible without your contribution!

To all the funders of this work, especially **Bröstcancerföreningarnas Riksorganisation (BRO)**, **John och Augusta Perssons stiftelse** and **Maggie Stephens stiftelse** for the support of my personal education that made much of this work possible. Thank you also to the other funders of this work: the Swedish Cancer Society (Cancerfonden), Region Skåne, Governmental Funding of Research within the Swedish National Health Service (ALF), Mrs Berta Kamprad Foundation, Anna-Lisa and Sven-Erik Lundgren Foundation, Magnus Bergvall Foundation, the Gunnar, Arvid, and Elisabeth Nilsson Cancer Foundation, the Anna and Edwin Berger Foundation, the Swedish Cancer and Allergy Foundation, Skåne County Research Foundation (FOU), Lund University Research Foundation, Skåne University Hospital Research Foundation, BioCARE, the King Gustaf V Jubilee Fund, the Marcus and Marianne Wallenberg Foundation and the Stockholm Cancer Society.

To my family, my **mum Ingrid** and **dad Christer** for your never-ending support and encouragement. I apologize for taking some extra time before becoming a real doctor. To **Erik** for being a true little brother, always being there when needed, and for filling my spare time with model aircrafts. To **Susanna**, for being an equally true little sister, following your own path and showing that not all things must be planned.

To **Josefin**, my BFF and soon to be wife. I could not have done this without your support, and I cannot understand how you have managed to be positive at all times during my work. With you, every day becomes great, and I truly look forward to our future journey.

References

1. Torre, L. A.; Bray, F.; Siegel, R. L., et al., Global cancer statistics, 2012. *CA Cancer J Clin* **2015,** 65, (2), 87-108.
2. Socialstyrelsen, Statistics on Cancer Incidence 2015 (official Swedish health statistics). http://www.socialstyrelsen.se/publikationer2017/2017-1-20/Sidor/default.aspx **2017**.
3. Socialstyrelsen (ed. Johansson, E., Cancer i sifrror 2013. Populärvetnskapliga fakta om cancer. http://www.socialstyrelsen.se/publikationer2013/2013-6-5 **2013**.
4. Torre, L. A.; Siegel, R. L.; Ward, E. M., et al., Global Cancer Incidence and Mortality Rates and Trends--An Update. *Cancer Epidemiol Biomarkers Prev* **2016,** 25, (1), 16-27.
5. Bosetti, C.; Bertuccio, P.; Levi, F., et al., The decline in breast cancer mortality in Europe: an update (to 2009). *Breast* **2012,** 21, (1), 77-82.
6. Autier, P.; Boniol, M.; Gavin, A., et al., Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. *Bmj* **2011,** 343, d4411.
7. Berry, D. A.; Cronin, K. A.; Plevritis, S. K., et al., Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* **2005,** 353, (17), 1784-92.
8. Narod, S. A.; Iqbal, J.; Miller, A. B., Why have breast cancer mortality rates declined? *Journal of Cancer Policy* **2015,** 5, 8-17.
9. Welch, H. G.; Prorok, P. C.; O'Malley, A. J., et al., Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *N Engl J Med* **2016,** 375, (15), 1438-1447.
10. Russo, J.; Russo, I. H., Development of the human breast. *Maturitas* **2004,** 49, (1), 2-15.
11. Hanahan, D.; Weinberg, R. A., The hallmarks of cancer. *Cell* **2000,** 100, (1), 57-70.
12. Hanahan, D.; Weinberg, R. A., Hallmarks of cancer: the next generation. *Cell* **2011,** 144, (5), 646-74.
13. Foulkes, W. D., Inherited susceptibility to common cancers. *N Engl J Med* **2008,** 359, (20), 2143-53.
14. Afghahi, A.; Kurian, A. W., The Changing Landscape of Genetic Testing for Inherited Breast Cancer Predisposition. *Curr Treat Options Oncol* **2017,** 18, (5), 27.
15. Perou, C. M.; Borresen-Dale, A. L., Systems biology and genomics of breast cancer. *Cold Spring Harb Perspect Biol* **2011,** 3, (2).

16. Nik-Zainal, S.; Davies, H.; Staaf, J., et al., Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **2016,** 534, (7605), 47-54.
17. Mertins, P.; Mani, D. R.; Ruggles, K. V., et al., Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **2016,** 534, (7605), 55-62.
18. Cancer Genome Atlas, N., Comprehensive molecular portraits of human breast tumours. *Nature* **2012,** 490, (7418), 61-70.
19. Curigliano, G.; Burstein, H. J.; E, P. W., 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.
20. Giordano, S. H.; Temin, S.; Kirshner, J. J., et al., Systemic therapy for patients with advanced human epidermal growth factor receptor 2-positive breast cancer: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol* **2014,** 32, (19), 2078-99.
21. Figueroa-Magalhaes, M. C.; Jelovac, D.; Connolly, R., et al., Treatment of HER2-positive breast cancer. *Breast* **2014,** 23, (2), 128-136.
22. Liang, K.; Lu, Y.; Jin, W., et al., Sensitization of breast cancer cells to radiation by trastuzumab. *Mol Cancer Ther* **2003,** 2, (11), 1113-20.
23. Guo, G.; Wang, T.; Gao, Q., et al., Expression of ErbB2 enhances radiation-induced NF-kappaB activation. *Oncogene* **2004,** 23, (2), 535-45.
24. Cao, N.; Li, S.; Wang, Z., et al., NF-kappaB-mediated HER2 overexpression in radiation-adaptive resistance. *Radiat Res* **2009,** 171, (1), 9-21.
25. Duru, N.; Fan, M.; Candas, D., et al., HER2-associated radioresistance of breast cancer stem cells isolated from HER2-negative breast cancer cells. *Clin Cancer Res* **2012,** 18, (24), 6634-47.
26. Hou, J.; Zhou, Z.; Chen, X., et al., HER2 reduces breast cancer radiosensitivity by activating focal adhesion kinase in vitro and in vivo. *Oncotarget* **2016,** 7, (29), 45186-45198.
27. Perou, C. M.; Sorlie, T.; Eisen, M. B., et al., Molecular portraits of human breast tumours. *Nature* **2000,** 406, (6797), 747-52.
28. Sorlie, T.; Perou, C. M.; Tibshirani, R., et al., Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **2001,** 98, (19), 10869-74.
29. Goldhirsch, A.; Winer, E. P.; Coates, A. S., et al., Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* **2013**.
30. Coates, A. S.; Winer, E. P.; Goldhirsch, A., et al., Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol* **2015,** 26, (8), 1533-46.
31. Krop, I.; Ismaila, N.; Andre, F., et al., Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. *J Clin Oncol* **2017,** 35, (24), 2838-2847.

32. Prat, A.; Perou, C. M., Deconstructing the molecular portraits of breast cancer. *Mol Oncol* **2011,** 5, (1), 5-23.
33. Parker, J. S.; Mullins, M.; Cheang, M. C., et al., Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **2009,** 27, (8), 1160-7.
34. Martin, M.; Gonzalez-Rivera, M.; Morales, S., et al., Prospective study of the impact of the Prosigna assay on adjuvant clinical decision-making in unselected patients with estrogen receptor positive, human epidermal growth factor receptor negative, node negative early-stage breast cancer. *Curr Med Res Opin* **2015,** 31, (6), 1129-37.
35. Nielsen, T.; Wallden, B.; Schaper, C., et al., 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.
36. Cardoso, F.; van't Veer, L. J.; Bogaerts, J., et al., 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* **2016,** 375, (8), 717-29.
37. Geradts, J.; Bean, S. M.; Bentley, R. C., et al., The oncotype DX recurrence score is correlated with a composite index including routinely reported pathobiologic features. *Cancer Invest* **2010,** 28, (9), 969-77.
38. Orucevic, A.; Bell, J. L.; McNabb, A. P., et al., Oncotype DX breast cancer recurrence score can be predicted with a novel nomogram using clinicopathologic data. *Breast Cancer Res Treat* **2017,** 163, (1), 51-61.
39. Zujewski, J. A.; Kamin, L., Trial assessing individualized options for treatment for breast cancer: the TAILORx trial. *Future Oncol* **2008,** 4, (5), 603-10.
40. Lal, S.; McCart Reed, A. E.; de Luca, X. M., et al., Molecular signatures in breast cancer. *Methods* **2017**.
41. Eden, P.; Ritz, C.; Rose, C., et al., "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur J Cancer* **2004,** 40, (12), 1837-41.
42. Viale, G.; de Snoo, F. A.; Slaets, L., et al., Immunohistochemical versus molecular (BluePrint and MammaPrint) subtyping of breast carcinoma. Outcome results from the EORTC 10041/BIG 3-04 MINDACT trial. *Breast Cancer Res Treat* **2017**.
43. Lundberg, A.; Lindstrom, L. S.; Harrell, J. C., et al., Gene Expression Signatures and Immunohistochemical Subtypes Add Prognostic Value to Each Other in Breast Cancer Cohorts. *Clin Cancer Res* **2017**.
44. Cardoso, F.; Piccart-Gebhart, M.; Van't Veer, L., et al., The MINDACT trial: the first prospective clinical validation of a genomic tool. *Mol Oncol* **2007,** 1, (3), 246-51.
45. Sparano, J. A., TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* **2006,** 7, (4), 347-50.
46. Cardoso, F.; Van't Veer, L.; Rutgers, E., et al., Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* **2008,** 26, (5), 729-35.
47. Wong, W. B.; Ramsey, S. D.; Barlow, W. E., et al., The value of comparative effectiveness research: projected return on investment of the RxPONDER trial (SWOG S1007). *Contemp Clin Trials* **2012,** 33, (6), 1117-23.

48. Sun, Z.; Prat, A.; Cheang, M. C., et al., Chemotherapy benefit for 'ER-positive' breast cancer and contamination of nonluminal subtypes-waiting for TAILORx and RxPONDER. *Ann Oncol* **2015,** 26, (1), 70-4.

49. Sparano, J. A.; Gray, R. J.; Makower, D. F., et al., Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med* **2015,** 373, (21), 2005-14.

50. Stemmer, S. M.; Steiner, M.; Rizel, S., et al., Clinical outcomes in patients with node-negative breast cancer treated based on the recurrence score results: evidence from a large prospectively designed registry. *NPJ Breast Cancer* **2017,** 3, 33.

51. Lindstrom, L. S.; Karlsson, E.; Wilking, U. M., et al., Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J Clin Oncol* **2012,** 30, (21), 2601-8.

52. Kurbasic, E.; Sjöström, M.; Krogh, M., et al., Changes in glycoprotein expression between primary breast tumour and synchronous lymph node metastases or asynchronous distant metastases. *Clinical proteomics* **2015,** 12, (1), 13.

53. Gruvberger, S.; Ringner, M.; Chen, Y., et al., Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* **2001,** 61, (16), 5979-84.

54. Mohammed, H.; Russell, I. A.; Stark, R., et al., Progesterone receptor modulates ERalpha action in breast cancer. *Nature* **2015,** 523, (7560), 313-7.

55. Carroll, J. S., Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer. *Eur J Endocrinol* **2016,** 175, (1), R41-9.

56. Clarke, R.; Tyson, J. J.; Dixon, J. M., Endocrine resistance in breast cancer--An overview and update. *Mol Cell Endocrinol* **2015,** 418 Pt 3, 220-34.

57. Alluri, P. G.; Speers, C.; Chinnaiyan, A. M., Estrogen receptor mutations and their role in breast cancer progression. *Breast Cancer Res* **2014,** 16, (6), 494.

58. Johansson, H. J.; Sanchez, B. C.; Forshed, J., et al., Proteomics profiling identify CAPS as a potential predictive marker of tamoxifen resistance in estrogen receptor positive breast cancer. *Clin Proteomics* **2015,** 12, (1), 8.

59. Johansson, H. J.; Sanchez, B. C.; Mundt, F., et al., Retinoic acid receptor alpha is associated with tamoxifen resistance in breast cancer. *Nat Commun* **2013,** 4, 2175.

60. De Marchi, T.; Timmermans, M. A.; Sieuwerts, A. M., et al., Phosphoserine aminotransferase 1 is associated to poor outcome on tamoxifen therapy in recurrent breast cancer. *Sci Rep* **2017,** 7, (1), 2099.

61. De Marchi, T.; Kuhn, E.; Dekker, L. J., et al., Targeted MS Assay Predicting Tamoxifen Resistance in Estrogen-Receptor-Positive Breast Cancer Tissues and Sera. *J Proteome Res* **2016,** 15, (4), 1230-42.

62. De Marchi, T.; Liu, N. Q.; Stingl, C., et al., 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Mol Oncol* **2016,** 10, (1), 24-39.

63. Umar, A.; Kang, H.; Timmermans, A. M., et al., Identification of a putative protein profile associated with tamoxifen therapy resistance in breast cancer. *Mol Cell Proteomics* **2009,** 8, (6), 1278-94.

64. Wang, Z. Y.; Yin, L., Estrogen receptor alpha-36 (ER-alpha36): A new player in human breast cancer. *Mol Cell Endocrinol* **2015,** 418 Pt 3, 193-206.
65. Gu, W.; Dong, N.; Wang, P., et al., Tamoxifen resistance and metastasis of human breast cancer cells were mediated by the membrane-associated estrogen receptor ER-alpha36 signaling in vitro. *Cell Biol Toxicol* **2017,** 33, (2), 183-195.
66. Warner, M.; Huang, B.; Gustafsson, J. A., Estrogen Receptor beta as a Pharmaceutical Target. *Trends Pharmacol Sci* **2017,** 38, (1), 92-99.
67. Tan, W.; Li, Q.; Chen, K., et al., Estrogen receptor beta as a prognostic factor in breast cancer patients: A systematic review and meta-analysis. *Oncotarget* **2016,** 7, (9), 10373-85.
68. Andersson, S.; Sundberg, M.; Pristovsek, N., et al., Insufficient antibody validation challenges oestrogen receptor beta research. *Nat Commun* **2017,** 8, 15840.
69. Barton, M., Not lost in translation: Emerging clinical importance of the G protein-coupled estrogen receptor GPER. *Steroids* **2016,** 111, 37-45.
70. Thomas, P.; Pang, Y.; Filardo, E. J., et al., Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. *Endocrinology* **2005,** 146, (2), 624-32.
71. Pupo, M.; Maggiolini, M.; Musti, A. M., GPER Mediates Non-Genomic Effects of Estrogen. *Methods Mol Biol* **2016,** 1366, 471-488.
72. Owman, C.; Blay, P.; Nilsson, C., et al., Cloning of human cDNA encoding a novel heptahelix receptor expressed in Burkitt's lymphoma and widely distributed in brain and peripheral tissues. *Biochem Biophys Res Commun* **1996,** 228, (2), 285-92.
73. Jacenik, D.; Cygankiewicz, A. I.; Krajewska, W. M., The G protein-coupled estrogen receptor as a modulator of neoplastic transformation. *Mol Cell Endocrinol* **2016,** 429, 10-8.
74. Molina, L.; Figueroa, C. D.; Bhoola, K. D., et al., GPER-1/GPR30 a novel estrogen receptor sited in the cell membrane: therapeutic coupling to breast cancer. *Expert Opin Ther Targets* **2017,** 21, (8), 755-766.
75. Scaling, A. L.; Prossnitz, E. R.; Hathaway, H. J., GPER Mediates Estrogen-Induced Signaling and Proliferation in Human Breast Epithelial Cells and Normal and Malignant Breast. *Horm Cancer* **2014**.
76. Marjon, N. A.; Hu, C.; Hathaway, H. J., et al., G protein-coupled estrogen receptor (GPER) regulates mammary tumorigenesis and metastasis. *Mol Cancer Res* **2014**.
77. Sanden, C.; Broselid, S.; Cornmark, L., et al., G protein-coupled estrogen receptor 1/G protein-coupled receptor 30 localizes in the plasma membrane and traffics intracellularly on cytokeratin intermediate filaments. *Mol Pharmacol* **2011,** 79, (3), 400-10.
78. Revankar, C. M.; Cimino, D. F.; Sklar, L. A., et al., A transmembrane intracellular estrogen receptor mediates rapid cell signaling. *Science* **2005,** 307, (5715), 1625-30.
79. Otto, C.; Rohde-Schulz, B.; Schwarz, G., et al., G protein-coupled receptor 30 localizes to the endoplasmic reticulum and is not activated by estradiol. *Endocrinology* **2008,** 149, (10), 4846-56.

80. Levin, E. R., G protein-coupled receptor 30: estrogen receptor or collaborator? *Endocrinology* **2009,** 150, (4), 1563-5.
81. Langer, G.; Bader, B.; Meoli, L., et al., A critical review of fundamental controversies in the field of GPR30 research. *Steroids* **2010,** 75, (8-9), 603-10.
82. Feldman, R. D.; Gros, R., Unraveling the mechanisms underlying the rapid vascular effects of steroids: sorting out the receptors and the pathways. *Br J Pharmacol* **2011,** 163, (6), 1163-9.
83. Funakoshi, T.; Yanai, A.; Shinoda, K., et al., G protein-coupled receptor 30 is an estrogen receptor in the plasma membrane. *Biochem Biophys Res Commun* **2006,** 346, (3), 904-10.
84. Gonzalez de Valdivia, E.; Broselid, S.; Kahn, R., et al., G protein-coupled estrogen receptor 1 (GPER1)/GPR30 increases ERK1/2 activity through PDZ motif-dependent and -independent mechanisms. *J Biol Chem* **2017,** 292, (24), 9932-9943.
85. Zhou, K.; Sun, P.; Zhang, Y., et al., Estrogen stimulated migration and invasion of estrogen receptor-negative breast cancer cells involves an ezrin-dependent crosstalk between G protein-coupled receptor 30 and estrogen receptor beta signaling. *Steroids* **2016,** 111, 113-120.
86. Pandey, D. P.; Lappano, R.; Albanito, L., et al., Estrogenic GPR30 signalling induces proliferation and migration of breast cancer cells through CTGF. *Embo j* **2009,** 28, (5), 523-32.
87. Ignatov, A.; Ignatov, T.; Roessner, A., et al., Role of GPR30 in the mechanisms of tamoxifen resistance in breast cancer MCF-7 cells. *Breast Cancer Res Treat* **2010,** 123, (1), 87-96.
88. Ahola, T. M.; Manninen, T.; Alkio, N., et al., G protein-coupled receptor 30 is critical for a progestin-induced growth inhibition in MCF-7 breast cancer cells. *Endocrinology* **2002,** 143, (9), 3376-84.
89. Ariazi, E. A.; Brailoiu, E.; Yerrum, S., et al., The G protein-coupled receptor GPR30 inhibits proliferation of estrogen receptor-positive breast cancer cells. *Cancer Res* **2010,** 70, (3), 1184-94.
90. Filardo, E.; Quinn, J.; Pang, Y., et al., Activation of the novel estrogen receptor G protein-coupled receptor 30 (GPR30) at the plasma membrane. *Endocrinology* **2007,** 148, (7), 3236-45.
91. Wei, W.; Chen, Z. J.; Zhang, K. S., et al., The activation of G protein-coupled receptor 30 (GPR30) inhibits proliferation of estrogen receptor-negative breast cancer cells in vitro and in vivo. *Cell Death Dis* **2014,** 5, e1428.
92. Filardo, E. J.; Graeber, C. T.; Quinn, J. A., et al., Distribution of GPR30, a seven membrane-spanning estrogen receptor, in primary breast cancer and its association with clinicopathologic determinants of tumor progression. *Clin Cancer Res* **2006,** 12, (21), 6359-66.
93. Ignatov, T.; Weissenborn, C.; Poehlmann, A., et al., GPER-1 expression decreases during breast cancer tumorigenesis. *Cancer Invest* **2013,** 31, (5), 309-15.
94. Kuo, W. H.; Chang, L. Y.; Liu, D. L., et al., The interactions between GPR30 and the major biomarkers in infiltrating ductal carcinoma of the breast in an Asian population. *Taiwan J Obstet Gynecol* **2007,** 46, (2), 135-45.

95. Vivacqua, A.; Romeo, E.; De Marco, P., et al., GPER mediates the Egr-1 expression induced by 17beta-estradiol and 4-hydroxitamoxifen in breast and endometrial cancer cells. *Breast Cancer Res Treat* **2012,** 133, (3), 1025-35.

96. Li, Y.; Chen, Y.; Zhu, Z. X., et al., 4-Hydroxytamoxifen-stimulated processing of cyclin E is mediated via G protein-coupled receptor 30 (GPR30) and accompanied by enhanced migration in MCF-7 breast cancer cells. *Toxicology* **2013,** 309, 61-5.

97. Ignatov, A.; Ignatov, T.; Weissenborn, C., et al., G-protein-coupled estrogen receptor GPR30 and tamoxifen resistance in breast cancer. *Breast Cancer Res Treat* **2011,** 128, (2), 457-66.

98. Catalano, S.; Giordano, C.; Panza, S., et al., Tamoxifen through GPER upregulates aromatase expression: a novel mechanism sustaining tamoxifen-resistant breast cancer cell growth. *Breast Cancer Res Treat* **2014,** 146, (2), 273-85.

99. Yu, T.; Yang, G.; Hou, Y., et al., Cytoplasmic GPER translocation in cancer-associated fibroblasts mediates cAMP/PKA/CREB/glycolytic axis to confer tumor cells with multidrug resistance. *Oncogene* **2017,** 36, (15), 2131-2145.

100. Yuan, J.; Liu, M.; Yang, L., et al., Acquisition of epithelial-mesenchymal transition phenotype in the tamoxifen-resistant breast cancer cell: a new role for G protein-coupled estrogen receptor in mediating tamoxifen resistance through cancer-associated fibroblast-derived fibronectin and beta1-integrin signaling pathway in tumor cells. *Breast Cancer Res* **2015,** 17, 69.

101. Mo, Z.; Liu, M.; Yang, F., et al., GPR30 as an initiator of tamoxifen resistance in hormone-dependent breast cancer. *Breast Cancer Res* **2013,** 15, (6), R114.

102. Darby, S.; McGale, P.; Correa, C., et al., Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet* **2011,** 378, (9804), 1707-16.

103. Pan, H.; Gray, R.; Braybrooke, J., et al., 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *N Engl J Med* **2017,** 377, (19), 1836-1846.

104. Turajlic, S.; Swanton, C., Metastasis as an evolutionary process. *Science* **2016,** 352, (6282), 169-75.

105. Nik-Zainal, S.; Van Loo, P.; Wedge, D. C., et al., The life history of 21 breast cancers. *Cell* **2012,** 149, (5), 994-1007.

106. Tang, M. H.; Dahlgren, M.; Brueffer, C., et al., Remarkable similarities of chromosomal rearrangements between primary human breast cancers and matched distant metastases as revealed by whole-genome sequencing. *Oncotarget* **2015,** 6, (35), 37169-84.

107. Hoadley, K. A.; Siegel, M. B.; Kanchi, K. L., et al., Tumor Evolution in Two Patients with Basal-like Breast Cancer: A Retrospective Genomics Study of Multiple Metastases. *PLoS Med* **2016,** 13, (12), e1002174.

108. Yates, L. R.; Knappskog, S.; Wedge, D., et al., Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* **2017,** 32, (2), 169-184.e7.

109. Robinson, D. R.; Wu, Y. M.; Lonigro, R. J., et al., Integrative clinical genomics of metastatic cancer. *Nature* **2017,** 548, (7667), 297-303.

110. Meric-Bernstam, F.; Frampton, G. M.; Ferrer-Lozano, J., et al., Concordance of genomic alterations between primary and recurrent breast cancer. *Mol Cancer Ther* **2014,** 13, (5), 1382-9.

111. Zurrida, S.; Bassi, F.; Arnone, P., et al., The Changing Face of Mastectomy (from Mutilation to Aid to Breast Reconstruction). *Int J Surg Oncol* **2011,** 2011, 980158.

112. Hosseini, A.; Khoury, A. L.; Esserman, L. J., Precision surgery and avoiding over-treatment. *Eur J Surg Oncol* **2017,** 43, (5), 938-943.

113. Edge, S. B.; Compton, C. C., The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* **2010,** 17, (6), 1471-4.

114. Elston, C. W.; Ellis, I. O., Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **1991,** 19, (5), 403-10.

115. Harris, L. N.; Ismaila, N.; McShane, L. M., et al., Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol* **2016,** 34, (10), 1134-50.

116. Fredholm, H.; Magnusson, K.; Lindstrom, L. S., et al., Long-term outcome in young women with breast cancer: a population-based study. *Breast Cancer Res Treat* **2016,** 160, (1), 131-143.

117. Bayani, J.; Yao, C. Q.; Quintayo, M. A., et al., Molecular stratification of early breast cancer identifies drug targets to drive stratified medicine. *NPJ Breast Cancer* **2017,** 3, 3.

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

119. Fredriksson, I.; Liljegren, G.; Arnesson, L. G., et al., Time trends in the results of breast conservation in 4694 women. *Eur J Cancer* **2001,** 37, (12), 1537-44.

120. Fredriksson, I.; Liljegren, G.; Palm-Sjovall, M., et al., Risk factors for local recurrence after breast-conserving surgery. *Br J Surg* **2003,** 90, (9), 1093-102.

121. Voduc, K. D.; Cheang, M. C.; Tyldesley, S., et al., Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol* **2010,** 28, (10), 1684-91.

122. Millar, E. K.; Graham, P. H.; O'Toole, S. A., et al., Prediction of local recurrence, distant metastases, and death after breast-conserving therapy in early-stage invasive breast cancer using a five-biomarker panel. *J Clin Oncol* **2009,** 27, (28), 4701-8.

123. Arvold, N. D.; Taghian, A. G.; Niemierko, A., et al., Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy. *J Clin Oncol* **2011,** 29, (29), 3885-91.

124. Nguyen, P. L.; Taghian, A. G.; Katz, M. S., et al., Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. *J Clin Oncol* **2008,** 26, (14), 2373-8.

125. Lowery, A. J.; Kell, M. R.; Glynn, R. W., et al., Locoregional recurrence after breast cancer surgery: a systematic review by receptor phenotype. *Breast Cancer Res Treat* **2012,** 133, (3), 831-41.
126. Yin, W.; Jiang, Y.; Shen, Z., et al., Trastuzumab in the adjuvant treatment of HER2-positive early breast cancer patients: a meta-analysis of published randomized controlled trials. *PLoS One* **2011,** 6, (6), e21030.
127. Mamounas, E. P.; Liu, Q.; Paik, S., et al., 21-Gene Recurrence Score and Locoregional Recurrence in Node-Positive/ER-Positive Breast Cancer Treated With Chemo-Endocrine Therapy. *J Natl Cancer Inst* **2017,** 109, (4).
128. Mamounas, E. P.; Tang, G.; Fisher, B., et al., 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), 1677-83.
129. Senkus, E.; Kyriakides, S.; Ohno, S., et al., Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **2015,** 26 Suppl 5, v8-30.
130. Fisher, B., Biological and clinical considerations regarding the use of surgery and chemotherapy in the treatment of primary breast cancer. *Cancer* **1977,** 40, (Supplement S1), 574-587.
131. Bidard, F. C.; Proudhon, C.; Pierga, J. Y., Circulating tumor cells in breast cancer. *Mol Oncol* **2016,** 10, (3), 418-30.
132. Fisher, B.; Anderson, S.; Bryant, J., et al., Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *N Engl J Med* **2002,** 347, (16), 1233-41.
133. Veronesi, U.; Cascinelli, N.; Mariani, L., et al., Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *N Engl J Med* **2002,** 347, (16), 1227-32.
134. van Maaren, M. C.; de Munck, L.; de Bock, G. H., et al., 10 year survival after breast-conserving surgery plus radiotherapy compared with mastectomy in early breast cancer in the Netherlands: a population-based study. *Lancet Oncol* **2016,** 17, (8), 1158-1170.
135. Lagendijk, M.; van Maaren, M. C.; Saadatmand, S., et al., Breast conserving therapy and mastectomy revisited: Breast cancer-specific survival and the influence of prognostic factors in 129,692 patients. *Int J Cancer* **2018,** 142, (1), 165-175.
136. Galimberti, V.; Cole, B. F.; Zurrida, S., et al., Axillary dissection versus no axillary dissection in patients with sentinel-node micrometastases (IBCSG 23-01): a phase 3 randomised controlled trial. *Lancet Oncol* **2013,** 14, (4), 297-305.
137. Giuliano, A. E.; McCall, L.; Beitsch, P., et al., Locoregional recurrence after sentinel lymph node dissection with or without axillary dissection in patients with sentinel lymph node metastases: the American College of Surgeons Oncology Group Z0011 randomized trial. *Ann Surg* **2010,** 252, (3), 426-32; discussion 432-3.
138. Giuliano, A. E.; Hunt, K. K.; Ballman, K. V., et al., Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *Jama* **2011,** 305, (6), 569-75.

139. Giuliano, A. E.; Ballman, K.; McCall, L., et al., Locoregional Recurrence After Sentinel Lymph Node Dissection With or Without Axillary Dissection in Patients With Sentinel Lymph Node Metastases: Long-term Follow-up From the American College of Surgeons Oncology Group (Alliance) ACOSOG Z0011 Randomized Trial. *Ann Surg* **2016,** 264, (3), 413-20.
140. Giuliano, A. E.; Ballman, K. V.; McCall, L., et al., Effect of Axillary Dissection vs No Axillary Dissection on 10-Year Overall Survival Among Women With Invasive Breast Cancer and Sentinel Node Metastasis: The ACOSOG Z0011 (Alliance) Randomized Clinical Trial. *Jama* **2017,** 318, (10), 918-926.
141. de Boniface, J.; Frisell, J.; Andersson, Y., et al., Survival and axillary recurrence following sentinel node-positive breast cancer without completion axillary lymph node dissection: the randomized controlled SENOMAC trial. *BMC Cancer* **2017,** 17, (1), 379.
142. Sydsvenska Bröstcancergruppen: Lathund för kirurgisk och onkologisk behandling av bröstcancer – Regional anpassning av nationellt vårdprogram, Region Väst och Syd. **2017**.
143. Kirwan, C. C.; Coles, C. E.; Bliss, J., It's PRIMETIME. Postoperative Avoidance of Radiotherapy: Biomarker Selection of Women at Very Low Risk of Local Recurrence. *Clin Oncol (R Coll Radiol)* **2016,** 28, (9), 594-6.
144. Killander, F.; Karlsson, P.; Anderson, H., et al., No breast cancer subgroup can be spared postoperative radiotherapy after breast-conserving surgery. Fifteen-year results from the Swedish Breast Cancer Group randomised trial, SweBCG 91 RT. *Eur J Cancer* **2016,** 67, 57-65.
145. Laurberg, T.; Tramm, T.; Nielsen, T., et al., Intrinsic subtypes and benefit from postmastectomy radiotherapy in node-positive premenopausal breast cancer patients who received adjuvant chemotherapy - results from two independent randomized trials. *Acta Oncol* **2017**, 1-6.
146. Nimeus-Malmström, E.; Krogh, M.; Malmström, P., et al., Gene expression profiling in primary breast cancer distinguishes patients developing local recurrence after breast-conservation surgery, with or without postoperative radiotherapy. *Breast Cancer Res* **2008,** 10, (2), R34.
147. Eschrich, S. A.; Fulp, W. J.; Pawitan, Y., et al., Validation of a radiosensitivity molecular signature in breast cancer. *Clin Cancer Res* **2012,** 18, (18), 5134-43.
148. Torres-Roca, J. F.; Fulp, W. J.; Caudell, J. J., et al., Integration of a Radiosensitivity Molecular Signature Into the Assessment of Local Recurrence Risk in Breast Cancer. *Int J Radiat Oncol Biol Phys* **2015,** 93, (3), 631-8.
149. Scott, J. G.; Berglund, A.; Schell, M. J., et al., A genome-based model for adjusting radiotherapy dose (GARD): a retrospective, cohort-based study. *Lancet Oncol* **2017,** 18, (2), 202-211.
150. Servant, N.; Bollet, M. A.; Halfwerk, H., et al., Search for a gene expression signature of breast cancer local recurrence in young women. *Clin Cancer Res* **2012,** 18, (6), 1704-15.

151. Nuyten, D. S.; Kreike, B.; Hart, A. A., et al., Predicting a local recurrence after breast-conserving therapy by gene expression profiling. *Breast Cancer Res* **2006,** 8, (5), R62.
152. Kreike, B.; Halfwerk, H.; Armstrong, N., et al., Local recurrence after breast-conserving therapy in relation to gene expression patterns in a large series of patients. *Clin Cancer Res* **2009,** 15, (12), 4181-90.
153. Kreike, B.; Halfwerk, H.; Kristel, P., et al., Gene expression profiles of primary breast carcinomas from patients at high risk for local recurrence after breast-conserving therapy. *Clin Cancer Res* **2006,** 12, (19), 5705-12.
154. Speers, C.; Zhao, S.; Liu, M., et al., Development and Validation of a Novel Radiosensitivity Signature in Human Breast Cancer. *Clin Cancer Res* **2015,** 21, (16), 3667-77.
155. Pavlopoulou, A.; Bagos, P. G.; Koutsandrea, V., et al., Molecular determinants of radiosensitivity in normal and tumor tissue: A bioinformatic approach. *Cancer Lett* **2017,** 403, 37-47.
156. Bernier, J., Precision medicine for early breast cancer radiotherapy: Opening up new horizons? *Crit Rev Oncol Hematol* **2017,** 113, 79-82.
157. Harbeck, N.; Rody, A., Lost in translation? Estrogen receptor status and endocrine responsiveness in breast cancer. *J Clin Oncol* **2012,** 30, (7), 686-9.
158. Davies, C.; Godwin, J.; Gray, R., et al., Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* **2011,** 378, (9793), 771-84.
159. Clarke, M.; Collins, R.; Darby, S., et al., Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet* **2005,** 366, (9503), 2087-106.
160. Davies, C.; Pan, H.; Godwin, J., et al., Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* **2013,** 381, (9869), 805-16.
161. Gray, R. G.; Rea, D.; Handley, K., aTTom: long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years in 6,953 women with early breast cancer. . *Journal of Clinical Oncology* **2013,** 31, (abstract 5).
162. Regionala Cancercentrum i samverkan: Bröstcancer - Nationellt vårdprogram. **2014**.
163. Peto, R.; Davies, C.; Godwin, J., et al., Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* **2012,** 379, (9814), 432-44.
164. Partridge, A. H.; Rumble, R. B.; Carey, L. A., et al., Chemotherapy and targeted therapy for women with human epidermal growth factor receptor 2-negative (or unknown) advanced breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol* **2014,** 32, (29), 3307-29.
165. Cardoso, F.; Costa, A.; Senkus, E., et al., 3rd ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 3). *Ann Oncol* **2017,** 28, (1), 16-33.

166. Wallwiener, M.; Hartkopf, A. D.; Riethdorf, S., et al., The impact of HER2 phenotype of circulating tumor cells in metastatic breast cancer: a retrospective study in 107 patients. *BMC Cancer* **2015,** 15, 403.

167. Jaeger, B. A. S.; Neugebauer, J.; Andergassen, U., et al., The HER2 phenotype of circulating tumor cells in HER2-positive early breast cancer: A translational research project of a prospective randomized phase III trial. *PLoS One* **2017,** 12, (6), e0173593.

168. Jordan, N. V.; Bardia, A.; Wittner, B. S., et al., HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* **2016,** 537, (7618), 102-106.

169. Finn, R. S.; Martin, M.; Rugo, H. S., et al., Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med* **2016,** 375, (20), 1925-1936.

170. Hortobagyi, G. N.; Stemmer, S. M.; Burris, H. A., et al., Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. *N Engl J Med* **2016,** 375, (18), 1738-1748.

171. Li, X.; Oprea-Ilies, G. M.; Krishnamurti, U., New Developments in Breast Cancer and Their Impact on Daily Practice in Pathology. *Arch Pathol Lab Med* **2017,** 141, (4), 490-498.

172. Kristensen, V. N.; Lingjaerde, O. C.; Russnes, H. G., et al., Principles and methods of integrative genomic analyses in cancer. *Nat Rev Cancer* **2014,** 14, (5), 299-313.

173. Rolland, D. C. M.; Basrur, V.; Jeon, Y. K., et al., Functional proteogenomics reveals biomarkers and therapeutic targets in lymphomas. *Proc Natl Acad Sci U S A* **2017,** 114, (25), 6581-6586.

174. Zhu, Y.; Engstrom, P. G.; Tellgren-Roth, C., et al., Proteogenomics produces comprehensive and highly accurate protein-coding gene annotation in a complete genome assembly of Malassezia sympodialis. *Nucleic Acids Res* **2017,** 45, (5), 2629-2643.

175. Zhang, Y.; Kwok-Shing Ng, P.; Kucherlapati, M., et al., A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations. *Cancer Cell* **2017,** 31, (6), 820-832.e3.

176. Conrads, T. P.; Petricoin, E. F., 3rd, The Obama Administration's Cancer Moonshot: A Call for Proteomics. *Clin Cancer Res* **2016,** 22, (18), 4556-8.

177. Singer, D. S.; Jacks, T.; Jaffee, E., A U.S. "Cancer Moonshot" to accelerate cancer research. *Science* **2016,** 353, (6304), 1105-6.

178. Bartlett, J. M. S.; Parelukar, W., Breast cancers are rare diseases-and must be treated as such. *NPJ Breast Cancer* **2017,** 3, 11.

179. Gao, H.; Korn, J. M.; Ferretti, S., et al., High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* **2015,** 21, (11), 1318-25.

180. Hudis, C. A.; Barlow, W. E.; Costantino, J. P., et al., Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol* **2007,** 25, (15), 2127-32.

181. Gourgou-Bourgade, S.; Cameron, D.; Poortmans, P., et al., Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials)dagger. *Ann Oncol* **2015,** 26, (5), 873-9.
182. Rutqvist, L. E.; Johansson, H.; Stockholm Breast Cancer Study, G., Long-term follow-up of the randomized Stockholm trial on adjuvant tamoxifen among postmenopausal patients with early stage breast cancer. *Acta Oncol* **2007,** 46, (2), 133-45.
183. Rutqvist, L. E.; Cedermark, B.; Glas, U., et al., The Stockholm trial on adjuvant tamoxifen in early breast cancer. Correlation between estrogen receptor level and treatment effect. *Breast Cancer Res Treat* **1987,** 10, (3), 255-66.
184. Ryden, L.; Jonsson, P. E.; Chebil, G., et al., Two years of adjuvant tamoxifen in premenopausal patients with breast cancer: a randomised, controlled trial with long-term follow-up. *Eur J Cancer* **2005,** 41, (2), 256-64.
185. Randomized trial of two versus five years of adjuvant tamoxifen for postmenopausal early stage breast cancer. Swedish Breast Cancer Cooperative Group. *J Natl Cancer Inst* **1996,** 88, (21), 1543-9.
186. Malmström, P.; Holmberg, L.; Anderson, H., et al., Breast conservation surgery, with and without radiotherapy, in women with lymph node-negative breast cancer: a randomised clinical trial in a population with access to public mammography screening. *Eur J Cancer* **2003,** 39, (12), 1690-7.
187. Prat, A.; Karginova, O.; Parker, J. S., et al., Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat* **2013,** 142, (2), 237-55.
188. Elliott, K.; McQuaid, S.; Salto-Tellez, M., et al., Immunohistochemistry should undergo robust validation equivalent to that of molecular diagnostics. *J Clin Pathol* **2015,** 68, (10), 766-70.
189. Kononen, J.; Bubendorf, L.; Kallioniemi, A., et al., Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* **1998,** 4, (7), 844-7.
190. Martelotto, L. G.; Ng, C. K.; Piscuoglio, S., et al., Breast cancer intra-tumor heterogeneity. *Breast Cancer Res* **2014,** 16, (3), 210.
191. Kyndi, M.; Sorensen, F. B.; Knudsen, H., et al., Tissue microarrays compared with whole sections and biochemical analyses. A subgroup analysis of DBCG 82 b&c. *Acta Oncol* **2008,** 47, (4), 591-9.
192. Crick, F., Central dogma of molecular biology. *Nature* **1970,** 227, (5258), 561-3.
193. Li, J. J.; Bickel, P. J.; Biggin, M. D., System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2014,** 2, e270.
194. Jovanovic, M.; Rooney, M. S.; Mertins, P., et al., Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **2015,** 347, (6226), 1259038.
195. Battle, A.; Khan, Z.; Wang, S. H., et al., Genomic variation. Impact of regulatory variation from RNA to protein. *Science* **2015,** 347, (6222), 664-7.

196. Chandra Gupta, S.; Nandan Tripathi, Y., Potential of long non-coding RNAs in cancer patients: From biomarkers to therapeutic targets. *Int J Cancer* **2017,** 140, (9), 1955-1967.

197. Nassar, F. J.; Nasr, R.; Talhouk, R., MicroRNAs as biomarkers for early breast cancer diagnosis, prognosis and therapy prediction. *Pharmacol Ther* **2017,** 172, 34-49.

198. Gruvberger-Saal, S. K.; Cunliffe, H. E.; Carr, K. M., et al., Microarrays in breast cancer research and clinical practice--the future lies ahead. *Endocr Relat Cancer* **2006,** 13, (4), 1017-31.

199. Kuhn, K.; Baker, S. C.; Chudin, E., et al., A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res* **2004,** 14, (11), 2347-56.

200. Ritchie, M. E.; Dunning, M. J.; Smith, M. L., et al., BeadArray expression analysis using bioconductor. *PLoS Comput Biol* **2011,** 7, (12), e1002276.

201. Tyekucheva, S.; Martin, N. E.; Stack, E. C., et al., Comparing Platforms for Messenger RNA Expression Profiling of Archival Formalin-Fixed, Paraffin-Embedded Tissues. *J Mol Diagn* **2015,** 17, (4), 374-81.

202. Taylor, S. C.; Laperriere, G.; Germain, H., Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci Rep* **2017,** 7, (1), 2409.

203. Geiss, G. K.; Bumgarner, R. E.; Birditt, B., et al., Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **2008,** 26, (3), 317-25.

204. Byron, S. A.; Van Keuren-Jensen, K. R.; Engelthaler, D. M., et al., Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* **2016,** 17, (5), 257-71.

205. Barbosa-Morais, N. L.; Dunning, M. J.; Samarajiwa, S. A., et al., A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res* **2010,** 38, (3), e17.

206. Tsang, H. F.; Xue, V. W.; Koh, S. P., et al., NanoString, a novel digital color-coded barcode technology: current and future applications in molecular diagnostics. *Expert Rev Mol Diagn* **2017,** 17, (1), 95-103.

207. Jung, S. H.; Sohn, I., Statistical Issues in the Design and Analysis of nCounter Projects. *Cancer Inform* **2014,** 13, (Suppl 7), 35-43.

208. Tyanova, S.; Albrechtsen, R.; Kronqvist, P., et al., Proteomic maps of breast cancer subtypes. *Nat Commun* **2016,** 7, 10259.

209. Panis, C.; Pizzatti, L.; Souza, G. F., et al., Clinical proteomics in cancer: Where we are. *Cancer Lett* **2016,** 382, (2), 231-239.

210. Mardamshina, M.; Geiger, T., Next-Generation Proteomics and Its Application to Clinical Breast Cancer Research. *Am J Pathol* **2017,** 187, (10), 2175-2184.

211. Aebersold, R.; Mann, M., Mass-spectrometric exploration of proteome structure and function. *Nature* **2016,** 537, (7620), 347-55.

212. Lam, S. W.; Jimenez, C. R.; Boven, E., Breast cancer classification by proteomic technologies: current state of knowledge. *Cancer Treat Rev* **2014,** 40, (1), 129-38.

213. Waldemarson, S.; Kurbasic, E.; Krogh, M., et al., Proteomic analysis of breast tumors confirms the mRNA intrinsic molecular subtypes using different classifiers: a large-scale analysis of fresh frozen tissue samples. *Breast Cancer Res* **2016,** 18, (1), 69.

214. Sjöholm, K., A holistic approach to host-pathogen interactions. Detecting the large to unravel the small. *Lund University, Faculty of Engineering, Dissertation Series* **2017**.

215. Rodriguez, J.; Gupta, N.; Smith, R. D., et al., Does trypsin cut before proline? *J Proteome Res* **2008,** 7, (1), 300-5.

216. Feist, P.; Hummon, A. B., Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **2015,** 16, (2), 3537-63.

217. Li, X.; Franz, T., Up to date sample preparation of proteins for mass spectrometric analysis. *Arch Physiol Biochem* **2014,** 120, (5), 188-91.

218. Li, S.; Tang, H., Computational Methods in Mass Spectrometry-Based Proteomics. *Adv Exp Med Biol* **2016,** 939, 63-89.

219. Murray, H. C.; Dun, M. D.; Verrills, N. M., Harnessing the power of proteomics for identification of oncogenic, druggable signalling pathways in cancer. *Expert Opin Drug Discov* **2017,** 12, (5), 431-447.

220. Nesvizhskii, A. I., Proteogenomics: concepts, applications and computational strategies. *Nat Methods* **2014,** 11, (11), 1114-25.

221. Vidova, V.; Spacil, Z., A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Anal Chim Acta* **2017,** 964, 7-23.

222. Kusebauch, U.; Campbell, D. S.; Deutsch, E. W., et al., Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* **2016,** 166, (3), 766-778.

223. Faria, S. S.; Morris, C. F.; Silva, A. R., et al., A Timely Shift from Shotgun to Targeted Proteomics and How It Can Be Groundbreaking for Cancer Research. *Front Oncol* **2017,** 7, 13.

224. Picotti, P.; Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **2012,** 9, (6), 555-66.

225. Sajic, T.; Liu, Y.; Aebersold, R., Using data-independent, high-resolution mass spectrometry in protein biomarker research: perspectives and clinical applications. *Proteomics Clin Appl* **2015,** 9, (3-4), 307-21.

226. Anjo, S. I.; Santa, C.; Manadas, B., SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications. *Proteomics* **2017,** 17, (3-4).

227. Sandin, M.; Teleman, J.; Malmstrom, J., et al., Data processing methods and quality control strategies for label-free LC-MS protein quantification. *Biochim Biophys Acta* **2014,** 1844, (1 Pt A), 29-41.

228. Lindemann, C.; Thomanek, N.; Hundt, F., et al., Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biol Chem* **2017,** 398, (5-6), 687-699.

229. Chen, Y.; Wang, F.; Xu, F., et al., Mass Spectrometry-Based Protein Quantification. *Adv Exp Med Biol* **2016,** 919, 255-279.
230. Reiter, L.; Rinner, O.; Picotti, P., et al., mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* **2011,** 8, (5), 430-5.
231. Teleman, J.; Karlsson, C.; Waldemarson, S., et al., Automated selected reaction monitoring software for accurate label-free protein quantification. *J Proteome Res* **2012,** 11, (7), 3766-73.
232. MacLean, B.; Tomazela, D. M.; Shulman, N., et al., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** 26, (7), 966-8.
233. Bereman, M. S.; MacLean, B.; Tomazela, D. M., et al., The development of selected reaction monitoring methods for targeted proteomics via empirical refinement. *Proteomics* **2012,** 12, (8), 1134-41.
234. Stenemo, M.; Teleman, J.; Sjostrom, M., et al., Cancer associated proteins in blood plasma: Determining normal variation. *Proteomics* **2016,** 16, (13), 1928-37.
235. Gianazza, E.; Miller, I.; Palazzolo, L., et al., With or without you - Proteomics with or without major plasma/serum proteins. *J Proteomics* **2016,** 140, 62-80.
236. Zhang, H.; Liu, A. Y.; Loriaux, P., et al., Mass spectrometric detection of tissue proteins in plasma. *Mol Cell Proteomics* **2007,** 6, (1), 64-71.
237. Zielinska, D. F.; Gnad, F.; Wisniewski, J. R., et al., Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010,** 141, (5), 897-907.
238. Tian, Y.; Zhang, H., Characterization of disease-associated N-linked glycoproteins. *Proteomics* **2013,** 13, (3-4), 504-11.
239. Tian, Y.; Zhou, Y.; Elliott, S., et al., Solid-phase extraction of N-linked glycopeptides. *Nat Protoc* **2007,** 2, (2), 334-9.
240. Lai, Z. W.; Nice, E. C.; Schilling, O., Glycocapture-based proteomics for secretome analysis. *Proteomics* **2013,** 13, (3-4), 512-25.
241. Altman, D. G.; Bland, J. M., Uncertainty and sampling error. *Bmj* **2014,** 349, g7064.
242. Altman, D. G.; Bland, J. M., Uncertainty beyond sampling error. *Bmj* **2014,** 349, g7065.
243. Thomas, E., An introduction to medical statistics for health care professionals: basic statistical tests. *Musculoskeletal Care* **2005,** 3, (4), 201-12.
244. Altman, D. G.; Bland, J. M., Statistics notes: variables and parameters. *Bmj* **1999,** 318, (7199), 1667.
245. Bland, J. M.; Altman, D. G., Survival probabilities (the Kaplan-Meier method). *BMJ* **1998,** 317, (7172), 1572.
246. Bland, J. M.; Altman, D. G., The logrank test. *BMJ* **2004,** 328, (7447), 1073.
247. Kaplan, E. L.; Meier, P., Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* **1958,** 53, (282), 457-481.

248. Cox, D. R., Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* **1972,** 34, (2), 187-220.
249. Schoenfeld, D., Partial residuals for the proportional hazards regression model. *Biometrika* **1982,** 69, (1), 239-241.
250. Dignam, J. J.; Zhang, Q.; Kocherginsky, M., The use and interpretation of competing risks regression models. *Clin Cancer Res* **2012,** 18, (8), 2301-8.
251. Koller, M. T.; Raatz, H.; Steyerberg, E. W., et al., Competing risks and the clinical community: irrelevance or ignorance? *Stat Med* **2012,** 31, (11-12), 1089-97.
252. Satagopan, J. M.; Ben-Porat, L.; Berwick, M., et al., A note on competing risks in survival data analysis. *Br J Cancer* **2004,** 91, (7), 1229-35.
253. Kim, H. T., Cumulative incidence in competing risks data and competing risks regression analysis. *Clin Cancer Res* **2007,** 13, (2 Pt 1), 559-65.
254. Dignam, J. J.; Kocherginsky, M. N., Choice and interpretation of statistical tests used when competing risks are present. *J Clin Oncol* **2008,** 26, (24), 4027-34.
255. Chappell, R., Competing risk analyses: how are they different and why should you care? *Clin Cancer Res* **2012,** 18, (8), 2127-9.
256. Andersen, P. K.; Geskus, R. B.; de Witte, T., et al., Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol* **2012,** 41, (3), 861-70.
257. Haller, B.; Schmidt, G.; Ulm, K., Applying competing risks regression models: an overview. *Lifetime Data Anal* **2013,** 19, (1), 33-58.
258. Putter, H.; Fiocco, M.; Geskus, R. B., Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* **2007,** 26, (11), 2389-430.
259. Fine, J. P.; Gray, R. J., A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* **1999,** 94, (446), 496-509.
260. Pintilie, M., Analysing and interpreting competing risk data. *Stat Med* **2007,** 26, (6), 1360-7.
261. Geskus, R. B., Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics* **2011,** 67, (1), 39-49.
262. Tai, B. C.; Wee, J.; Machin, D., Analysis and design of randomised clinical trials involving competing risks endpoints. *Trials* **2011,** 12, 127.
263. Kuk, D.; Varadhan, R., Model selection in competing risks regression. *Stat Med* **2013,** 32, (18), 3077-88.
264. Altman, D. G.; Bland, J. M., Interaction revisited: the difference between two estimates. *Bmj* **2003,** 326, (7382), 219.
265. Altman, D. G.; Matthews, J. N., Statistics notes. Interaction 1: Heterogeneity of effects. *Bmj* **1996,** 313, (7055), 486.
266. Matthews, J. N.; Altman, D. G., Statistics notes. Interaction 2: Compare effect sizes not P values. *Bmj* **1996,** 313, (7060), 808.
267. Matthews, J. N.; Altman, D. G., Interaction 3: How to examine heterogeneity. *Bmj* **1996,** 313, (7061), 862.

268. Altman, D. G.; Bland, J. M., Diagnostic tests 3: receiver operating characteristic plots. *Bmj* **1994,** 309, (6948), 188.

269. Zweig, M. H.; Campbell, G., Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* **1993,** 39, (4), 561-77.

270. Altman, D. G.; Bland, J. M., Diagnostic tests. 1: Sensitivity and specificity. *Bmj* **1994,** 308, (6943), 1552.

271. Altman, D. G.; Bland, J. M., Diagnostic tests 2: Predictive values. *Bmj* **1994,** 309, (6947), 102.

272. James, G.; Witten, D.; Hastie, T., et al., *An Introduction to Statistical Learning - with implications in R*. Springer: New York, 2013.

273. Ferte, C.; Trister, A. D.; Huang, E., et al., Impact of bioinformatic procedures in the development and translation of high-throughput molecular classifiers in oncology. *Clin Cancer Res* **2013,** 19, (16), 4315-25.

274. Quackenbush, J., Microarray data normalization and transformation. *Nat Genet* **2002,** 32 Suppl, 496-501.

275. Welle, S., What statisticians should know about microarray gene expression technology. *Methods Mol Biol* **2013,** 972, 1-13.

276. Piccolo, S. R.; Sun, Y.; Campbell, J. D., et al., A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics* **2012,** 100, (6), 337-44.

277. Dunning, M.; Lynch, A.; Eldridge, M., illuminaHumanv4.db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4) *R package version 1.26.0.* **2015**.

278. Goh, W. W. B.; Wang, W.; Wong, L., Why Batch Effects Matter in Omics Data, and How to Avoid Them. *Trends Biotechnol* **2017,** 35, (6), 498-507.

279. Nygaard, V.; Rodland, E. A.; Hovig, E., Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics* **2016,** 17, (1), 29-39.

280. Paquet, E. R.; Hallett, M. T., Absolute assignment of breast cancer intrinsic molecular subtype. *J Natl Cancer Inst* **2015,** 107, (1), 357.

281. Kuhn, M.; Johnson, K., *Applied Predictive Modeling*. Springer: New York, 2013.

282. Afsari, B.; Fertig, E. J.; Geman, D., et al., switchBox: an R package for k-Top Scoring Pairs classifier development. *Bioinformatics* **2015,** 31, (2), 273-4.

283. Shi, P.; Ray, S.; Zhu, Q., et al., Top scoring pairs for feature selection in machine learning and applications to cancer outcome prediction. *BMC Bioinformatics* **2011,** 12, 375.

284. Liberzon, A.; Birger, C.; Thorvaldsdottir, H., et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **2015,** 1, (6), 417-425.

285. Lauss, M.; Ringner, M.; Hoglund, M., Prediction of stage, grade, and survival in bladder cancer using genome-wide expression data: a validation study. *Clin Cancer Res* **2010,** 16, (17), 4421-33.

286. Irizarry, R. A.; Wang, C.; Zhou, Y., et al., Gene set enrichment analysis made simple. *Stat Methods Med Res* **2009,** 18, (6), 565-75.

287. Mi, H.; Huang, X.; Muruganujan, A., et al., PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* **2017,** 45, (D1), D183-d189.

288. Szklarczyk, D.; Morris, J. H.; Cook, H., et al., The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **2017,** 45, (D1), D362-d368.

289. Fredlund, E.; Staaf, J.; Rantala, J. K., et al., The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. *Breast Cancer Res* **2012,** 14, (4), R113.

290. Ringner, M.; Fredlund, E.; Hakkinen, J., et al., GOBO: gene expression-based outcome for breast cancer online. *PLoS One* **2011,** 6, (3), e17911.

291. Varma, S.; Simon, R., Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **2006,** 7, 91.

292. Ambroise, C.; McLachlan, G. J., Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci U S A* **2002,** 99, (10), 6562-6.

293. Kimbung, S.; Kovacs, A.; Danielsson, A., et al., Contrasting breast cancer molecular subtypes across serial tumor progression stages: biological and prognostic implications. *Oncotarget* **2015,** 6, (32), 33306-18.

294. Tramm, T.; Mohammed, H.; Myhre, S., et al., Development and validation of a gene profile predicting benefit of postmastectomy radiotherapy in patients with high-risk breast cancer: a study of gene expression in the DBCG82bc cohort. *Clin Cancer Res* **2014,** 20, (20), 5272-80.

295. van de Vijver, M. J.; He, Y. D.; van't Veer, L. J., et al., A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **2002,** 347, (25), 1999-2009.

296. Strom, T.; Harrison, L. B.; Giuliano, A. R., et al., Tumour radiosensitivity is associated with immune activation in solid tumours. *Eur J Cancer* **2017,** 84, 304-314.

297. Nagalla, S.; Chou, J. W.; Willingham, M. C., et al., Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome Biol* **2013,** 14, (4), R34.

298. Cance, W. G.; Liu, E. T., Protein kinases in human breast cancer. *Breast Cancer Res Treat* **1995,** 35, (1), 105-14.

299. Gyorffy, B.; Lanczky, A.; Eklund, A. C., et al., An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **2010,** 123, (3), 725-31.

300. Thorat, M. A.; Cuzick, J., Preventing invasive breast cancer using endocrine therapy. *Breast* **2017,** 34 Suppl 1, S47-s54.

301. Chlebowski, R. T.; Geller, M. L., Adherence to endocrine therapy for breast cancer. *Oncology* **2006,** 71, (1-2), 1-9.

302. Makubate, B.; Donnan, P. T.; Dewar, J. A., et al., Cohort study of adherence to adjuvant endocrine therapy, breast cancer recurrence and mortality. *Br J Cancer* **2013,** 108, (7), 1515-24.

303. Blamey, R. W.; Bates, T.; Chetty, U., et al., Radiotherapy or tamoxifen after conserving surgery for breast cancers of excellent prognosis: British Association of Surgical Oncology (BASO) II trial. *Eur J Cancer* **2013,** 49, (10), 2294-302.
304. Fisher, B.; Bryant, J.; Dignam, J. J., et al., Tamoxifen, radiation therapy, or both for prevention of ipsilateral breast tumor recurrence after lumpectomy in women with invasive breast cancers of one centimeter or less. *J Clin Oncol* **2002,** 20, (20), 4141-9.
305. Fyles, A. W.; McCready, D. R.; Manchul, L. A., et al., Tamoxifen with or without breast irradiation in women 50 years of age or older with early breast cancer. *N Engl J Med* **2004,** 351, (10), 963-70.
306. Potter, R.; Gnant, M.; Kwasny, W., et al., Lumpectomy plus tamoxifen or anastrozole with or without whole breast irradiation in women with favorable early breast cancer. *Int J Radiat Oncol Biol Phys* **2007,** 68, (2), 334-40.
307. Hughes, K. S.; Schnaper, L. A.; Bellon, J. R., et al., Lumpectomy plus tamoxifen with or without irradiation in women age 70 years or older with early breast cancer: long-term follow-up of CALGB 9343. *J Clin Oncol* **2013,** 31, (19), 2382-7.
308. Winzer, K. J.; Sauerbrei, W.; Braun, M., et al., Radiation therapy and tamoxifen after breast-conserving surgery: updated results of a 2 x 2 randomised clinical trial in patients with low risk of recurrence. *Eur J Cancer* **2010,** 46, (1), 95-101.
309. Kunkler, I. H.; Williams, L. J.; Jack, W. J., et al., Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): a randomised controlled trial. *Lancet Oncol* **2015,** 16, (3), 266-73.
310. van 't Veer, L. J.; Yau, C.; Yu, N. Y., et al., Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Res Treat* **2017**.
311. Esserman, L. J.; Yau, C.; Thompson, C. K., et al., Use of Molecular Tools to Identify Patients With Indolent Breast Cancers With Ultralow Risk Over 2 Decades. *JAMA Oncol* **2017**.
312. Weissenborn, C.; Ignatov, T.; Nass, N., et al., GPER Promoter Methylation Controls GPER Expression in Breast Cancer Patients. *Cancer Invest* **2017,** 35, (2), 100-107.
313. Box, G. E. P., Science and Statistics. *Journal of the American Statistical Association* **1976,** 71, (356), 791-799.
314. Altman, D. G.; Bland, J. M., Statistics notes. Treatment allocation in controlled trials: why randomise? *Bmj* **1999,** 318, (7192), 1209.
315. Rundle, A. G.; Vineis, P.; Ahsan, H., Design options for molecular epidemiology research within cohort studies. *Cancer Epidemiol Biomarkers Prev* **2005,** 14, (8), 1899-907.
316. Shen, Y.; Zhang, S.; Zhou, J., et al., Cohort Research in "Omics" and Preventive Medicine. *Adv Exp Med Biol* **2017,** 1005, 193-220.
317. Peto, R., Current misconception 3: that subgroup-specific trial mortality results often provide a good basis for individualising patient care. *Br J Cancer* **2011,** 104, (7), 1057-8.
318. Bland, J. M.; Altman, D. G., Comparisons within randomised groups can be very misleading. *Bmj* **2011,** 342, d561.

319. Bland, J. M.; Altman, D. G., Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials* **2011,** 12, 264.

320. Altman, D. G.; Bland, J. M., Absence of evidence is not evidence of absence. *Bmj* **1995,** 311, (7003), 485.

321. Vaidyanathan, G., Redefining clinical trials: the age of personalized medicine. *Cell* **2012,** 148, (6), 1079-80.

322. Sjostrom, M.; Hartman, L.; Honeth, G., et al., Stem cell biomarker ALDH1A1 in breast cancer shows an association with prognosis and clinicopathological variables that is highly cut-off dependent. *J Clin Pathol* **2015**.

323. Denkert, C.; Loibl, S.; Muller, B. M., et al., Ki67 levels as predictive and prognostic parameters in pretherapeutic breast cancer core biopsies: a translational investigation in the neoadjuvant GeparTrio trial. *Ann Oncol* **2013,** 24, (11), 2786-93.

324. Denkert, C.; Budczies, J.; von Minckwitz, G., et al., Strategies for developing Ki67 as a useful biomarker in breast cancer. *Breast* **2015,** 24 Suppl 2, S67-72.

325. Storey, J. D.; Tibshirani, R., Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **2003,** 100, (16), 9440-5.

326. Benjamini, Y.; Hochberg, Y., Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **1995,** 57, (1), 289-300.

327. Benjamini, Y., Discovering the false discovery rate. *Journal of the Royal Statistical Society. Series B (Methodological)* **2010,** 72, (Part 4.), 405-416.

328. Ludbrook, J., Statistics in biomedical laboratory and clinical science: applications, issues and pitfalls. *Med Princ Pract* **2008,** 17, (1), 1-13.

329. Bennet, C. M.; Baird, A. A.; Miller, M. B., et al., Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: An argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results.* **2010,** 1, (1), 1-5.

330. Sachs, M. C., Statistical principles for omics-based clinical trials. *Chin Clin Oncol* **2015,** 4, (3), 29.

331. Sekula, P.; Mallett, S.; Altman, D. G., et al., Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles. *PLoS One* **2017,** 12, (6), e0178531.

332. McShane, L. M.; Altman, D. G.; Sauerbrei, W., et al., REporting recommendations for tumour MARKer prognostic studies (REMARK). *Eur J Cancer* **2005,** 41, (12), 1690-6.

333. Boja, E. S.; Rodriguez, H., Mass spectrometry-based targeted quantitative proteomics: achieving sensitive and reproducible detection of proteins. *Proteomics* **2012,** 12, (8), 1093-110.

334. Gallego Romero, I.; Pai, A. A.; Tung, J., et al., RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol* **2014,** 12, 42.

335. Dowsett, M.; Goldhirsch, A.; Hayes, D. F., et al., International Web-based consultation on priorities for translational breast cancer research. *Breast Cancer Res* **2007,** 9, (6), R81.

336. Abrams, J.; Conley, B.; Mooney, M., et al., National Cancer Institute's Precision Medicine Initiatives for the new National Clinical Trials Network. *Am Soc Clin Oncol Educ Book* **2014**, 71-6.
337. Dietel, M.; Johrens, K.; Laffert, M. V., et al., A 2015 update on predictive molecular pathology and its role in targeted cancer therapy: a review focussing on clinical relevance. *Cancer Gene Ther* **2015,** 22, (9), 417-30.
338. Paquet, E. R.; Lesurf, R.; Tofigh, A., et al., Detecting gene signature activation in breast cancer in an absolute, single-patient manner. *Breast Cancer Res* **2017,** 19, (1), 32.
339. Diamandis, E. P., Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics* **2004,** 3, (4), 367-78.
340. Alfaro, J. A.; Sinha, A.; Kislinger, T., et al., Onco-proteogenomics: cancer proteomics joins forces with genomics. *Nat Methods* **2014,** 11, (11), 1107-13.
341. Woo, S.; Cha, S. W.; Na, S., et al., Proteogenomic strategies for identification of aberrant cancer peptides using large-scale next-generation sequencing data. *Proteomics* **2014,** 14, (23-24), 2719-30.
342. Rivers, R. C.; Kinsinger, C.; Boja, E. S., et al., Linking cancer genome to proteome: NCI's investment into proteogenomics. *Proteomics* **2014,** 14, (23-24), 2633-6.
343. Edwards, N. J.; Oberti, M.; Thangudu, R. R., et al., The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *J Proteome Res* **2015**.
344. Shukla, H. D.; Mahmood, J.; Vujaskovic, Z., Integrated proteo-genomic approach for early diagnosis and prognosis of cancer. *Cancer Lett* **2015**.
345. Ruggles, K. V.; Tang, Z.; Wang, X., et al., An analysis of the sensitivity of proteogenomic mapping of somatic mutations and novel splicing events in cancer. *Mol Cell Proteomics* **2015**.
346. Ruggles, K. V.; Krug, K.; Wang, X., et al., Methods, Tools and Current Perspectives in Proteogenomics. *Mol Cell Proteomics* **2017,** 16, (6), 959-981.

Printed by Media-Tryck, Lund 2017 NORDIC ECOLABEL, 3041 0903

Martin Sjöström, M.D., was born in 1988 in Linköping in Östergötland. Besides working towards precision medicine in oncology, he enjoys training for road cycling races and triathlons, building and flying model aircrafts, being in the nature (especially fly fishing or scuba diving), and travelling. He also holds a bachelor's degree in Latin, with an essay on Virgil's use of physiological metaphors for psychological phenomena in the Aeneid (Själens spegel – psykets kroppsliga uttryck i Vergilius Aeneiden).

***

Precision medicine will transform healthcare by using individualized treatment strategies and approaches. In oncology, this means that the exact changes in individual tumors will be exploited to assess the risk of recurrence, guide treatment decisions, and ultimately targeted with new treatment strategies. This thesis presents five studies towards better characterization of primary breast cancer by using high-throughput techniques, and further explores how the characterization can lead to individualized adjuvant treatment after surgery.

LUND UNIVERSITY
Faculty of Medicine

Department of Clinical Sciences Lund
Division of Oncology and Pathology
Division of Surgery
Lund University, Faculty of Medicine
Doctoral Dissertation Series 2018:8
ISBN 978-91-7619-575-8
ISSN 1652-8220