

# **Discovery of Protein Biomarkers Associated to Tamoxifen Resistance**

**Tommaso De Marchi**

The studies described in this thesis were performed at the Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands. The research described here was performed within the framework of the Erasmus Postgraduate School of Molecular Medicine.



This research project was supported by the Dutch Cancer Society (KWF).

Support for the printing was obtained by the Department of Medical Oncology, Erasmus University, Carl Zeiss BV, and New England Peptide Inc.

Cover design, printing and binding by Ridderprint BV, Ridderkerk.

# **Discovery of Protein Biomarkers Associated to Tamoxifen Resistance**

**De ontdekking van eiwit biomerkers geassocieerd aan tamoxifen resistentie**

**Thesis**

**To obtain the degree of Doctor from the**

**Erasmus University Rotterdam**

**By command of the**

**Rector Magnificus**

**Prof.dr. H.A.P. Pols**

**and in accordance with the decision of the Doctorate Board**

**The public defense shall be held on**

*Thursday 13<sup>th</sup> October 2016, at 15:30 hours*

**Tommaso De Marchi**

**born in San Vito al tagliamento, Italy**

**Erasmus University Rotterdam**

The logo of Erasmus University Rotterdam, featuring the word "Erasmus" in a stylized, cursive script.

**Doctoral committee**

**Supervisor:** Prof.dr. J.A. Foekens

**Other members:** Prof.dr. E.C. Zwarthoff

Prof.dr. C. Verhoef

Prof.dr. C.G.J. Sweep

**Co-supervisors:** Dr. A. Umar

Dr.ir. J.W.M. Martens

**SPONSOR OF THE THESIS**



*This thesis is dedicated to my family.*

# CONTENTS

<b>Chapter 1</b>	<b>Introduction</b>	<b>7-31</b>
	1.1 Breast cancer classification	
	1.2 ER positive breast cancer	
	1.3 General mechanisms of endocrine therapy resistance	
	1.4 Prediction of tamoxifen therapy resistance	
	1.5 The proteomic approach	
	Part of this chapter is derived from: <i>Drug Discov Today</i> 2016; epub ahead of print	
<b>Chapter 2</b>	<b>Aim and outline of this thesis</b>	<b>33-36</b>
<b>Chapter 3</b>	<b>The advantage of laser-capture microdissection over whole tissue analysis in proteomic profiling studies</b>	<b>37-58</b>
	<i>Proteomics</i> 2016; 16, 10, 1474-85	
<b>Chapter 4</b>	<b>4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer</b>	<b>59-90</b>
	<i>Mol Oncol</i> 2016; 10, 1: 24-39	
<b>Chapter 5</b>	<b>Antibody-Based Capture of Target Peptides in Multiple Reaction Monitoring Experiments</b>	<b>91-108</b>
	<i>Methods Mol Biol</i> 2015;1293:123-35	
<b>Chapter 6</b>	<b>Targeted MS assay quantifies proteins predicting tamoxifen resistance in estrogen receptor positive breast cancer</b>	<b>109-138</b>
	<i>J Proteome Res</i> , 2016; 15, 4, 1230-42	
<b>Chapter 7</b>	<b>Annexin-A1 and Caldesmon are associated with resistance to tamoxifen in estrogen receptor positive recurrent breast cancer</b>	<b>139-164</b>
	<i>Oncotarget</i> 2016; 17, 7: 3098-110	

<b>Chapter 8</b>	<b>Phosphoserine aminotransferase 1 is associated to poor outcome on tamoxifen therapy in recurrent breast cancer</b>	<b>165-190</b>
	<i>Submitted for publication</i>	
<b>Chapter 9</b>	<b>Discussion</b>	<b>191-212</b>
<b>Chapter 10</b>	<b>Samenvatting/Summary/Sommario</b>	<b>213-222</b>
<b>Appendices</b>	Acknowledgements	<b>223-236</b>
	List of Publications	
	PhD Portfolio	
	Curriculum Vitae	

# Chapter 1

## Introduction

Part of this chapter is derived from:

### **Endocrine therapy resistance in estrogen receptor (ER) positive breast cancer**

Tommaso De Marchi, John A. Foekens, Arzu Umar, John W. Martens.

*Drug Discov Today*. 2016; epub ahead of print

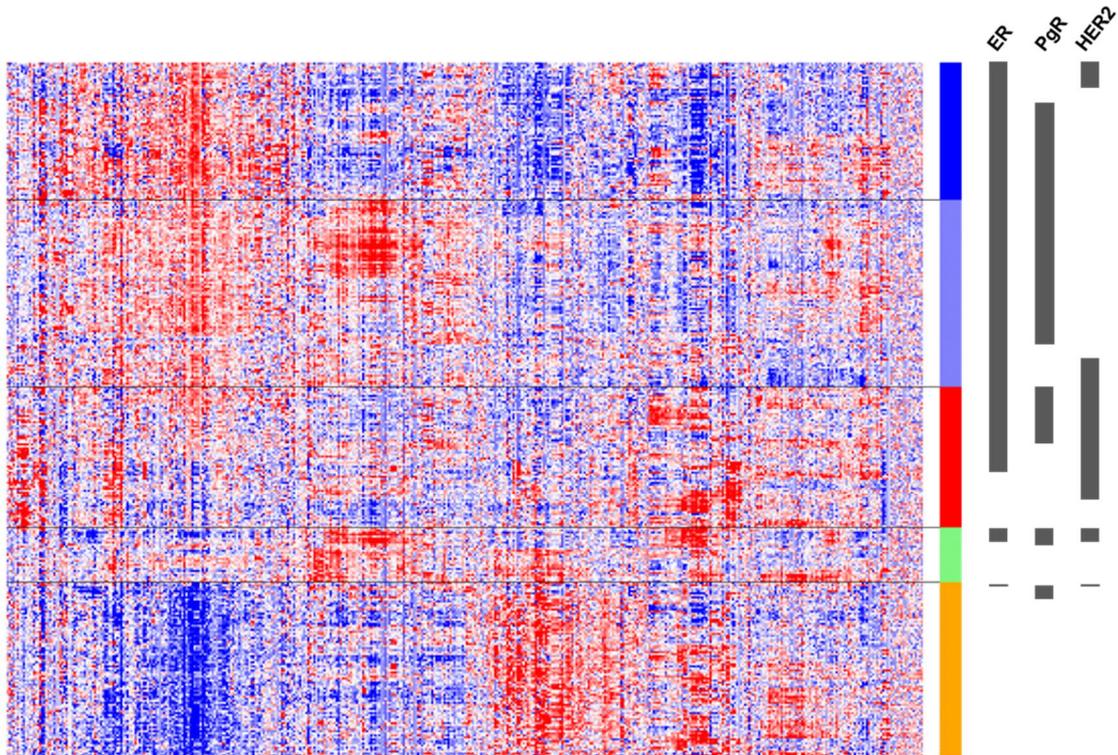
## **1.0 - Introduction**

### *1.1 – Breast cancer classification.*

Breast cancer is the most common malignancy among women [1]. At all tumor stages, treatment of breast cancer is determined based on the status of several markers, first of all the transcription factors estrogen receptor  $\alpha$  (ER, ESR1 gene) and progesterone receptor (PgR, PGR gene), and the receptor tyrosine kinase human epidermal growth factor receptor-2 (HER2, ERBB2 gene). Still, many more pathological and molecular features play a role in breast cancer development, metastasis formation and therapy resistance.

#### *1.1.1 – Intrinsic breast cancer subtypes*

By using gene expression array analysis several major molecular subtypes of breast cancer, defined as intrinsic subtypes, were identified (Figure 1.1) [2]. ER positive tumors are subdivided into luminal A and B subtypes, which are characterized by the expression of estrogen regulated genes, such as GATA3, NAT1 or XBP1, and by a cell morphology similar to that of the mammary gland luminal epithelium (i.e. ducts and acini). In particular, luminal B tumors display higher expression of cell cycle genes and lower expression of luminal genes (e.g. PGR) [3]. The ER negative subgroup is further divided into tumors that show HER2 overexpression and those that are basal or normal breast-like. While the HER2 positive subgroup is associated with expression of genes such as GATA4 and GRB7, distinction between the basal and normal breast-like subtypes relates to the overexpression of genes such as CDH3 and CXCL1 (specific to the basal subtype), and PIK3R1, AKR1C1 and FACL2 (specific to the normal breast-like subtype) [4]. In addition, while basal cancer morphology resembles the basal cell layer of the mammary gland, the one of the normal breast-like subtype is more similar to healthy breast epithelium.



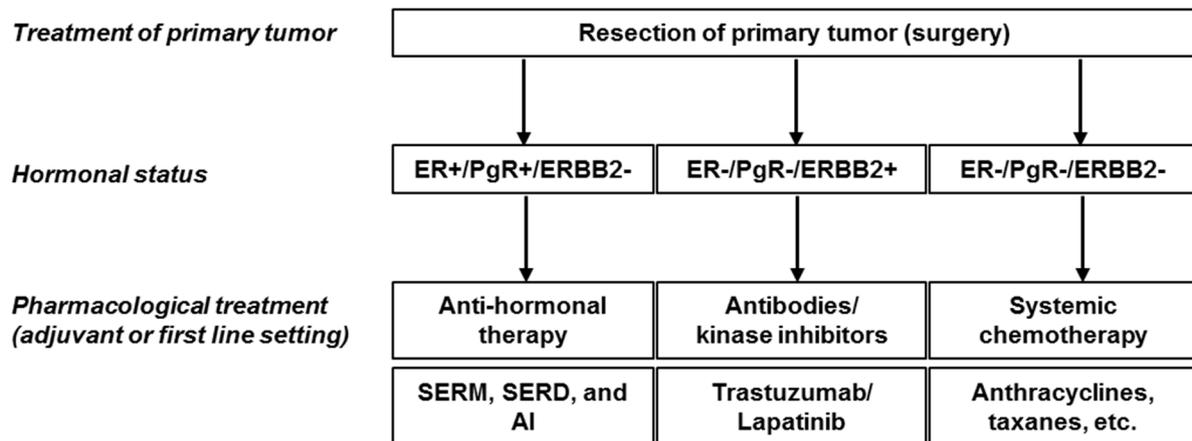
**Figure 1.1. Hierarchical clustering defining the intrinsic subtypes of breast cancer.**

Colors are indicative of breast cancer intrinsic subtypes: luminal B (dark blue), luminal A (light blue), HER2 (red), normal-like (green), and basal (orange). Hormonal (ER and PgR) and HER2 receptor statuses are displayed in grey. Modified from [5].

### *1.1.2 – Hormonal and HER2 status and treatment options*

Nearly three quarters of all breast cancer cases display ER positivity, and constitute a class of relatively less aggressive tumors which generally proliferate due to ER signaling [6,7]. In ER positive breast cancers, PgR is often co-expressed with ER, though some (20-25%) ER positive tumors do not display PgR positivity. A small percentage of tumors (i.e. 3-5%) displays ER negativity and PgR positivity. The HER2 gene is found amplified in nearly 20% of all breast cancer cases and its protein HER2 can be concomitantly expressed along with ER and PgR. For both ER positive and HER2 positive breast cancer patients, treatment options include targeted therapies. Anti-estrogens (selective estrogen receptor modulators [SERM], selective estrogen receptor degraders [SERD]) and aromatase inhibitors (AI), that respectively block ER signaling and reduce systemic estrogen levels, are used to treat ER positive malignancies, the choice of which largely depends on the menopausal status of the patient. On the other hand, trastuzumab (monoclonal antibody) and lapatinib (tyrosine kinase inhibitor) are given to ERBB2 amplified breast cancer patients [8–10]. Around 10-15% of breast cancers does not express ER, PgR or HER2 and are

therefore being classified as triple-negative breast cancer [11]. This subgroup of patients harbors highly aggressive malignancies, with a higher relapse rate than non-triple-negative breast cancers [12]. Furthermore, in the clinic no targeted systemic treatment is currently available for these tumors, which are then treated with standard chemotherapy (Figure 1.2) [13].



**Figure 1.2. Pharmacological treatment options according to breast cancer receptor status.**

In accordance with breast cancers having different hormonal receptor and HER2 status, tumors of various intrinsic subtypes differ not only in gene expression patterns but, as a consequence, also display significant differences in risk of developing metastases, with basal, HER2 positive and luminal B tumors showing a significantly shorter relapse-free-survival (RFS) when compared to normal-like and luminal A subtypes [14,15].

## 1.2 – ER positive breast cancer

### 1.2.1 – Mechanism of ER-signaling

Estrogens are a class of steroidal hormones that regulate cell growth and differentiation of tissues (e.g. mammary gland, ovary, bone, and uterus) and are present in the human body as estrone, estriol and 17 $\beta$ -estradiol, the latter being the most abundant. These compounds bind the ER, promoting its dimerization and a conformational change in the ligand binding domain (LBD) of the receptor that allows the recruitment of transcriptional co-activators for downstream gene expression of target genes containing an ER binding site (i.e. estrogen responsive element) [16]. ER can also modulate the expression of target genes without direct binding, but by regulating transcription factors (e.g.

AP-1 [17]) through protein-protein interactions [18]. The ligand-dependent activation of ER relies on the activation of C terminal activation function (AF) 2, located in the LBD of the receptor. Conversely, ligand-independent activation relies on AF-1, and is activated through ER phosphorylation (described in more detail below). Subsequently to its activation (irrespective of its modality) ER associates to chromatin in concert with pioneer factors such as GATA3 and FOXA1 to activate transcription [19]. Important ER-regulated genes include PGR [20,21], and TFF1 [22,23], among others.

### *1.2.2 – Major endocrine therapies*

The first anti-hormonal drug to be widely introduced in the clinic was tamoxifen, a SERM that acts as an antagonist to estrogens for binding to ER by recruiting transcriptional co-repressors (e.g. NCoR) instead of co-activators [6,24], leading ultimately to tumor growth inhibition. Tamoxifen treatment in the adjuvant setting (i.e. immediately after surgical treatment of the primary tumor and radiotherapy) reduces disease recurrence and breast cancer mortality by 39% and 31%, respectively [25–27]. Furthermore, although estrogen levels vary largely between pre- and post-menopausal women, early studies showed similar survival benefits in both groups [6,25]. In addition to its action in ER positive cases, a meta-analysis study showed that a small group (i.e. 5-10%) of ER negative tumors also responded to tamoxifen treatment, though long-term survival benefits were not observed [25]. Additional studies largely confirmed that only patients with ER positive disease benefitted from tamoxifen and other endocrine agents [27]. From a clinical perspective, tamoxifen is a well-tolerated drug because its side effects are generally mild. However, long-term use of tamoxifen has been associated to an increased risk of endometrial cancer [10,28,29].

Another competitive binder of ER is fulvestrant, a SERD and pure ER antagonist that prevents its dimerization and facilitates its proteasomal degradation [30]. Due to this mechanism of action, fulvestrant treatment efficacy is unlikely to suffer from cross-resistance with other anti-estrogenic treatments, nor has it been shown to increase endometrial cancer risk [31], though further clinical confirmation in large patient cohorts still needs to be provided.

AIs (e.g. letrozole, anastrozole, exemestane) constitute a class of drugs that inhibit estrogen signaling through a different mechanism as compared to tamoxifen or fulvestrant. Inhibition of the CYP450 family enzyme aromatase leads to systemic downregulation of estrogen levels due to the blockade of testosterone conversion into estrogens. In recent years, AI-based therapy has become one of the mainstays of ER positive breast cancer treatment in post-menopausal women, in whom it

has shown increased efficacy (30% recurrence rate reduction) when compared to tamoxifen [27,32,33]. Study reports have further shown that AIs are either equivalent or superior to tamoxifen as first-line treatment for recurrent breast cancer in post-menopausal women [34].

### **1.3 – General mechanisms of endocrine therapy resistance.**

Tamoxifen and AIs currently constitute the two most common endocrine treatments of ER positive breast cancer. However, their effectiveness is severely reduced by the tumor's intrinsic or acquired resistance (Figure 1.3). Although resistance occurs during all stages of the disease, it is especially a clinical problem in the recurrent setting, where nearly half of the patients already manifests resistance upon the start of treatment while the remainder develops it during therapy [35]. Several mechanisms (discussed in more detail below) have been associated with endocrine therapy resistance, such as ER gene (ESR1) mutations, epigenetic aberrations or signaling cross-talk.

#### *1.3.1 – Estrogen receptor mutations*

Somatic mutations in the ESR1 gene such as A1587G, which leads to Tyr537Ser amino acid modification in the LBD of the receptor, have been shown to have a direct adverse impact on patient survival in the recurrent setting [36–39]. These mutations generally are observed in 10-30% of all endocrine resistant recurrent breast cancers and have been linked to enhanced sensitivity to estradiol as well as to constitutive activation of transcriptional activity of ER in absence of ER agonists [36,37,40,41]. Strikingly, these mutations seem to present themselves only after exposure to one or more lines of endocrine treatment (in particular AIs) in the recurrent setting [36,42], as concluded from paired analyses of primary tumors and their metastatic therapy resistant counterparts [42,43]. Taken together, these observations support the idea that ESR1 gene mutations, especially the ones in the LBD, prevalently arise due to endocrine therapy selection of resistant clones. Functional characterization as well as therapeutic targeting of these mutants is just in its infancy. However, due to their predictive value, monitoring of these mutations through the course of endocrine therapy – namely in metastatic lesions, circulating tumor cells, or cell-free DNA (cfDNA) [38,42–44] - may help clinicians to identify and monitor resistant patients.

### *1.3.2 –Dysregulation of gene expression and cross-talk mechanisms*

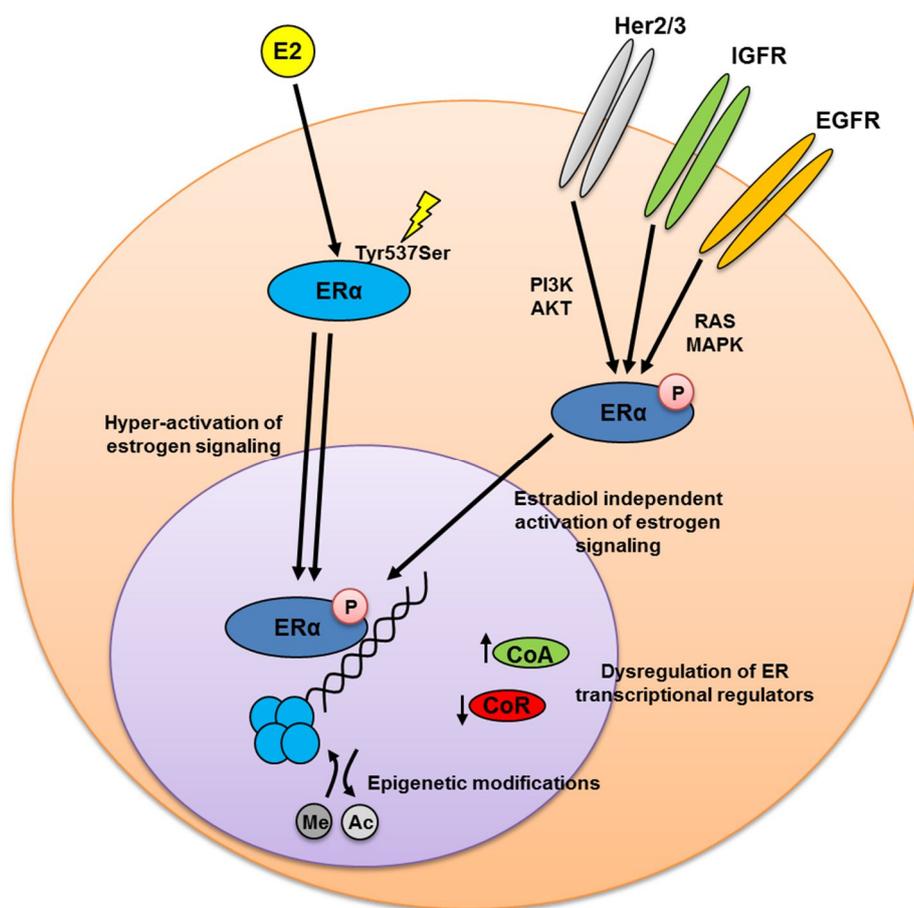
A plethora of mechanisms can influence ER-mediated gene expression, such as changes in expression or occurrence of mutations in ER transcriptional modulators. Numerous studies have pointed out the association of the nuclear receptor co-activator AIB1 to tamoxifen therapy outcome, though it remains unclear whether it promotes therapy benefit or resistance [45–48]. Other co-activators that have been linked to tamoxifen resistance are RIP140, which regulates genes that have been linked to tamoxifen resistance [49], and SRC-1, which acts as a limiting co-factor for the transactivation of ESR1 and PGR gene expression [50]. Conversely the downregulation of ER transcriptional co-repressors, such as NCoR1, have also been shown to contribute to resistance to tamoxifen therapy [51,52]. These molecular alterations have been associated with promotion of cell cycle progression despite tamoxifen treatment, and as a consequence they possibly constitute alternative therapeutic targets [51]. In addition to this, transcription factors (e.g. GATA3) and transcriptional complex stabilizing factors (e.g. GREB1) have been shown to contribute to expression dysregulation of ER-related genes leading to endocrine resistance [19,53–56].

Recent studies have pointed out that epigenetic changes, in particular DNA hypo-/hypermethylation at CpG islands and histone modifications (e.g. methylation at lysine residues) [57–60], can also affect outcome to endocrine therapy. Alterations in the DNA methylation pattern has been previously associated to resistance to tamoxifen therapy, such as the hypomethylation at the promoter region of PSAT1, a gene coding for an enzyme involved in serine biosynthesis [61]. Furthermore, endocrine treatment resistance has been associated with epigenetic gene silencing such as in the case of HOXC10, a homeobox gene involved in apoptosis and cell growth inhibition [62], or the recruitment of histone-related proteins such as EZH2, a polycomb protein with histone methyltransferase properties [62,63], which has been found upregulated in many breast cancers [64–66].

Several kinases of the MAPK family, such as ERK1 and ERK3, have been shown to phosphorylate ER (e.g. at Ser-118), prompting ligand-independent activation of the receptor, and altering response to ER agonists and antagonists [67–69]. Furthermore, the expression of HER2, EGFR or IGFR can ultimately induce phosphorylation of ER and AIB1 through cross-talk mechanisms, which have been shown to empower estrogen signaling and induce resistance to tamoxifen [70–73]. Phosphoinositide 3 kinase (PI3K) and protein kinase B (PKB; also known as Akt) also play a role in activation of ER-related transcription: these kinases activate the receptor in the absence of estrogens through phosphorylation of the AF-1 (PI3K) and AF-2 (PI3K and PKB) domains of the receptor

[74]. Furthermore, activating phosphorylations of the ER can be enacted by other kinases, such as c-Jun terminal kinase or p38 [75–77].

Signaling pathway cross-talk mechanisms have also been associated to the expression of long non-coding (lnc) RNAs, such as BCAR4 (co-expressed with ERBB2) [78–80], which have been shown to promote resistance to endocrine therapies. Taken together, these molecules may not only be used as biomarkers of outcome to endocrine therapy, but could also constitute alternative therapeutic targets in endocrine therapy resistant patients.



**Figure 1.3. Schematic representation of major endocrine resistance mechanisms.**

## 1.4 – Prediction of tamoxifen therapy resistance

### 1.4.1 – Traditional clinical predictive factors

Several clinical and molecular markers have been associated with response to tamoxifen treatment, both in the adjuvant and in the recurrent setting. Although such information is not always easily

defined in the clinic, it has been established that post-menopausal women do benefit less from tamoxifen therapy in the adjuvant setting and their regimen is therefore switched to AI [81].

Among the molecular markers, ER and PgR levels are so far the best clinical markers predictive of tamoxifen outcome [82,83]. In addition to this, mitotic index and Ki-67 levels are also in use and have been associated to poor prognosis [3,84–86].

#### *1.4.2 – Biomarkers of tamoxifen resistance*

With the introduction of gene expression analysis (e.g. hybridization chips) and new techniques to study gene expression in vitro (e.g. RNA silencing), new methods of biomarker discovery and investigation have become available. While classifiers predictive of recurrence for ER positive breast cancer (e.g. Mammaprint® [87]) found gradual introduction into clinical practice, tamoxifen resistance predictive signatures still needed development [88,89]. Subsequently, either single genes (e.g. BCAR4) or entire gene lists (e.g. 44- and 76-gene signatures) were used as classifiers of tamoxifen resistance in the adjuvant and/or recurrent settings [64,79,80,90,91]. Furthermore, with the advent of epigenomics and the dissection of DNA transcription mechanisms, new perspectives have been added to the investigation of tamoxifen resistance. Signatures developed by analysis of ER and histone proteins binding sites (e.g. H3K4me3) not only significantly predicted patient outcome in AI and tamoxifen treated patients, but outperformed previous gene expression classifiers [49,92].

### **1.5 – The proteomic approach**

#### *1.5.1 – Mass spectrometry: a cutting-edge technique for the study of proteins*

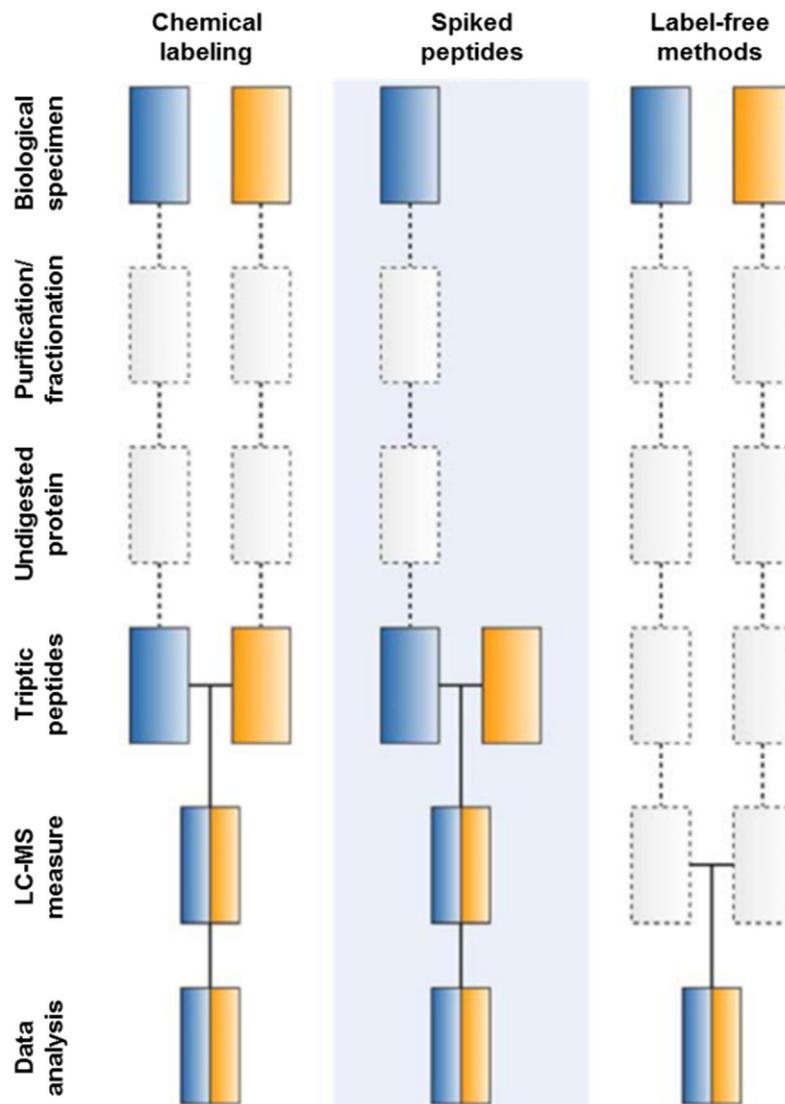
With the improvement of chromatographic and ion separation techniques, mass spectrometry (MS) has become one of the most advanced techniques to address biological and clinical issues, being able to identify and quantify > 10,000 protein species in a single biological specimen, rendering this technique a robust additional tool complementary to gene expression analysis [93–95]. Furthermore, the elucidation of the human proteome showed the full capabilities of the proteomic approach in analyzing the wide dynamic range of protein expression present throughout different tissues [96,97]. In addition, with the possibility to detect post-translational modifications (PTMs), a second functional layer of information can be provided [98,99].

### *1.5.2 – Common quantitation methods in proteomics*

In MS data analysis, protein identification is the first step, which can be followed by quantitation. The latter can be achieved through a plethora of methods (Figure 1.4). Chemical labeling techniques (e.g. isotope tags for relative and absolute quantitation [iTRAQ], tandem mass tags [TMT]) constitute one of the most common approaches for proteomic analyses of large sample sets, offer the possibility to accurately quantify proteins in a robust and reproducible way [100,101]. However, these techniques require relatively high amounts of sample material, and pre-labeling protein enrichment is impractical when working with small amounts of sample material. Furthermore, peptide quantitation through these methods may suffer from interference derived from co-eluting peptides of a similar mass, though this can be minimized by employment of fractionation techniques. In this perspective, an iTRAQ or TMT experiment would rely on several sample preparation steps, which would be time-consuming and susceptible to variation in protein quantitation due to the extensive sample manipulation [102]. In this perspective, algorithms that enable quantification without previous labeling are becoming attractive alternatives. Label-free quantification (LFQ), which is provided within the MaxQuant environment for example, does not rely on sample pre-processing, and protein levels are computationally derived from the intensity of their peptide-spectrum matches (PSMs), which are compared between samples. Another approach consists of counting fragmentation spectra belonging to a protein-specific peptide, the so-called spectral count [103]. While the first approach uses peptide extracted ion chromatograms and integrates them through the chromatographic run (in the time scale), the second counts peptide fragmentation spectra, which are based on MS2 scans, and compares them for every peptide to achieve relative quantitation [102]. Despite the fact that label-free methods provide less accurate quantification compared to chemical labeling methods, the number of experiments that can be compared is virtually unlimited, rendering such approach well suited for the analysis of large sample cohorts with two or more experimental conditions. Furthermore, LFQ has proven to identify a relatively higher amount of proteins when compared to labeling strategies, thus potentially providing a higher analytical depth. Recent studies have demonstrated that combination of tissue enrichment strategies with sample fractionation and LFQ allows the quantification of thousands of proteins, especially in the low abundance range [102,104–107]. Furthermore, with the continuous improvements in LFQ softwares (e.g. MaxQuant), protein quantification has become progressively more accurate [108–111].

While these methods are generally used in global proteomic studies, other quantitation strategies are used to determine highly accurate abundances of target peptides/proteins [102,112]. These targeted

approaches are generally more suited for biomarker verification and clinical assay development strategies, due to the fact they provide precise quantitation of (theoretically) hundreds of proteotypic peptides in a sensitive and reproducible way. Protein quantitation methods for these targeted assays, such as selected reaction monitoring (SRM), comprise label-free and chemical labeling approaches, as well as isotopically labeled standard spikes. Since isotopically labeled versions of target analytes are used as internal standards, they retain the same physical and chemical properties of their endogenous counterparts, minimizing interference derived from various sources (e.g. co-eluting species) [113]. Furthermore, when compared to immuno-assays (e.g. enzyme-linked immunosorbent assay [ELISA]), SRM MS provides comparable selectivity and lower development costs [114,115].



**Figure 1.4. Most common protein quantitation methods used in MS-based proteomics.**

Blue and yellow boxes represent different experimental conditions, while dashed lines indicate steps at which experimental variation can occur. Modified from [102].

### *1.5.3 – Proteomic approaches for biomarkers*

Proteomic approaches for the development and validation of clinically useful biomarkers are diverse, though they can be summarized into two main categories: global (or shotgun) and targeted proteomic analyses. While in the former approach the whole (or a subset of the) dynamic range of proteins expressed in a biological sample is measured, the latter employs previous biological knowledge to specifically identify and accurately quantify a putative marker or a subset of proteins (e.g. N-glycosylated proteins). In the biomarker discovery phase, statistical methods are applied to define significant differences in (modified and/or unmodified) protein levels between two or more experimental or biological/clinical conditions (t-test, ANOVA, etc.) [116]. These differentially expressed proteins constitute a list of putative biomarkers from which a predictor can be developed, which are then verified in an independent set of samples (i.e. validation set). Both global and targeted approaches are used in biomarker discovery studies, though a general consensus remains: while global proteomic analysis is generally used to identify putative biomarkers, ELISA assays or targeted MS approaches are then used to provide accurate and quantitative measurements, which provide biological and technical validations as well as a more clinically feasible assay [105,111,117,118].

### *1.5.4 – Phosphoproteomics*

The most common post-translational modification, dysregulation of protein phosphorylation, has been reported as one of the critical factors associated to cancer development, metastasis formation, and therapy resistance (e.g. ERBB2, EGFR) [119,120]. On the technical side, the study of phosphorylated proteins has so extensively evolved through the development of multiplexable techniques (e.g. reverse-phase protein arrays, peptide arrays, MS), that assessment of thousands of phosphorylation events is nowadays standard practice in many laboratories [121]. On the biological side, not only the analysis of phosphorylated proteins enables the identification of dysregulated (e.g. hyper-activated, mutated) signaling pathways in diseases such as cancer (e.g. effects of BRAF mutations), but also provide clarifying information on drug on- and off-target effects, as well as new pharmacologically targetable biomarkers (e.g. kinase inhibitors) [122–125]. Measurement of post-translational modification, though, come at the cost of increasing quantitation variation due to the additional purification step required (i.e. increased sample manipulation). Furthermore, while modified peptides are less ionizable compared to their unmodified counterparts, modifications are also more labile, impacting overall sensitivity of the MS measurement [126]. Alternative ionization

techniques (e.g. combined electron transfer and collision induced dissociation EThcD) have been developed to solve this issue in the field of phosphoproteomics [127], though the laborious sample preparation impacts negatively on the analysis of large cohorts of clinical specimens.

#### *1.5.5 - Glycoproteomics*

While phosphorylation changes elucidate the status of kinases in a given biological condition, glycosylation patterns provide additional information on cell-to-cell adhesion or protein folding [128,129]. Furthermore, glycosylated protein function may also change depending on the nature or the position of the attached group. Elucidation of the function of glycosylated proteins in complex diseases has pointed out that alteration of glycosylation pathways are hallmarks of cancer progression and metastasis [130,131]. Furthermore, glycosylated proteins are also secreted in the bloodstream and may constitute viable disease biomarkers for diagnostic purposes. Despite this, extensive glyco-proteomic analyses remain challenging due to extensive sample preparation, lability of modifications, and difficulties in glycosylated peptide spectra annotations [117,132].

#### *1.5.6- Protein markers of tamoxifen resistance*

In recent years, several studies reported global MS-derived biomarkers for tamoxifen resistance in the adjuvant setting. Proteomic studies in both ER positive breast cancer cell lines and patient-derived tumors showed that senescence inducing (i.e. RARA) and proliferation-related (i.e. CAPS) proteins were associated to short RFS and poor outcome to adjuvant tamoxifen therapy [133,134]. Other studies investigated changes in serum protein levels between tumors either responsive or resistant to tamoxifen treatment (e.g. Apo-lipoprotein E) [135].

Only one study investigated tamoxifen therapy resistance in the recurrent setting through MS, in which a panel of 100 proteins was derived from the analysis of 51 ER positive breast cancers. Despite these initial findings, only CD147 (or EMMPRIN; i.e. the most significant protein) was validated in an independent sample set through immunohistochemistry (IHC) [136].

Taken together these studies paved the way for clinical breast cancer proteomics. However, each of them suffered from either relatively small sample numbers or tumor heterogeneity unrepresentative (i.e. cell lines) discovery sets. Furthermore, most studies were small sized or lacked of validation

cohorts, necessitating large independent verification sets prior to biomarker introduction into a clinical setting.

#### *1.5.7 – Tissue proteomic workflow for biomarker discovery*

In the search for prognostic and predictive cancer biomarkers, both genomic and proteomic analyses are hampered by the presence of multiple cell types in primary tissues, which can alter accurate quantitation due to signals derived from tumor-surrounding tissues (e.g. stromal cells, adipocytes, leucocytes) [137–139]. To overcome this issue, cell population enrichment techniques, such as LCM, have proven to be powerful tools in elucidating tumor biology and approaches to biomarker discovery [140–142]. Despite the high purity of sample material derived from cell enrichment procedures, only minute amounts of tissue can be derived in a time- and cost-effective manner, which can limit the number of protein identifications on the MS level. Improvements in liquid chromatography and with the next generation of high resolution MS (i.e. Orbitrap series), the amount of identified and quantified proteins from LCM material was highly improved [104].

## References:

- [1] Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2015. *CA Cancer J Clin* 2015;65:5–29. doi:10.3322/caac.21254.
- [2] Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees C a, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52. doi:10.1038/35021093.
- [3] Esposito A, Criscitiello C, Curigliano G. Highlights from the 14(th) St Gallen International Breast Cancer Conference 2015 in Vienna: Dealing with classification, prognostication, and prediction refinement to personalize the treatment of patients with early breast cancer. *Ecancermedicalsecience* 2015;9:518. doi:10.3332/ecancer.2015.518.
- [4] Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100:8418–23. doi:10.1073/pnas.0932692100.
- [5] Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JGM, et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 2008;68:3108–14. doi:10.1158/0008-5472.CAN-07-5644.
- [6] Osborne CK. Tamoxifen in the treatment of breast cancer. *NEJM* 1998;339:1609–18.
- [7] Lumachi F, Brunello a, Maruzzo M, Basso U, Basso SMM. Treatment of estrogen receptor-positive breast cancer. *Curr Med Chem* 2013;20:596–604. doi:10.2174/092986713804999303.
- [8] Burki TK. Adjuvant treatment for HER2-positive breast cancer. *Lancet Oncol* 2015;16:e59. doi:10.1016/S1470-2045(15)70019-7.
- [9] Larsen PB, Kümler I, Nielsen DL. A systematic review of trastuzumab and lapatinib in the treatment of women with brain metastases from HER2-positive breast cancer. *Cancer Treat Rev* 2013;39:720–7. doi:10.1016/j.ctrv.2013.01.006.
- [10] Schiavon G, Smith IE. Endocrine therapy for advanced/metastatic breast cancer. *Hematol Oncol Clin North Am* 2013;27:715–36, viii. doi:10.1016/j.hoc.2013.05.004.
- [11] Foulkes W, Smith I, Reis-Filho JS. Triple-Negative Breast Cancer. *NEJM* 2010;363:1938–48.
- [12] Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka C a., et al. Triple-negative breast cancer: Clinical features and patterns of recurrence. *Clin Cancer Res* 2007;13:4429–34. doi:10.1158/1078-0432.CCR-06-3045.
- [13] Silver DP, Richardson AL, Eklund AC, Wang ZC, Szallasi Z, Li Q, et al. Efficacy of neoadjuvant cisplatin in triple-negative breast cancer. *J Clin Oncol* 2010;28:1145–53. doi:10.1200/JCO.2009.22.4725.
- [14] Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Mol Oncol* 2011;5:5–23. doi:10.1016/j.molonc.2010.11.003.
- [15] Creighton CJ. The molecular profile of luminal B breast cancer. *Biol Targets Ther* 2012;6:289–97. doi:10.2147/BTT.S29923.

- [16] Rosenfeld MG, Glass CK. Coregulator codes of transcriptional regulation by nuclear receptors. *J Biol Chem* 2001;276:36865–8. doi:10.1074/jbc.R100041200.
- [17] Uht RM, Webb P, Nguyen P, Price Jr Jr. RH, Valentine C, Favre H, et al. A conserved lysine in the estrogen receptor DNA binding domain regulates ligand activation profiles at AP-1 sites, possibly by controlling interactions with a modulating repressor. *Nucl Recept* 2004;2:2. doi:10.1186/1478-1336-2-2 [pii].
- [18] Nilsson S, Ma S, Treuter E, Tujague M, Thomsen J. Mechanisms of Estrogen Action. *Physiol Rev* 2001;81:1535–65.
- [19] Theodorou V, Stark R, Menon S, Carroll JS. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res* 2013;23:12–22. doi:10.1101/gr.139469.112.
- [20] Ylikomi T, Bocquel MT, Berry M, Gronemeyer H, Chambon P. Cooperation of proto-signals for nuclear accumulation of estrogen and progesterone receptors. *EMBO J* 1992;11:3681–94.
- [21] McGuire WL, Horwitz KB, Pearson OH, Segaloff A. Current status of estrogen and progesterone receptors in breast cancer. *Cancer* 1977;39:2934–47.
- [22] Spyrtos F, Andrieu C, Hacène K, Chambon P, Rio MC. pS2 and response to adjuvant hormone therapy in primary breast cancer. *Br J Cancer* 1994;69:394–7.
- [23] Jeltsch JM, Roberts M, Schatz C, Garnier JM, Brown AM, Chambon P. Structure of the human oestrogen-responsive gene pS2. *Nucleic Acids Res* 1987;15:1401–14.
- [24] Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M. Cofactor Dynamics and Sufficiency in Estrogen Receptor–Regulated Transcription. *Cell* 2000;103:843–52. doi:10.1016/S0092-8674(00)00188-4.
- [25] (EBCTCG) EBCTCG. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;351:1451–67. doi:10.1016/S0140-6736(97)11423-4.
- [26] (EBCTCG) EBCTCG. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;365:1687–717. doi:10.1016/S0140-6736(05)66544-0.
- [27] (EBCTCG) EBCTCG. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet* 2015;303:1341–52. doi:10.1016/S0140-6736(15)61074-1.
- [28] Jordan VC. Tamoxifen: toxicities and drug resistance during the treatment and prevention of breast cancer. *Annu Rev Pharmacol Toxicol* 1995;35:195–211. doi:10.1146/annurev.pa.35.040195.001211.
- [29] Neven P. Tamoxifen and endometrial lesions. *Lancet* 1993;342:452. doi:10.1016/0140-6736(93)91590-I.
- [30] Wardell SE, Marks JR, McDonnell DP. The turnover of estrogen receptor  $\alpha$  by the selective estrogen receptor degrader (SERD) fulvestrant is a saturable process that is not required for antagonist efficacy. *Biochem Pharmacol* 2011;82:122–30. doi:10.1016/j.bcp.2011.03.031.

- [31] Valachis A, Mauri D, Polyzos NP, Mavroudis D, Georgoulas V, Casazza G. Fulvestrant in the treatment of advanced breast cancer: A systematic review and meta-analysis of randomized controlled trials. *Crit Rev Oncol Hematol* 2010;73:220–7. doi:10.1016/j.critrevonc.2009.03.006.
- [32] Bonnetterre J, Thurlimann B, Robertson JFR, Krzakowski M, Mauriac L, Koralewski P, et al. Anastrozole Versus Tamoxifen as First-Line Therapy for Advanced Breast Cancer in 668 Postmenopausal Women : Results of the Tamoxifen or Arimidex Randomized Group Efficacy and Tolerability Study. *J Clin Oncol* 2000;18:3748–57.
- [33] Bonnetterre J, Buzdar A, Nabholz JM, Robertson JF, Thürlimann B, von Euler M, et al. Anastrozole is superior to tamoxifen as first-line therapy in hormone receptor positive advanced breast carcinoma. *Cancer* 2001;92:2247–58. doi:10.1002/1097-0142(20011101)92:9<2247::AID-CNCR1570>3.0.CO;2-Y [pii].
- [34] Cardoso F, Harbeck N, Fallowfield L, Kyriakides S, Senkus E. Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2012;23 Suppl 7:vii11–9. doi:10.1093/annonc/mds232.
- [35] Clarke R, Tyson JJ, Dixon JM. Endocrine resistance in breast cancer - An overview and update. *Mol Cell Endocrinol* 2015. doi:10.1016/j.mce.2015.09.035.
- [36] Toy W, Shen Y, Won H, Green B, Sakr R a, Will M, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* 2013;45:1439–45. doi:10.1038/ng.2822.
- [37] Robinson DR, Wu Y-M, Vats P, Su F, Lonigro RJ, Cao X, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* 2013;45:1446–51. doi:10.1038/ng.2823.
- [38] Segal C V, Dowsett M. Estrogen receptor mutations in breast cancer - new focus on an old target. *Clin Cancer Res* 2014;25:1724–6. doi:10.1158/1078-0432.CCR-14-0067.
- [39] Alluri PG, Speers C, Chinnaiyan AM. Estrogen receptor mutations and their role in breast cancer progression. *Breast Cancer Res* 2014;16:494. doi:10.1186/s13058-014-0494-7.
- [40] Li S, Shen D, Shao J, Crowder R, Liu W, Prat A, et al. Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep* 2013;4:1116–30. doi:10.1016/j.celrep.2013.08.022.
- [41] Jeselsohn R, Yelensky R, Buchwalter G, Frampton G, Meric-Bernstam F, Gonzalez-Angulo AM, et al. Emergence of constitutively active estrogen receptor- $\alpha$  mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res* 2014;20:1757–67. doi:10.1158/1078-0432.CCR-13-2332.
- [42] Schiavon G, Hrebien S, Garcia-murillas I, Cutts RJ, Pearson A, Tarazona N, et al. Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* 2015;7:1–9. doi:10.1126/scitranslmed.aac7551.
- [43] Wang P, Bahreini A, Gyanchandani R, Lucas P, Hartmaier RJ, Watters RJ, et al. Sensitive detection of mono- and polyclonal ESR1 mutations in primary tumors, metastatic lesions and cell free DNA of breast cancer patients. *Clin Cancer Res* 2015. doi:10.1158/1078-0432.CCR-15-1534.

- [44] Butler TM, Johnson-Camacho K, Peto M, Wang NJ, Macey TA, Korkola JE, et al. Exome sequencing of cell-free DNA from metastatic cancer patients identifies clinically actionable mutations distinct from primary disease. *PLoS One* 2015;10:1–14. doi:10.1371/journal.pone.0136407.
- [45] Burandt E, Jens G, Holst F, Jänicke F, Müller V, Quaas A, et al. Prognostic relevance of AIB1 (NCoA3) amplification and overexpression in breast cancer. *Breast Cancer Res Treat* 2013;137:745–53. doi:10.1007/s10549-013-2406-4.
- [46] Weiner M, Skoog L, Fornander T, Nordenskjöld B, Sgroi DC, Stål O. Oestrogen receptor co-activator AIB1 is a marker of tamoxifen benefit in postmenopausal breast cancer. *Ann Oncol* 2013;24:1994–9. doi:10.1093/annonc/mdt159.
- [47] Alkner S, Bendahl P-O, Grabau D, Lövgren K, Stål O, Rydén L, et al. AIB1 is a predictive factor for tamoxifen response in premenopausal women. *Ann Oncol* 2010;21:238–44. doi:10.1093/annonc/mdp293.
- [48] Osborne CK, Bardou V, Hopp T a, Chamness GC, Hilsenbeck SG, Fuqua S a W, et al. Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J Natl Cancer Inst* 2003;95:353–61. doi:10.1093/jnci/95.5.353.
- [49] Rosell M, Nevedomskaya E, Stelloo S, Nautiyal J, Poliandri A, Steel JH, et al. Complex formation and function of estrogen receptor alpha in transcription requires RIP140. *Cancer Res* 2014;74:5469–79. doi:10.1158/0008-5472.CAN-13-3429.
- [50] Berns EMJJ, Van Staveren IL, Klijn JGM, Foekens JA. Predictive value of SRC-1 for tamoxifen response of recurrent breast cancer. *Breast Cancer Res Treat* 1998;48:87–92. doi:10.1023/A:1005903226483.
- [51] Keeton EK, Brown M. Cell cycle progression stimulated by tamoxifen-bound estrogen receptor-alpha and promoter-specific effects in breast cancer cells deficient in N-CoR and SMRT. *Mol Endocrinol* 2005;19:1543–54. doi:10.1210/me.2004-0395.
- [52] Zhang X, Jeyakumar M, Petukhov S, Bagchi MK. A nuclear receptor corepressor modulates transcriptional activity of antagonist-occupied steroid hormone receptor. *Mol Endocrinol* 1998;12:513–24. doi:10.1210/mend.12.4.0089.
- [53] Wilson BJ, Giguere V. Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. *Mol Cancer* 2008;7:49. doi:10.1186/1476-4598-7-49.
- [54] Liu H, Shi J, Wilkerson ML, Lin F. Immunohistochemical evaluation of GATA3 expression in tumors and normal tissues: A useful immunomarker for breast and urothelial carcinomas. *Am J Clin Pathol* 2012;138:57–64. doi:10.1309/AJCP5UAFMSA9ZQBZ.
- [55] Mohammed H, D’Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep* 2013;3:342–9. doi:10.1016/j.celrep.2013.01.010.

- [56] Nimmrich I, Sieuwerts AM, Meijer-Van Gelder ME, Schwöpe I, Bolt-De Vries J, Harbeck N, et al. DNA hypermethylation of PITX2 is a marker of poor prognosis in untreated lymph node-negative hormone receptor-positive breast cancer patients. *Breast Cancer Res Treat* 2008;111:429–37. doi:10.1007/s10549-007-9800-8.
- [57] Barrow TM, Michels KB. Epigenetic epidemiology of cancer. *Biochem Biophys Res Commun* 2014;455:70–83. doi:10.1016/j.bbrc.2014.08.002.
- [58] Rodríguez-Rodero S, Delgado-Álvarez E, Fernández AF, Fernández-Morera JL, Menéndez-Torre E, Fraga MF. Epigenetic alterations in endocrine-related cancer. *Endocr Relat Cancer* 2014;21:R319–30. doi:10.1530/ERC-13-0070.
- [59] Pathiraja TN, Stearns V, Oesterreich S. Epigenetic Regulation in Estrogen Receptor Positive Breast Cancer—Role in Treatment Response. *J Mammary Gland Biol Neoplasia* 2010;15:35–47. doi:10.1007/s10911-010-9166-0.
- [60] Harbeck N, Nimmrich I, Hartmann A, Ross JS, Cufér T, Grützmann R, et al. Multicenter study using paraffin-embedded tumor tissue testing PITX2 DNA methylation as a marker for outcome prediction in tamoxifen-treated, node-negative breast cancer patients. *J Clin Oncol* 2008;26:5036–42. doi:10.1200/JCO.2007.14.1697.
- [61] Martens JWM, Nimmrich I, Koenig T, Look MP, Harbeck N, Model F, et al. Association of DNA methylation of phosphoserine aminotransferase with response to endocrine therapy in patients with recurrent breast cancer. *Cancer Res* 2005;65:4101–7. doi:10.1158/0008-5472.CAN-05-0064.
- [62] Pathiraja TN, Nayak SR, Xi Y, Jiang S, Garee JP, Edwards DP, et al. Epigenetic reprogramming of HOXC10 in endocrine-resistant breast cancer. *Sci Transl Med* 2014;6:229ra41. doi:10.1126/scitranslmed.3008326.
- [63] Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. *Cell* 2007;128:735–45. doi:10.1016/j.cell.2007.02.009.
- [64] Jansen MPH, Foekens JA, van Staveren IL, Dirkzwager-Kiel MM, Ritstier K, Look MP, et al. Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol* 2005;23:732–40. doi:10.1200/JCO.2005.05.145.
- [65] Reijm EA, Jansen MPH, Ruigrok-Ritstier K, van Staveren IL, Look MP, van Gelder MEM, et al. Decreased expression of EZH2 is associated with upregulation of ER and favorable outcome to tamoxifen in advanced breast cancer. *Breast Cancer Res Treat* 2011;125:387–94. doi:10.1007/s10549-010-0836-9.
- [66] Reijm EA, Timmermans AM, Look MP, Meijer-van Gelder ME, Stobbe CK, van Deurzen CHM, et al. High protein expression of EZH2 is related to unfavorable outcome to tamoxifen in metastatic breast cancer. *Ann Oncol* 2014;25:2185–90. doi:10.1093/annonc/mdu391.
- [67] Ring A, Dowsett M. Mechanisms of tamoxifen resistance. *Endocr Relat Cancer* 2004;11:643–58. doi:10.1677/erc.1.00776.
- [68] Li Z, Wang N, Fang J, Huang J, Tian F, Li C, et al. Role of PKC-ERK signaling in tamoxifen-induced apoptosis and tamoxifen resistance in human breast cancer cells. *Oncol Rep* 2012;27:1879–86. doi:10.3892/or.2012.1728.

- [69] Heckler MM, Thakor H, Schafer CC, Riggins RB. ERK/MAPK regulates ERR $\gamma$  expression, transcriptional activity and receptor-mediated tamoxifen resistance in ER+ breast cancer. *FEBS J* 2014;281:2431–42. doi:10.1111/febs.12797.
- [70] Shou J, Massarweh S, Osborne CK, Wakeling AE, Ali S, Weiss H, et al. Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. *J Natl Cancer Inst* 2004;96:926–35. doi:10.1093/jnci/djh166.
- [71] Massarweh S, Osborne CK, Creighton CJ, Qin L, Tsimelzon A, Huang S, et al. Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. *Cancer Res* 2008;68:826–33. doi:10.1158/0008-5472.CAN-07-2707.
- [72] Osborne CK, Coronado-Heinsohn EB, Hilsenbeck SG, McCue BL, Wakeling a E, McClelland RA, et al. Comparison of the effects of a pure steroidal antiestrogen with those of tamoxifen in a model of human breast cancer. *J Natl Cancer Inst* 1995;87:746–50. doi:10.1093/jnci/87.10.746.
- [73] Gee JM, Robertson JF, Gutteridge E, Ellis IO, Pinder SE, Rubini M, et al. Epidermal growth factor receptor/HER2/insulin-like growth factor receptor signalling and oestrogen receptor activity in clinical breast cancer. *Endocr Relat Cancer* 2005;12 Suppl 1:S99–111. doi:10.1677/erc.1.01005.
- [74] Campbell RA, Bhat-Nakshatri P, Patel NM, Constantinidou D, Ali S, Nakshatri H. Phosphatidylinositol 3-kinase/AKT-mediated activation of estrogen receptor  $\alpha$ : A new model for anti-estrogen resistance. *J Biol Chem* 2001;276:9817–24. doi:10.1074/jbc.M010840200.
- [75] Feng W, Webb P, Nguyen P, Liu X, Li J, Karin M, et al. Potentiation of estrogen receptor activation function 1 (AF-1) by Src/JNK through a serine 118-independent pathway. *Mol Endocrinol* 2001;15:32–45. doi:10.1210/mend.15.1.0590.
- [76] Lee H, Bai W. Regulation of Estrogen Receptor Nuclear Export by Ligand-Induced and p38-Mediated Receptor Phosphorylation Regulation of Estrogen Receptor Nuclear Export by Ligand-Induced and p38-Mediated Receptor Phosphorylation. *Mol Cell Biol* 2002;22:5835–45. doi:10.1128/MCB.22.16.5835.
- [77] Osborne CK, Schiff R. Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer. *The Breast* 2003;12:362–7. doi:10.1016/S0960-9776(03)00137-1.
- [78] van Agthoven T, Sieuwerts AM, Meijer-van Gelder ME, Look MP, Smid M, Veldscholte J, et al. Relevance of breast cancer antiestrogen resistance genes in human breast cancer progression and tamoxifen resistance. *J Clin Oncol* 2009;27:542–9. doi:10.1200/JCO.2008.17.1462.
- [79] Godinho MFE, Sieuwerts AM, Look MP, Meijer D, Foekens JA, Dorsers LCJ, et al. Relevance of BCAR4 in tamoxifen resistance and tumour aggressiveness of human breast cancer. *Br J Cancer* 2010;103:1284–91. doi:10.1038/sj.bjc.6605884.
- [80] Godinho MFE, Wulfkühle JD, Look MP, Sieuwerts AM, Sleijfer S, Foekens JA, et al. BCAR4 induces antioestrogen resistance but sensitises breast cancer to lapatinib. *Br J Cancer* 2012;107:947–55. doi:10.1038/bjc.2012.351.

- [81] De Vos FYFL, van Laarhoven HWM, Laven JSE, Themmen APN, Beex LVAM, Sweep CGJ, et al. Menopausal status and adjuvant hormonal therapy for breast cancer patients: A practical guideline. *Crit Rev Oncol Hematol* 2012;84:252–60. doi:10.1016/j.critrevonc.2012.06.005.
- [82] Elledge RM, Green S, Pugh R, Allred DC, Clark GM, Hill J, et al. Estrogen receptor (ER) and progesterone receptor (PgR), by ligand-binding assay compared with ER, PgR and pS2, by immuno-histochemistry in predicting response to tamoxifen in metastatic breast cancer: a Southwest Oncology Group Study. *Int J Cancer* 2000;89:111–7. doi:10.1002/(SICI)1097-0215(20000320)89:2<111::AID-IJC2>3.0.CO;2-W [pii].
- [83] Allred DC. Issues and updates: evaluating estrogen receptor-alpha, progesterone receptor, and HER2 in breast cancer. *Mod Pathol* 2010;23 Suppl 2:S52–9. doi:10.1038/modpathol.2010.55.
- [84] Steck K, El-Naggar AK. Comparative flow cytometric analysis of Ki-67 and proliferating cell nuclear antigen (PCNA) in solid neoplasms. *Cytometry* 1994;17:258–65. doi:10.1002/cyto.990170309.
- [85] Beelen K, Zwart W, Linn SC. Can predictive biomarkers in breast cancer guide adjuvant endocrine therapy? *Nat Rev Clin Oncol* 2012;9:529–41. doi:10.1038/nrclinonc.2012.121.
- [86] Jirstrom K. Pathology parameters and adjuvant tamoxifen response in a randomised premenopausal breast cancer trial. *J Clin Pathol* 2005;58:1135–42. doi:10.1136/jcp.2005.027185.
- [87] van t'Veer LJ, Dai H, Vijver MJ Van De, Kooy K Van Der, Marton MJ, Witteveen AT, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415.
- [88] Cobleigh MA, Tabesh B, Bitterman P, Baker J, Cronin M, Liu ML, et al. Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clin Cancer Res* 2005;11:8623–31. doi:10.1158/1078-0432.CCR-05-0735.
- [89] Glas AM, Floore A, Delahaye LJMJ, Witteveen AT, Pover RCF, Bakx N, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006;7. doi:10.1186/1471-2164-7-278.
- [90] Zhang Y, Sieuwerts AM, McGreevy M, Casey G, Cufer T, Paradiso A, et al. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat* 2009;116:303–9. doi:10.1007/s10549-008-0183-2.
- [91] Kok M, Koornstra RH, Mook S, Hauptmann M, Fles R, Jansen MP, et al. Additional value of the 70-gene signature and levels of ER and PR for the prediction of outcome in tamoxifen-treated ER-positive breast cancer. *Breast* 2012;21:769–78. doi:10.1016/j.breast.2012.04.010.
- [92] Jansen MPH, Knijnenburg T, Reijm EA, Simon I, Kerkhoven R, Droog M, et al. Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. *Cancer Res* 2013;73:6632–41. doi:10.1158/0008-5472.CAN-13-0704.
- [93] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198–207. doi:10.1038/nature01511.
- [94] Cox J, Mann M. Is proteomics the new genomics? *Cell* 2007;130:395–8. doi:10.1016/j.cell.2007.07.032.

- [95] Mann M, Kulak NA, Nagaraj N, Cox J. The coming age of complete, accurate, and ubiquitous proteomes. *Mol Cell* 2013;49:583–90. doi:10.1016/j.molcel.2013.01.029.
- [96] Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. *Nature* 2014;509:575–81. doi:10.1038/nature13302.
- [97] Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, et al. Mass-spectrometry-based draft of the human proteome. *Nature* 2014;509:582–7. doi:10.1038/nature13319.
- [98] Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003;21:255–61. doi:10.1038/nbt0303-255.
- [99] Jensen ON. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* 2004;8:33–41. doi:10.1016/j.cbpa.2003.12.009.
- [100] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999;17:994–9. doi:10.1038/13690.
- [101] Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 2003;75:1895–904.
- [102] Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 2007;389:1017–31. doi:10.1007/s00216-007-1486-6.
- [103] Wang M, You J, Bemis KG, Tegeler TJ, Brown DPG. Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. *Briefings Funct Genomics Proteomics* 2008;7:329–39. doi:10.1093/bfgp/eln031.
- [104] Wiśniewski JR, Ostasiewicz P, Mann M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J Proteome Res* 2011;10:3040–9. doi:10.1021/pr200019m.
- [105] Liu NQ, Braakman RBH, Stingl C, Luidert TM, Martens JWM, Foekens JA, et al. Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J Mammary Gland Biol Neoplasia* 2012;17:155–64. doi:10.1007/s10911-012-9252-6.
- [106] Braakman RBH, Tilanus-Linthorst MMA, Liu NQ, Stingl C, Dekker LJM, Luidert TM, et al. Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. *J Proteomics* 2012;75:2844–54. doi:10.1016/j.jprot.2012.01.022.
- [107] Liu NQ, Dekker LJM, Stingl C, Güzel C, De Marchi T, Martens JWM, et al. Quantitative proteomic analysis of microdissected breast cancer tissues: Comparison of label-free and SILAC-based quantification with shotgun, directed, and targeted MS approaches. *J Proteome Res* 2013;12:4627–41. doi:10.1021/pr4005794.
- [108] Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* 2010;2010. doi:10.1155/2010/840518.

- [109] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008;26:1367–72. doi:10.1038/nbt.1511.
- [110] Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 2011;10:1794–805. doi:10.1021/pr101065j.
- [111] Sandin M, Chawade A, Levander F. Is label-free LC-MS/MS ready for biomarker discovery? *Proteomics - Clin Appl* 2015;9:289–94. doi:10.1002/prca.201400202.
- [112] Gerber S a, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 2003;100:6940–5. doi:10.1073/pnas.0832254100.
- [113] Harlan R, Zhang H. Targeted proteomics: a bridge between discovery and validation. *Expert Rev Proteomics* 2014;11:657–61. doi:10.1586/14789450.2014.976558.
- [114] Wang P, Whiteaker J, Paulovich A. The evolving role of mass spectrometry in cancer biomarker discovery. *Cancer Biol Ther* 2009;8:1083–94.
- [115] Whiteaker JR, Zhao L, Zhang HY, Feng L-C, Piening BD, Anderson L, et al. Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal Biochem* 2007;362:44–54. doi:10.1016/j.ab.2006.12.023.
- [116] Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 2006;24:971–83. doi:10.1038/nbt1235.
- [117] Schiess R, Wollscheid B, Aebersold R. Targeted proteomic strategy for clinical biomarker discovery. *Mol Oncol* 2009;3:33–44. doi:10.1016/j.molonc.2008.12.001.
- [118] Chung L, Baxter RC. Breast cancer biomarkers: proteomic discovery and translation to clinically relevant assays. *Expert Rev Proteomics* 2012;9:599–614. doi:10.1586/epr.12.62.
- [119] Rexer BN, Arteaga CL. Intrinsic and Acquired Resistance to HER2-Targeted Therapies in HER2 Gene-Amplified Breast Cancer: Mechanisms and Clinical Implications. *Crit Rev Oncog* 2012;17:1–16. doi:10.1615/CritRevOncog.v17.i1.20.
- [120] Chong CR, Jänne PA. The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat Med* 2013;19:1389–400. doi:10.1038/nm.3388.
- [121] Mumby M, Brekken D. Phosphoproteomics: new insights into cellular signaling. *Genome Biol* 2005;6:230. doi:10.1186/gb-2005-6-9-230.
- [122] Smit MA, Maddalo G, Greig K, Raaijmakers LM, Possik PA, Breukelen B Van, et al. ROCK 1 is a potential combinatorial drug target for BRAF mutant melanoma 2014:1–15.
- [123] Knight ZA, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer* 2010;10:130–7. doi:10.1038/nrc2787.Targeting.

- [124] Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, et al. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 2005;4:1240–50. doi:10.1074/mcp.M500089-MCP200.
- [125] Harsha HC, Pandey A. Phosphoproteomics in cancer. *Mol Oncol* 2010;4:482–95. doi:10.1016/j.molonc.2010.09.004.
- [126] Roepstorff P. Mass spectrometry based proteomics, background, status and future needs. *Protein Cell* 2012;3:641–7. doi:10.1007/s13238-012-2079-5.
- [127] Frese CK, Zhou H, Taus T, Altelaar AM, Mechtler K, Heck AJR, et al. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (ET<sub>h</sub>CD). *J Proteome Res* 2013;12:1520–5. doi:10.1021/pr301130k.
- [128] Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 2002;72:43R – 56R. doi:10.1093/glycob/12.4.43R.
- [129] Aebi M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* 2013;1833:2430–7. doi:10.1016/j.bbamcr.2013.04.001.
- [130] Pan S, Chen R, Aebersold R, Brentnall T a. Mass spectrometry based glycoproteomics-from a proteomics perspective. *Mol Cell Proteomics* 2011;10:R110.003251. doi:10.1074/mcp.R110.003251.
- [131] Shriver Z, Raguram S, Sasisekharan R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat Rev Drug Discov* 2004;3:863–73.
- [132] Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol Cell Proteomics* 2004;3:311–26. doi:10.1074/mcp.M300127-MCP200.
- [133] Johansson HJ, Sanchez BC, Mundt F, Forshed J, Kovacs A, Panizza E, et al. Retinoic acid receptor alpha is associated with tamoxifen resistance in breast cancer. *Nat Commun* 2013;4:2175. doi:10.1038/ncomms3175.
- [134] Johansson HJ, Sanchez BC, Forshed J, Stål O, Fohlin H, Lewensohn R, et al. Proteomics profiling identify CAPS as a potential predictive marker of tamoxifen resistance in estrogen receptor positive breast cancer. *Clin Proteomics* 2015;12:4–13. doi:10.1186/s12014-015-9080-y.
- [135] Majidzadeh-A K, Gharechahi J. Plasma proteomics analysis of tamoxifen resistance in breast cancer. *Med Oncol* 2013;30:753. doi:10.1007/s12032-013-0753-y.
- [136] Umar A, Kang H, Timmermans AM, Look MP, Meijer-van Gelder ME, den Bakker M a, et al. Identification of a putative protein profile associated with tamoxifen therapy resistance in breast cancer. *Mol Cell Proteomics* 2009;8:1278–94. doi:10.1074/mcp.M800493-MCP200.
- [137] Jr. DJJ, Rodriguez-Canales J, Mukherjee S, Prieto DA, C.Hanson J, Emmert-Buck M, et al. Approaching solid tumor heterogeneity on a cellular basis by tissue proteomics using laser capture microdissection and biological mass spectrometry. *J Proteome Res* 2010;8:2310–8. doi:10.1021/pr8009403.Approaching.

- [138] Longo D. Tumor Heterogeneity and Personalized Medicine. *NEJM* 2012;366:956–7.
- [139] Kondo T. Inconvenient truth: cancer biomarker development by using proteomics. *Biochim Biophys Acta* 2014;1844:861–5. doi:10.1016/j.bbapap.2013.07.009.
- [140] Yang F, Foekens JA, Yu J, Sieuwerts a M, Timmermans M, Klijn JGM, et al. Laser microdissection and microarray analysis of breast tumors reveal ER-alpha related genes and pathways. *Oncogene* 2006;25:1413–9. doi:10.1038/sj.onc.1209165.
- [141] Emmert-buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser Capture Microdissection. *Science* (80- ) 1996;274:8–11.
- [142] Xu BJ. Combining laser capture microdissection and proteomics: methodologies and clinical applications. *Proteomics Clin Appl* 2010;4:116–23. doi:10.1002/prca.200900138.



## Chapter 2

### Aim and outline of this thesis

Resistance to endocrine therapy is a major clinical problem. In the recurrent setting, nearly half of the patients with ER positive tumors that are treated with tamoxifen manifest intrinsic resistance to the drug, the other half develops resistance later on. Several studies have elucidated various mechanisms of resistance to tamoxifen therapy, though none of these observations have – so far - been translated into the clinic in terms of patient stratification for treatment or new treatment options.

With the advancements in proteomic technologies, in particular MS, changes in protein abundance and post-translational modifications are now being measured with high sensitivity and accuracy. Furthermore, targeted MS approaches provide multiplexable and high precision platforms to measure analytes from different kinds of samples (e.g. blood, primary tumor tissue). In the light of this, proteomic technologies offer a sensitive and accurate approach that is not only suitable for biomarker discovery studies, but may also provide a platform for potential introduction into clinical diagnostics.

In our laboratory, we developed and optimized a tissue proteomic workflow by combining laser capture microdissection (LCM) to high resolution MS analysis (1, 2) for the analysis of large patient cohorts and the discovery of prognostic and predictive biomarkers in breast cancer (e.g. (3)).

In the light of this, the scope of this thesis is:

- Evaluation of the advantage of cell enrichment techniques in proteomics based biomarker discovery.
- Development and validation of a protein based classifier for tamoxifen resistance.
- Design and development of a targeted MS based assay to quantify classifier proteins.
- Analysis of pathways and potential biomarkers present in subgroups of ER positive breast cancers.

Due to the fact that primary cancer tissues, and in particular breast cancer, are composed of various cell types (i.e. epithelial tumor cells, adipocytes, leukocytes, fibroblasts, etc.), we have successfully developed a tissue proteomic pipeline that combines LCM cell purification with high resolution MS (1, 2). Despite this, clear-cut differences between proteomic analysis of breast cancer LCM derived and whole tissue material was still unclear. In Chapter 3 we have elucidated the differences between proteomic analyses of whole tissue and LCM enriched material analyzed by high resolution MS in terms of identified MS/MS spectra, peptides and proteins. Furthermore, we evaluated the differences in the quantitative measurement of key breast cancer markers (e.g. ER, HER2).

Having established the viability of LCM in elucidating proteomic changes in breast cancer tissues, we applied this tissue proteomics pipeline in Chapter 4 for the development and validation of a 4-protein signature (PDCD4, OCIAD1, G3BP2 and CGN) predictive of tamoxifen resistance in recurrent ER positive breast cancer. In order to forward the measurement of the predictive 4-protein signature towards a clinical setting, a reproducible, accurate, fast, and preferably multiplexable assay would be advantageous over shotgun proteomics or single-plex IHC. Since the introduction of high resolution MS in clinical diagnostics is not easily feasible due to extensive sample measurement times, and due to the fact that only PDCD4 was validated at IHC level, we aimed at developing a multiplexable targeted MS assay. An example of such a technique is multiple reaction monitoring (MRM), which is capable of analyzing target proteins in a faster and more accurate way when compared to high resolution MS methods. In addition, we combined immunoaffinity enrichment through anti-peptide antibodies with MRM in order to overcome interference from tissue heterogeneity. This approach, defined as immuno-MRM (iMRM), is able to quantify proteins out of complex sample matrixes (e.g. plasma) in the sub-femtomolar range. In Chapter 5 we describe the development of an iMRM approach for targeted analysis of the protein signature in breast cancer primary tissue lysates. This workflow was then applied to measure peptide abundances representing the 4-protein predictor in breast cancer patient-derived material (e.g. whole tissue lysates). Chapter 6 describes the generation of a 4-protein classifier based on iMRM data in breast cancer tissues and its comparison to classifiers based on high resolution MS measurements of whole tissue and LCM enriched material. The 4 proteins were also measured in an independent cohort of breast cancer patient derived sera in order to assess whether a serum-based classifier could be developed.

Having developed a tamoxifen resistance predictive signature for recurrent ER positive breast cancer through high resolution MS, and a targeted MS assay to measure the 4 proteins with high precision, we aimed to investigate molecular mechanisms of the most differentially expressed

proteins in our proteomic dataset. Although the 4 signature proteins significantly predicted tamoxifen outcome groups, ANXA1 and CALD1 displayed the most notable changes in expression levels. In Chapter 7, we first assessed the clinical relevance of these proteins by IHC, and then performed pathway analysis on ANXA1, CALD1, and their correlated proteins, to elucidate which mechanisms of resistance these molecules might be involved in.

A plethora of biological pathways have been associated to tamoxifen resistance in breast cancer. Among these, metabolic changes in tumor cells, such as the diversion of glucose into the serine biosynthesis pathway, have been associated with cancer cell proliferation and oncogenesis (4, 5). Phosphoserine aminotransferase 1 (PSAT1) is an aminotransferase that catalyzes the transformation of phospho-pyruvate into phospho-serine, and its mRNA expression has been associated with poor outcome to tamoxifen treatment in ER positive recurrent breast cancer (6). In Chapter 8 we aimed at assessing the association of PSAT1 protein levels with tumor progression and, based on gene expression data, elucidate its role in tamoxifen resistance through pathway analysis.

In conclusion, this thesis demonstrates the viability of proteomic technologies in biomarker discovery studies, in the development of clinically feasible assays, and in the elucidation of proteomic changes of cancer cells in tamoxifen resistant ER positive breast cancer. Furthermore, we here point out that different mechanisms of resistance may arise in subsets of tumors, which are linked to cytokine signaling and cancer cell metabolic changes.

## References:

1. Braakman, R. B. H., Tilanus-Linthorst, M. M. A., Liu, N. Q., Stingl, C., Dekker, L. J. M., Luider, T. M., Martens, J. W. M., Foekens, J. A., and Umar, A. (2012) Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. *J. Proteomics* 75, 2844–54
2. Liu, N. Q., Braakman, R. B. H., Stingl, C., Luider, T. M., Martens, J. W. M., Foekens, J. A., and Umar, A. (2012) Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J. Mammary Gland Biol. Neoplasia* 17, 155–64
3. Liu, N. Q., Stingl, C., Look, M. P., Smid, M., Braakman, R. B. H., De Marchi, T., Sieuwerts, A. M., Span, P. N., Sweep, F. C. G. J., Linderholm, B. K., Mangia, A., Paradiso, A., Dirix, L. Y., Van Laere, S. J., Luider, T. M., Martens, J. W. M., Foekens, J. A., and Umar, A. (2014) Comparative Proteome Analysis Revealing an 11-Protein Signature for Aggressive Triple-Negative Breast Cancer. *J. Natl. Cancer Inst.* 106, Epub 2014
4. Possemato, R., Marks, K. M., Shaul, Y. D., Pacold, M. E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H.-K., Jang, H. G., Jha, A. K., Chen, W. W., Barrett, F. G., Stransky, N., Tsun, Z.-Y., Cowley, G. S., Barretina, J., Kalaany, N. Y., Hsu, P. P., Ottina, K., Chan, A. M., Yuan, B., Garraway, L. a, Root, D. E., Mino-Kenudson, M., Brachtel, E. F., Driggers, E. M., and Sabatini, D. M. (2011) Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* 476, 346–50
5. Locasale, J. W., Grassian, A. R., Melman, T., Lyssiotis, C. a, Mattaini, K. R., Bass, A. J., Heffron, G., Metallo, C. M., Muranen, T., Sharfi, H., Sasaki, A. T., Anastasiou, D., Mullarky, E., Vokes, N. I., Sasaki, M., Beroukhim, R., Stephanopoulos, G., Ligon, A. H., Meyerson, M., Richardson, A. L., Chin, L., Wagner, G., Asara, J. M., Brugge, J. S., Cantley, L. C., and Vander Heiden, M. G. (2011) Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nat. Genet.* 43, 869–74
6. Martens, J. W. M., Nimmrich, I., Koenig, T., Look, M. P., Harbeck, N., Model, F., Kluth, A., Bolt-De Vries, J., Sieuwerts, A. M., Portengen, H., Meijer-Van Gelder, M. E., Piepenbrock, C., Olek, A., Höfler, H., Kiechle, M., Klijn, J. G. M., Schmitt, M., Maier, S., and Foekens, J. A. (2005) Association of DNA methylation of phosphoserine aminotransferase with response to endocrine therapy in patients with recurrent breast cancer. *Cancer Res.* 65, 4101–7

## **Chapter 3**

### **The advantage of laser-capture microdissection over whole tissue analysis in proteomic profiling studies**

Tommaso De Marchi, Rene B. Braakman, Christoph Stingl, Martijn M. van Duijn, Marcel Smid, John A. Foekens, Theo M. Luider, John W. Martens, Arzu Umar.

*Proteomics*, 2016; 16, 10, 1474-85

## **Abstract**

Laser capture microdissection (LCM) offers a reliable cell population enrichment tool and has been successfully coupled to MS analysis. Despite this, most proteomic studies employ whole tissue lysate (WTL) analysis in the discovery of disease biomarkers and in profiling analyses. Furthermore, the influence of tissue heterogeneity in WTL analysis, nor its impact in biomarker discovery studies have been completely elucidated. In order to address this, we compared previously obtained high resolution MS data from a cohort of 38 breast cancer tissues, of which both LCM enriched tumor epithelial cells and whole tissue lysate samples were analyzed. Label-free quantification (LFQ) analysis through MaxQuant software showed a significantly higher number of identified and quantified proteins in LCM enriched samples (3,404) compared to whole tissue lysates (2,837). Furthermore, WTL samples displayed a higher amount of missing data compared to LCM both at peptide and protein levels ( $p\text{-value} < 0.001$ ). 2D analysis on co-expressed proteins revealed discrepant expression of immune system and lipid metabolisms related proteins between LCM and WTL samples. We hereby show that LCM better dissected the biology of breast tumor epithelial cells, possibly due to lower interference from surrounding tissues and highly abundant proteins.

## Introduction

Current proteomics technologies are capable of quantifying almost the entire human proteome due to improvements in sample preparation, liquid chromatography, and MS instruments [1–3]. Furthermore, MS based techniques have proven to add another layer of information - when compared to gene expression data -in the profiling of complex diseases, such as cancer [4]. Despite this, proteomic characterization of tumor tissues is often thwarted by the presence of diverse cell subpopulations that increase tissue morphological heterogeneity [5], such as stromal cells, leukocytes, endothelial cells, adipocytes and other cell types. In the analysis of breast carcinomas, complexity can be further enhanced by the presence of normal epithelial tissue, which, if not segregated prior to analysis, may influence protein quantification [6,7]. Cell enrichment techniques such as LCM offer a robust and reliable tool to isolate specific cell types from their harboring tissues, and can then be subjected to molecular analyses that allow elucidation of biological features with associated molecular mechanisms [8,9]. As an example, LCM based enrichment is well suited for both downstream genomic and proteomic analyses, allowing detection of thousands of transcripts and proteins out of heterogeneous biological samples [10–13]. However, due to the minute amounts of material yielded, quantitative proteomics through peptide labeling strategies is very challenging in such a workflow. In recent years, label-free quantification (LFQ) algorithms have been extensively improved, allowing high sensitivity and robustness in the identification of large numbers of proteins [14]. In particular, MaxQuant software enables protein identification and quantitation with high mass accuracy (i.e. p.p.b. range) and precision based on the assembly of 3D peaks (i.e. counts/s, m/z, and retention time) [15]. Furthermore, MaxQuant enables identification transfer across samples through the “match-between-runs” option, which aligns peptide retention times between different runs, where alignment order is determined by hierarchical clustering, and applies a Gaussian kernel smoothing to mass matches between different runs. This way, the amount of overall identifications in a dataset is generally increased more than 2-fold while missing data are significantly reduced [16,17]. As a consequence, an increase in number of aligned samples would likely maximize peptide identifications and reduce the presence of missing values. In our laboratory, we established a tissue proteomic pipeline that not only allows robust quantification of thousands of proteins from minute amounts of microdissected epithelial tumor cells using MaxQuant [12,13], but has also proven to be a valuable tool in the development and validation of predictive and prognostic biomarkers [18,19]. Due to the time-consuming sample preparation procedure for LCM, whole tissue analysis is often preferred over LCM enrichment in profiling and biomarker discovery studies. So far, no in-depth comparison between the two approaches has been described with regard to its performance in a biomarker discovery pipeline.

We previously generated and validated a protein predictor for tamoxifen therapy resistance in a cohort of 112 breast cancer tissues, from which epithelial tumor cells were enriched by LCM and analyzed through high resolution MS [19]. For 38 of these tumors, whole tissue lysate (WTL) shotgun proteomic data were generated as well. We hereby compare previously generated proteomic profiles of matching LCM and WTL samples to evaluate their performance in total protein identification, and in the perspective of biomarker discovery studies. To assess the influence of identification transfer between samples using MaxQuant, LCM and WTL mass spectrometer RAW files were searched as one and as two separate data sets. Differences between the LCM and WTL sets were evaluated at the MS, peptide, and protein levels.

## **Materials and methods**

### *Patient samples*

A total of 38 estrogen receptor (ER) positive primary breast cancer tissues constituting a subset of a previously described cohort (De Marchi et al, under review, and [19,20] were available for this study. Samples were selected out of a total of 112 ER positive fresh frozen primary breast cancer tissue samples derived from three academic medical centers in the Netherlands (years of collection: 1980-1996): Erasmus MC University Medical Center (EMC), Rotterdam, the National Cancer Institute – Antoni van Leeuwenhoek hospital (NKI-AVL), Amsterdam, and Radboud University Medical Center (RadboudUMC), Nijmegen. ER positivity in tumor cytosols was assessed by quantitative biochemical assays (EMC), reverse-transcriptase quantitative polymerase chain reaction (RT-PCR; RadboudUMC), or IHC (NKI-AVL). ERBB2 status was determined by RT-PCR (EMC and RadboudUMC) and IHC (NKI-AVL). Tumors were selected based on a tumor area (> 50%) after histological evaluation, resulting in a set of 38 samples (EMC: n = 15; RadboudUMC: n = 4; NKI-AVL: n = 19). All patients underwent surgery of their primary tumor, developed recurrent disease, and were treated with tamoxifen as first line therapy. Good and poor outcome groups for tamoxifen treatment of patients with recurrent breast cancer were defined based on whether tumor progression occurred before ( $\leq$ ; poor) or after (>; good) 6 months after start of therapy. The study cohort comprised a total of 25 good and 13 poor outcome patients.

### *LCM and WTL sample preparation*

All frozen tissue samples were cut on a Microm HM 560 cryostat (Thermo Scientific), collecting 1 (5  $\mu\text{m}$ ) section for downstream hematoxylin and eosin (HE) staining, 10 x 10  $\mu\text{m}$  sections for WTL preparation, and 8 x 8  $\mu\text{m}$  sections for downstream LCM. HE staining was performed as follows: distilled water (1 min), hematoxylin (30 sec), distilled water (1 min), eosin (30 sec), distilled water (1 min), 50% ethanol (1 min), 70% ethanol (1 min), 100% ethanol (1 min), 100% ethanol (1 min), 100% ethanol (1 min), Xylene (1 min), Xylene (1 min), Xylene (1 min). Sections for WTL processing were collected into a LoBind™ Eppendorf tube and immediately stored at  $-80^{\circ}\text{C}$ . Sections for downstream LCM were collected on polyethylene naphthalate (PEN) coated glass slides (Carl Zeiss Microsystems GmbH, Göttingen, Germany), which were previously sterilized by UV exposure (1 min). Collected sections were immediately dehydrated with 95% ethanol and stored at  $-80^{\circ}\text{C}$  for further processing. Prior to LCM procedure, slides were thawed (room temperature) and hematoxylin staining was performed, as follows: distilled water (1 min), hematoxylin (30 sec), distilled water (1 min), 50% ethanol (1 min), 70% ethanol (1 min), 95% ethanol (1 min), 100% ethanol (1 min), 100% ethanol (1 min). All ethanol solutions contained Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific Inc, Rockford, IL, USA) at a 1:100 v/v concentration. From each specimen, an area of approximately 500,000  $\mu\text{m}^2$  was collected from each tissue using a photo-activated localization microscopy Micro Beam Axio Observer A1 device and gathered in an adhesive cap (Carl Zeiss Microsystems GmbH, Göttingen, Germany). A volume of 20  $\mu\text{l}$  of 0.1% w/v Rapigest surfactant (Waters Corporation, Milford, MA, USA) in 50 mM ammonium bicarbonate was used to transfer the collected LCM samples into LoBind™ Eppendorf tubes (Eppendorf AG, Hamburg, Germany). Tissue containing buffer was frozen after collection and stored at  $-80^{\circ}\text{C}$ .

### *Tissue disruption and protein digestion*

LCM collected material and sections for WTL were thawed at room temperature. A total of 100  $\mu\text{l}$  of 0.1% w/v Rapigest surfactant in 50 mM ammonium bicarbonate solution was added to collected WTL sections. LCM and WTL material was sonicated using a horn sonifier bath (Ultrasonic Disruptor Sonifier II, Bransons Ultrasonics, Danbury, CT, USA) at 70% amplitude. Prior protein digestion, LCM and WTL sonicated material was then spun down at 14,000 g. Supernatants were then collected and transferred into a new tube. Protein concentration of WTL material was assessed by bicinchoninic acid (BCA) assay (Pierce - Thermo Fisher Scientific).

Proteins were denatured at 95°C (5 min), and subsequently reduced with 100 mM DTT (30 min) at room temperature. Alkylation was performed by adding 300 mM iodoacetamide to each sample and incubating them (30 min) in the dark at room temperature. Samples were then digested for 4 h (LCM) or overnight (WTL) at 37°C after addition of MS grade trypsin at a 1:4 enzyme-protein ratio (i.e. 100 ng/μl). Samples were acidified with TFA, and spun down at 14,000 g. Supernatants were collected and transferred to HPLC vials (Sigma-Aldrich Corporation, St. Louis, MO, USA) for downstream MS analysis.

### *High resolution mass spectrometry analysis*

The LCM cohort was analyzed by high resolution MS as previously reported [19,20]. Before LC-MS analysis of the WTL cohort, test LC-MS runs were performed by injecting different amounts of digested protein lysate (i.e. 1 μl, 2 μl, 3 μl, 4 μl, 5 μl). Ultra-violet traces at 214 nm of WTL sample test run chromatograms showing the same intensity (measured in mAU) as the matching LCM samples were then selected as final load volume, which roughly corresponded to ~250 ng of digested proteins as per previous BCA measurements.

All MS measurements were performed using a nano liquid chromatography system (Ultimate 3000, Dionex, Amsterdam, The Netherlands), which was coupled online to a linear Ion Trap – Orbitrap XL™ mass spectrometer (Thermo Electron, Bremen, Germany). Samples were loaded on a trap column (PepMap C18, 300 μm ID × 5 mm length, 5 μm particle size, 100 Å pore size; Dionex), and subsequently washed and desalted in 0.1% TFA acidified water. Trap column and analytical column (PepMap C18, 75 μm ID × 50 cm, 3 μm particle size and 100 Å pore size; Dionex) were then coupled and peptides were eluted in a 3 h binary gradient (flow: 300 nl/min; mobile phase A: 2% acetonitrile and 0.1% formic acid in H<sub>2</sub>O; mobile phase B: 80% acetonitrile and 0.08% formic acid). Gradient was run as follows: 0% to 25% mobile phase B for 2 h, increase to 50% mobile phase B in 1 h. For ESI, metal-coated nano ESI emitters (New Objective, Woburn, MA) were used and a spray voltage of 1.6 kV was applied. High-resolution scan with a resolution of 30,000 (at 400 m/z) was acquired from 400 to 1,800 Th and was used for MS detection. Automatic gain was set at 106 ions and lock mass was set at 445.120025 u protonated with (Si(CH<sub>3</sub>)<sub>2</sub>O)<sub>6</sub>. Charge state screening was enabled and unassigned charge states and single charged precursor ions were rejected. The 5 most intense peaks in full scan were selected (isolation width = 2 u) and fragmented by collision induced dissociation (CID) applying 35% normalized collision energy and detected in

the ion trap. Precursor masses within ( $\pm$ ) 5 ppm tolerance range that were selected once for MS/MS were excluded for further fragmentation for 3 min (dynamic exclusion enabled).

### *Mass spectrometry files analysis*

High resolution proteome profiles were previously generated from ~250 ng protein lysate both from LCM and WTL samples. Orbitrap.RAW files were previously deposited to ProteomeXchange (dataset identifier: PXD002381) via the PRIDE partner repository. MaxQuant (version 1.5.2.8) [15] with Andromeda as search engine [16] was used for protein identification/quantitation [21]. LCM and WTL sets were analyzed both as one set (aligned) and separately (separate; Fig.S1). Analysis of MS spectra was performed by selecting acetylation of the protein N-terminus and oxidation of methionine as variable modifications, while carbamidomethylation was kept as fixed modification. UniProt-SwissProt human canonical database (version 2015-02, canonical proteome; 20,198 identifiers) was selected as FASTA file. Seven amino acids were selected as minimum peptide length. Match between runs option was kept as default (match time window: 0.7 minutes; alignment time window: 20 minutes). LFQ was enabled and LFQ minimum ratio count was set to 1. Remaining options were kept as default.

### *Data analysis*

“ModificationSpecificPeptides.txt” and “ProteinGroups.txt” were imported and analyzed in Microsoft Excel. Peptides were filtered for Posterior Error Probability (PEP;  $< 0.05$ ), and proteins for Q-value ( $< 0.05$ ). Possible contaminants and reversed sequences were excluded. Intensities (peptides) and LFQ intensities (proteins) were Log10 transformed. Further peptide and protein filtering for missing data (defined as number of missing observation per sample) was performed in order to assess global identifications. “MS/MS submitted”, “MS/MS identified”, “Peaks sequenced” and “Peaks repeatedly sequenced” were extracted from “Summary.txt”. “Retention length” of shared peptides between LCM and WTL samples and “MS/MS counts” of all repeatedly sequenced peptides (i.e. MS/MS counts  $> 1$ ) were extracted from “Evidence.txt”. Significant differences in retention lengths, MS/MS counts, mean abundances, and missing data were assessed by Mann-Whitney (unpaired) and Wilcoxon ranked sum (paired) tests in GraphPad (v5.1). Pearson correlation was performed on abundances of shared peptides and proteins.

## Statistics and pathway analyses

Gene Ontology (GO) [22] cellular component annotations were retrieved for all quantified proteins in both WTL and LCM sets from the Uniprot database [23], and protein intensities were plotted according to organelle distribution. Differences in protein distributions within subcellular compartments between the two sets were assessed by  $\chi^2$  test. Hierarchical clustering analysis was performed using Cluster (v3.0; distance metric: correlation [uncentered], analysis: complete linkage on protein levels only) [24].

Mean abundances of all shared proteins between the two sets after aligned search were imported in Perseus (v 1.5.1.6). Furthermore, difference in mean abundances of shared proteins between LCM and WTL samples and significant proteins between good and poor outcome patients in each dataset were assessed by t-test (unpaired; unequal variances assumed). Concordance of expression between differentially expressed proteins in both LCM and WTL samples was assessed after transforming log ratios (i.e. poor vs good outcome) into dichotomous values, defined as Log ratio scores. Log ratio scores showing a negative expression direction ( $< 0$ ) were given a negative value (i.e. -1) while a positive value (i.e. +1) was given to Log ratios displaying positive expression directions ( $> 0$ ). Concordance score was calculated according to the following formula:

Concordance score = Log ratio score [WTL] - Log ratio score [LCM].

All concordant expression directions displayed a null concordance score (i.e. 0), while discordant ones displayed a positive or negative (i.e.  $\pm 2$ ) score.

All shared significant proteins between the two sets were also imported in Perseus with their Log ratios. Both protein lists (shared and shared differentially expressed) were annotated for GO biological process [22], and 2D analysis [25] was performed. Delta scores derived from 2D analysis were plotted in Microsoft Excel.

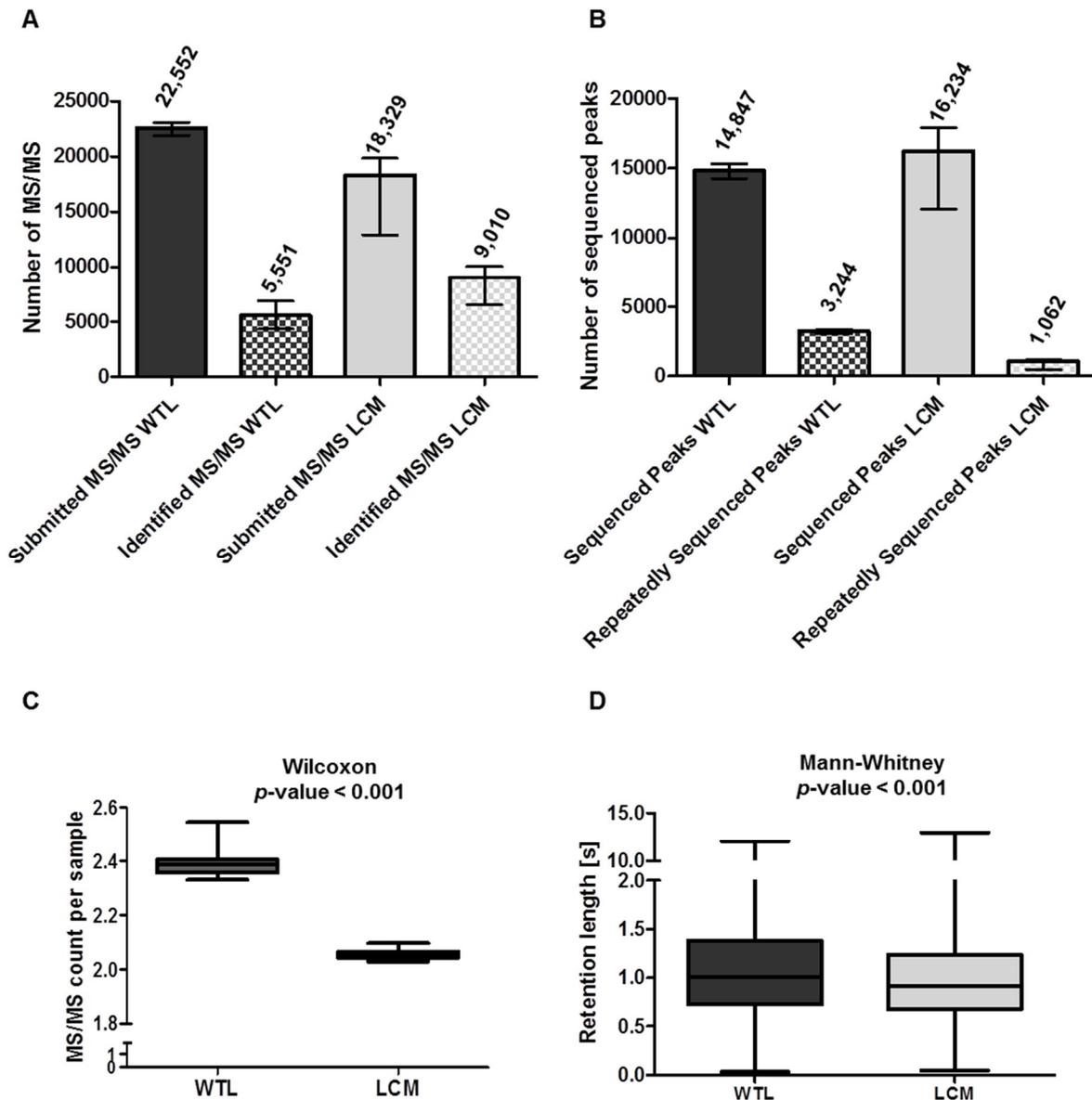
## Results

### *LCM enrichment results in increased MS feature identification and less peak resampling*

In order to evaluate the effect of identification transfer by MaxQuant, LCM and WTL .RAW files were searched both as one as well as separate data sets. After aligned search, a total of 214,194 and 325,230 MS/MS spectra were identified in the WTL and LCM sets, respectively (Fig. S2A). The WTL set showed a lower number of identified fragmentation spectra/sample (median: 5,551

[24.72% of total]) when compared to the LCM set (median: 9,010 [49.2% of total]; Fig. 3.1A). Upon testing for differences, a median of 14,847 sequenced chromatographic peaks was observed in WTL samples in an aligned search, while a median of 16,234 peaks was observed in LCM samples ( $p$ -value = 0.165). A higher amount of repeatedly sequenced peaks was, however, observed in WTL samples (median aligned: 3,244 [21.8%]) respective to the LCM set (median aligned: 1,062 [6.5%]; Fig. 3.1B). Repeatedly sequenced peaks matched to a total of 13,760 features (6,985 peptides, 1,544 proteins) in the WTL (Table S1), and 11,107 features (7,331 peptides, 1,855 proteins) in the LCM set (Table S2). Of these, median MS/MS counts per sample were 2.39 and 2.06 for the WTL and LCM sets (Wilcoxon rank  $p$ -value < 0.001), respectively (Fig. 3.1C). Subcellular protein distribution in LCM and WTL sets were significantly different ( $\chi^2$   $p$ -value < 0.001), with several GO-annotated plasma (e.g. APOA1) and extracellular matrix (ECM; e.g. SERPINA1) proteins detected in the WTL set only (Fig. S3A-B). Overall, these data suggested a higher redundancy in MS/MS spectra identifications in the WTL set. Furthermore, for shared sequenced peaks, WTL samples displayed a significantly higher retention time length compared to LCM (median WTL: 1.00 s; median LCM: 0.91 s;  $p$ -value < 0.001; Fig. 3.1D), though such evidence may not be sufficient to explain peak resampling.

Similar results were displayed after separate search: a total of 211,862 (WTL) and 327,276 (LCM) MS/MS were identified (Fig. S2B). Out of submitted MS/MS spectra, 24.8% and 49.0% were identified as peptides in the WTL and LCM sets, respectively (Fig. S4A). When assessing differences between number of sequenced peaks, no significant difference was found between sets (median WTL: 14,848; median LCM: 16,219;  $p$ -value = 0.258), though the WTL set still displayed a higher number of repeatedly sequenced peaks (median: 3,246 [21.9%]) when compared to LCM samples (median: 1,060 [6.5%]; Fig. S4B). Such analysis after separate search yielded similar results as the aligned search (data not shown), and for shared sequenced peaks, the WTL set displayed a significantly higher retention length (median: 1.29 s) when compared to LCM samples (median: 0.85 s;  $p$ -value < 0.001; Fig. S4C). These data suggest that WTL samples suffer from higher redundancy in MS identifications respect to LCM specimens.



**Figure 3.1. MS/MS level analysis of LCM and WTL sets after aligned MaxQuant search.**

Comparison between submitted and identified MS/MS spectra (A; bars represent median level and interquartile range) and number of sequenced and repeatedly sequenced chromatographic peaks in WTL and LCM sets (B; bars represent median level and interquartile range). When compared to LCM, WTL displayed significantly higher median MS/MS counts (C) and retention time length (D).

Acronyms: LCM: laser capture microdissection; WTL: whole tissue lysate; MS: mass spectrometry.

### *Peptide identification is increased in LCM samples*

Based on the differences found between WTL and LCM samples at the MS level, we next investigated whether this translated to the peptide level. After aligned search, while a high number of peptides (14,960; median Pearson  $r$  between sets = 0.05) and unique peptides (13,616; median Pearson  $r$  = 0.06) were identified in both sets, the WTL cohort displayed a lower number of unique

identifications (peptides: 2,682; unique peptides: 2,475) when compared to the LCM set (peptides: 10,952; unique peptides: 10,368; Fig. 3.2A-B). Waterfall plots of shared total and unique peptides between the two sets showed only a slight offset between abundances when aligned search was performed (Fig. 3.2C-D).

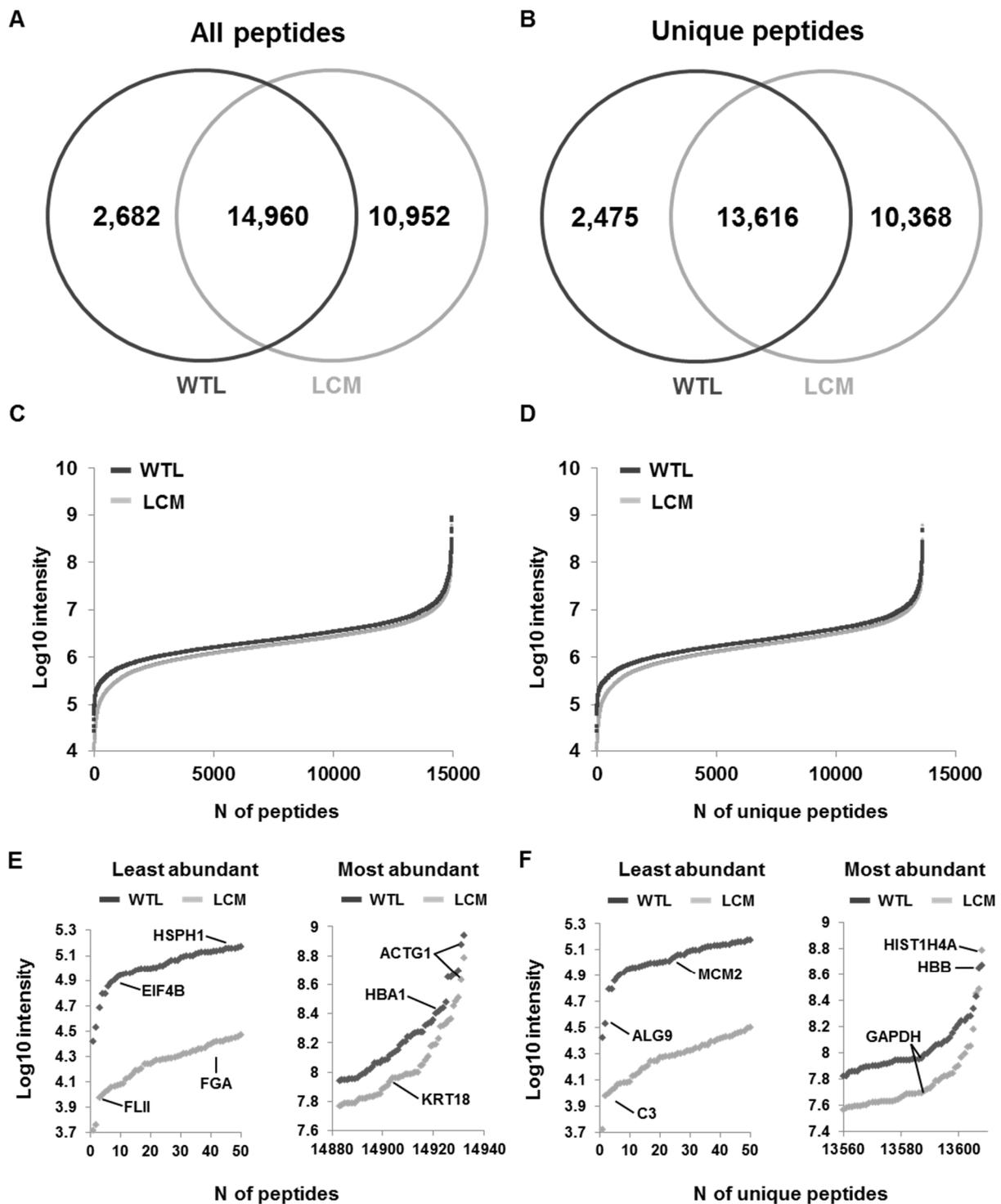


Figure 3.2. Peptide identifications in the LCM and WTL sets after aligned MaxQuant search.

A large number of total overlapping peptides (A) and unique peptides (B) were found between the LCM and WTL sets. Waterfall plots for shared total peptides (C) and unique peptides (D) showed a comparable dynamic range in both sets. Fifty least and most abundant peptides are also shown (E-F).

Among least and most abundant peptides, plasma protein-derived peptides showed higher abundance in the WTL (e.g. HBB) and lower intensity (e.g. C3) in LCM (Fig. 3.2E-F). Furthermore, the WTL set showed a significantly higher number of missing data for both total identified ( $p$ -value < 0.001; Fig. S5A) and unique peptides ( $p$ -value < 0.001; Fig. S5B).

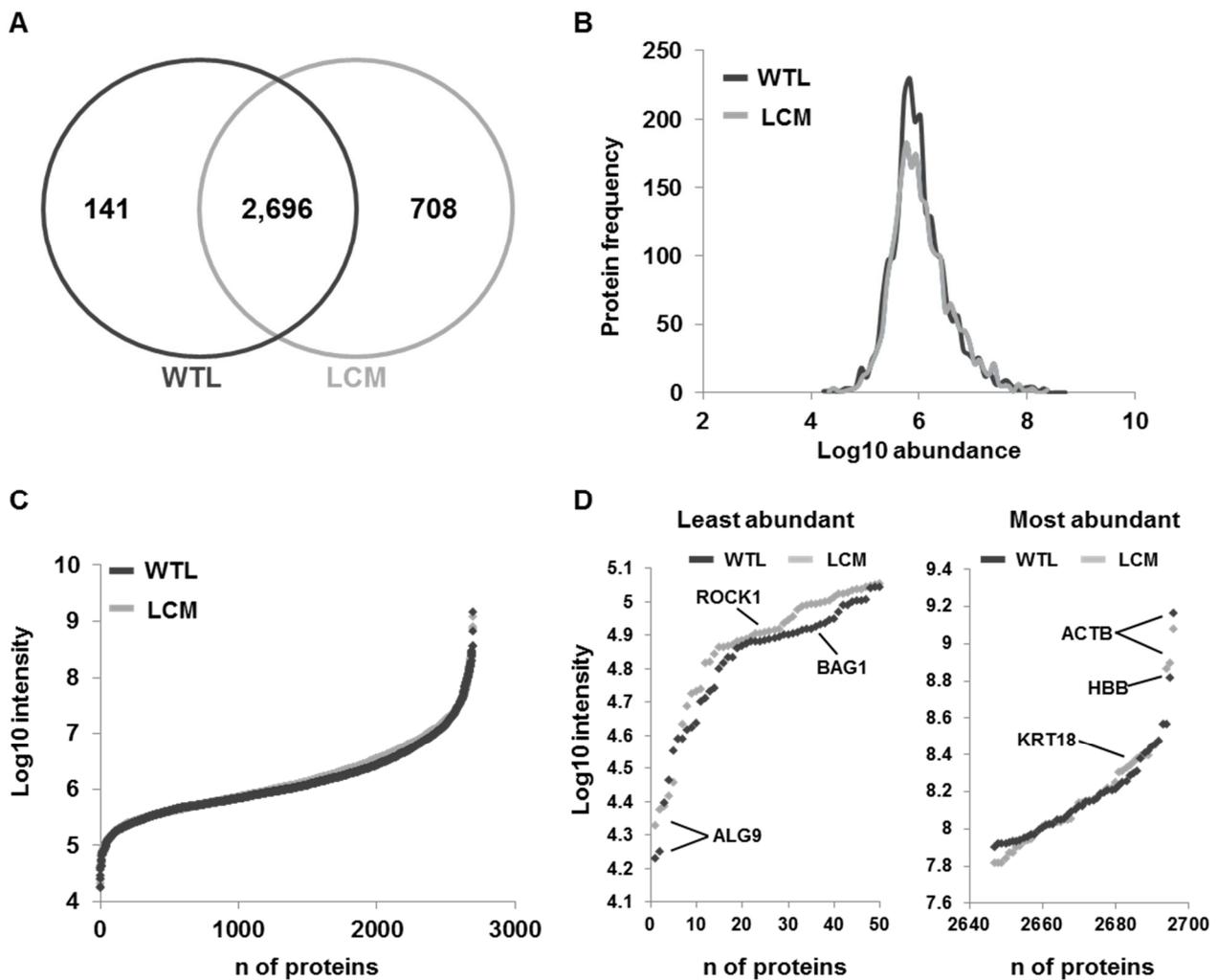
Separate search yielded similar results but less total identifications: 12,927 (unique: 11,671) and 25,454 (unique: 23,552) peptides were identified in WTL and LCM samples, respectively (Fig. S6A-B). Mean Pearson correlation of shared peptides (9,414) and unique peptides (8,434) between sets were -0.06 and -0.05, respectively. Furthermore, offsets between protein abundances (Fig. S6C-D) and missing data ( $p$ -value < 0.001; Fig. S5C-D) were more pronounced. These data suggest that alignment not only increased identification rates (aligned vs separate; peptides: +36.5%; unique peptides: +37.9%), but also decreased protein missing data in the WTL set. The LCM set did not greatly benefit from aligned search (aligned vs. separate; peptides: +1.80%; unique peptides: +1.83%).

#### *Proteomic analysis of LCM material yields higher amounts of identified proteins and lower missing observations*

Similar to our analysis at the MS/MS and peptide levels, we assessed differences at protein level. From aligned search a total of 2,837 and 3,404 proteins were identified in WTL and LCM sets, respectively (shared proteins: 2,696; median Pearson  $r$  = 0.45; Fig. 3.3A). Of these, a higher frequency of protein abundances along the intensity median was observed in WTL samples (Fig. 3.3B). However, waterfall plots showed similar dynamic ranges between sets (i.e. over 6 orders of magnitude; Fig. 3.3C). As observed at the peptide level, plasma proteins (e.g. HBB) were observed at high levels in the WTL set (Fig. 3.3D). In addition, the median number of missing data points in the WTL set was significantly higher than in LCM samples (median WTL = 26.32%; median LCM = 5.26%;  $p$ -value < 0.001; Fig. S7A).

From the separate search a total of 2,047 and 3,394 proteins were identified in the WTL and LCM sets, respectively, of which 1,815 were in common (median Pearson  $r$  = 0.52; Fig. S8A). Discrepancies in protein abundance frequency (Fig. S8B), and in mean abundance (Fig. S8C) were

observed. Least and most abundant proteins in the LCM and WTL sets also differed, confirming our results at the peptide level (Fig. S8D). In addition, the difference in missing data between WTL and LCM samples was more pronounced (median LCM = 0.0%; median WTL = 60.53%; *p*-value < 0.001; Fig. S7B) compared to aligned search. These data suggest that, also at the protein level, alignment increased identifications (+38.6%), and reduced missing values (for shared 1,815 identified proteins: -50.0%) in the WTL set, while no major change (identified proteins: +0.3%; missing data: +0.0%) was observed in the LCM set.

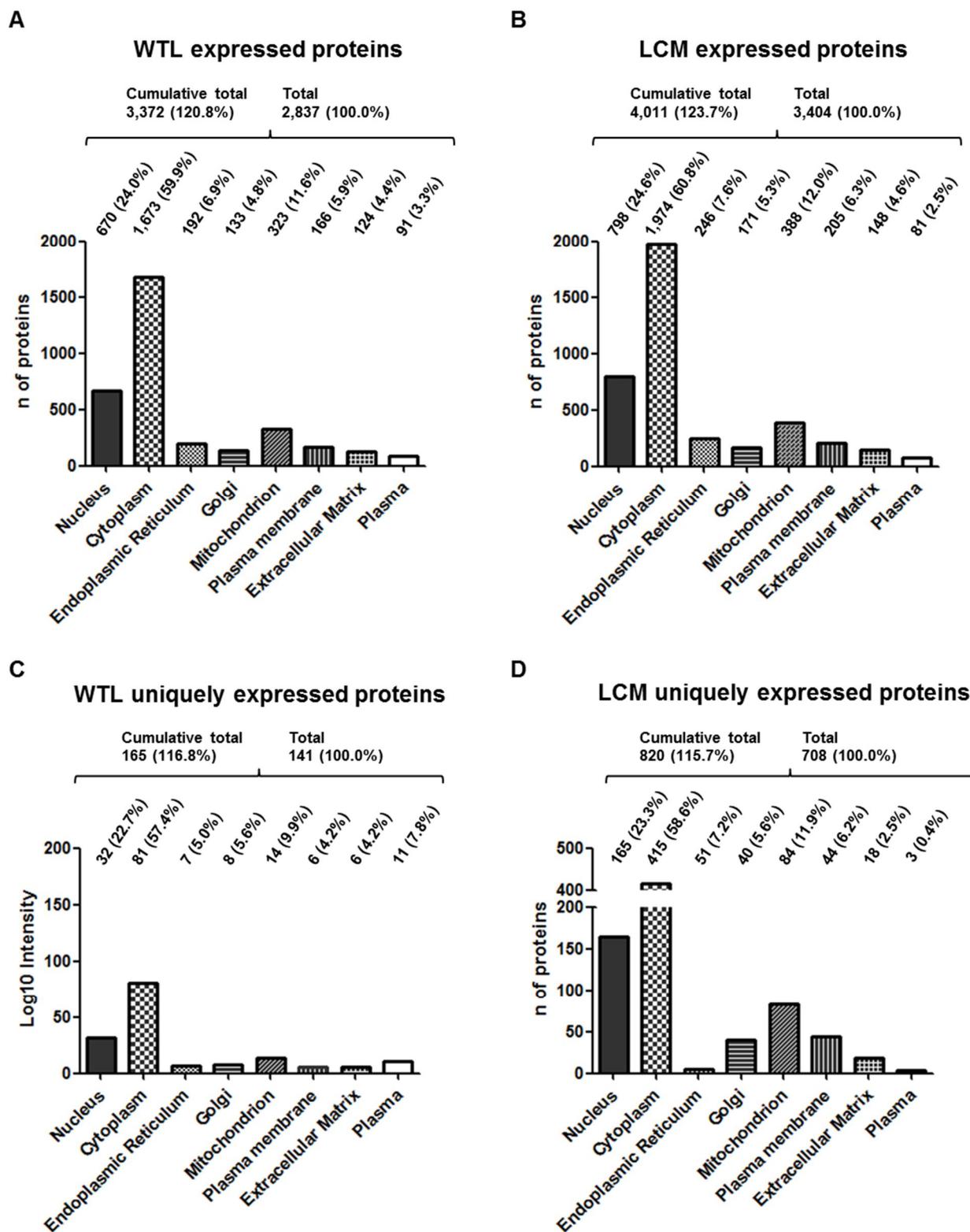


**Figure 3.3. Protein level analysis of the LCM and WTL sets after aligned MaxQuant search.**

Proteins identified in both LCM and WTL sets (A). WTL samples displayed a higher number of protein abundances localized around the median (B), however dynamic ranges were comparable (C). Fifty least and most abundant proteins are also shown (D).

### *LCM enrichment minimizes interference from the background proteome*

In order to assess whether LCM would minimize ECM and plasma protein identifications, all quantified proteins in both sets were annotated for GO cellular component. After aligned search, both sets showed a comparable number of identifications across cellular compartments ( $\chi^2$  *p-value* = 0.674; Fig. 3.4A-B). A relatively small percentage of proteins (< 5%) belonged to the ECM and plasma compartments with no difference between sets ( $\chi^2$  *p-value* = 0.156). Proteins unique to each set also encompassed all subcellular compartments; however a significant difference in distribution was observed ( $\chi^2$  *p-value* < 0.001), due to enrichment in plasma proteins in WTL samples ( $\chi^2$  *p-value* < 0.001; Fig. 3.4C-D). A significantly different distribution of proteins in cellular compartments was also observed after separate search ( $\chi^2$  *p-value* < 0.001; Fig. S9A-B), due to a higher percentage of ECM ( $\chi^2$  *p-value* < 0.001) and plasma ( $\chi^2$  *p-value* < 0.001) proteins in WTL samples. Compartment distributions of proteins unique to each set were also significantly different ( $\chi^2$  *p-value* < 0.001; Fig. S9C-D). In fact, while no endoplasmic reticulum ( $\chi^2$  *p-value* < 0.001) or Golgi-related ( $\chi^2$  *p-value* < 0.001) proteins were observed in WTL samples, the LCM set showed a small number of ECM ( $\chi^2$  *p-value* < 0.001) and plasma ( $\chi^2$  *p-value* < 0.001) proteins. These data suggest that LCM enrichment minimizes ECM and plasma protein interference when investigating the tumor cell proteome.



**Figure 3.4. Subcellular localization of all proteins identified in WTL and LCM datasets after aligned MaxQuant search.**

All identified proteins in LCM and WTL sets after aligned search were annotated for subcellular localization. Both the WTL (A) and LCM (B) datasets showed a large number of identified proteins belonging to intracellular organelles. Proteins identified only in the WTL (C) and LCM (D) sets both spanned across all subcellular compartments. Proteins identified per each subcellular compartment are reported as: number (%).

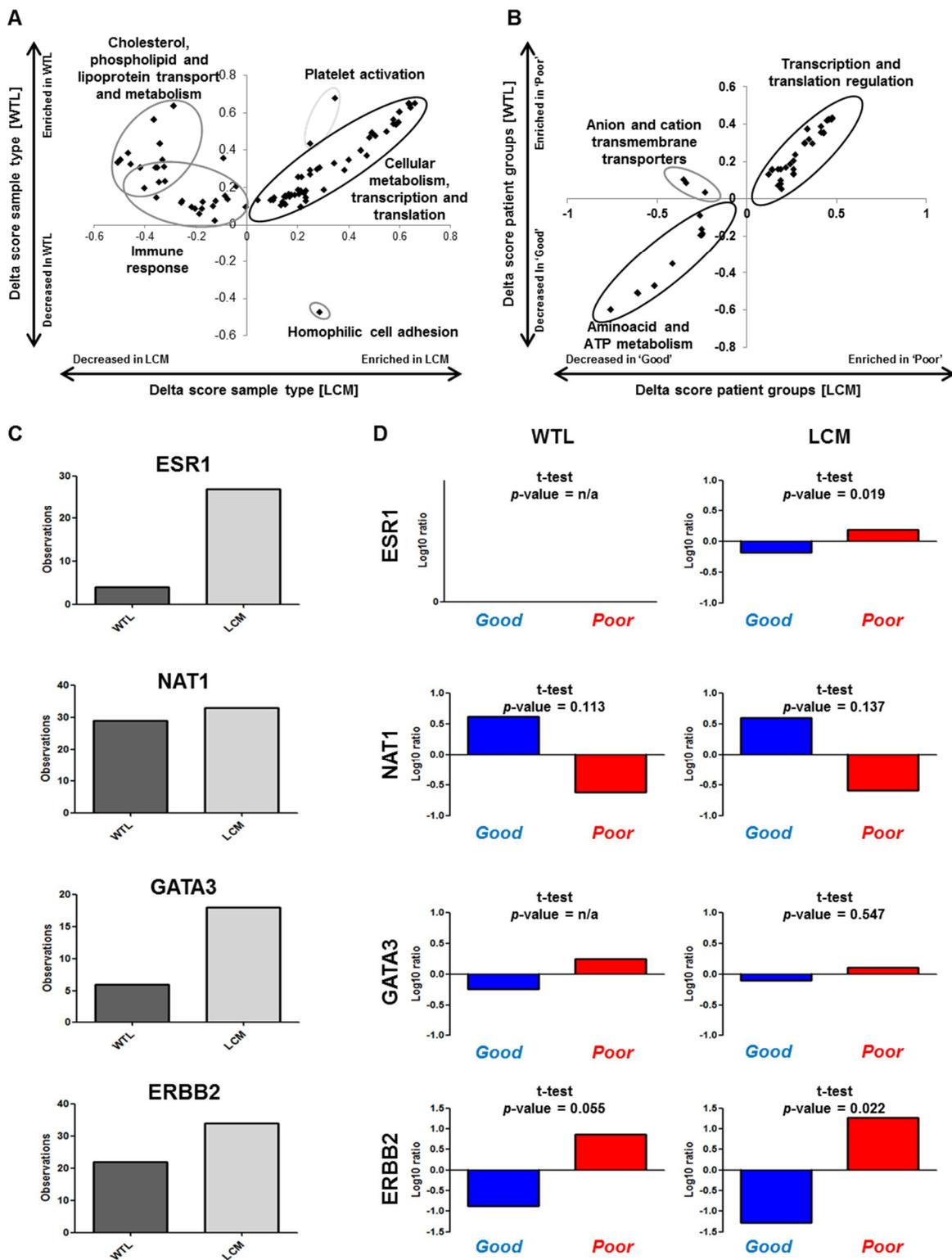
To assess differences in expression levels between shared ECM and plasma proteins between the two sets, hierarchical clustering was performed on a total of 62 ECM and 34 plasma proteins (aligned search) after filtering for missing data (i.e. < 30%). Among the ECM proteins, proteins specific to the basal layer (e.g. PRELP) and serum transporters (e.g. ITIH1) displayed a higher abundance in WTL compared to LCM, while proteins co-localized at the plasma membrane (e.g. HSDL2) showed higher abundance in LCM compared to WTL. A small set of proteins characterized by multiple subcellular localizations (e.g. ISOC1) did not display enrichment in either set (Fig.S10A). These abundance differences were confirmed at the peptide level (Fig. S10B). A similar pattern was observed for plasma proteins: immunoglobulins (e.g. IGHM) and coagulation-related proteins (e.g. FGA) showed higher abundance in WTL compared to LCM samples, while proteins also belonging to cell cytoplasm and plasma membrane (e.g. KRT19) displayed higher abundance in LCM compared to WTL samples. A subset comprising proteins localized multiple cellular and extracellular compartments (e.g. FN1) did not display enrichment in any sample set (Fig. S10C-D). These data indicate a specific enrichment in ECM and plasma protein subsets in each sample set, with intracellular proteins enriched in LCM.

#### *LCM allows more precise quantitation of the tumor proteome*

In order to assess whether certain protein groups were enriched in either the WTL or the LCM set, 2D analysis was performed on the 2,696 shared proteins after aligned search (Fig. 3.5A, Table S3). Proteins involved in cellular metabolism, transcription and translation showed expression concordance in both sets, while proteins involved in immune response, and lipid and cholesterol metabolisms displayed anti-correlating abundances and were found highly enriched in the WTL samples compared to the LCM samples, confirming hierarchical clustering analysis results. Platelet activation-related proteins were highly expressed in both sets but with a higher expression in WTL. On the other hand, homophilic cell adhesion-involved proteins were found highly expressed in LCM and anti-correlated to WTL samples.

In order to assess whether sample type would influence findings related to tamoxifen resistance, we investigated differentially expressed proteins between poor and good outcome patients in the aligned WTL and LCM datasets. A total of 542 and 374 proteins displayed differential expression between patient groups in the LCM and WTL sets, respectively, with 135 proteins found in both sets. Of these, 119 proteins showed identical expression direction in both sets (mean Pearson  $r = 0.63$ ), while 16 showed discordant expression (mean Pearson  $r = -0.14$ ; Table S4). Concordant

proteins were predominantly involved in transcription, translation, and in aminoacid and ATP metabolism. 2D analysis showed that ion transmembrane transporters displayed discordant expression between WTL and LCM (Fig. 3.5B, Table S5), being enriched in the former and not in the latter.



### **Figure 3.5. Analysis of differences in protein quantitation between WTL and LCM.**

Panel A displays 2D pathway analysis of shared proteins between LCM and WTL samples after aligned search: while proteins involved in platelet activation were enriched in WTL samples (light grey), cholesterol, lipid metabolism, immune response, and homophilic cell adhesion related proteins displayed opposite expression between WTL and LCM samples (grey). Transcription, translation and cellular metabolism involved proteins showed concordance of expression between LCM and WTL samples (black). Panel B shows 2D analysis for shared differentially expressed proteins: concordance of expression in transcription, translation and metabolism related proteins (black), while opposite expression was observed for ion transporters (grey). Panels C and D display the number of observations and differential expression between poor and good outcome patients of key breast cancer markers (ESR1, NAT1, GATA3, and ERBB2), respectively.

Acronyms: GO: gene ontology

We then sought to evaluate the identification rate and differences in expression levels of key breast cancer proteins in WTL and LCM samples. For ESR1 (Pearson  $r = 0.78$ ), NAT1 (Pearson  $r = 0.97$ ), GATA3 (Pearson  $r = n/a$  due to missing paired data points), and ERBB2 (Pearson  $r = 0.79$ ) a lower number of observations was found in the WTL set when compared to LCM (Fig. 3.5C-D). A significant difference was found between the levels of ESR1 ( $p\text{-value} = 0.019$ ) and ERBB2 ( $p\text{-value} = 0.022$ ) when comparing poor and good outcome patients in the LCM set, while no significance was observed in the WTL group. These data suggest that proteomic analysis of LCM-derived material results in more accurate quantitation of tumor proteins.

## **Discussion**

LCM based cell enrichment has been successfully coupled to high resolution MS and LFQ in the analysis of morphologically heterogeneous tissues such as colon and breast carcinomas, allowing assessment of thousands of proteins from pure cell populations [23]. Despite this, proteomic analysis of whole tissue lysates is often preferred to cell enrichment techniques due to the possibility of using chemical labeling techniques and relatively less laborious sample preparation. In the current study we used data of a set of 38 ER positive breast cancer tissues, from which LCM and WTL fractions were previously analyzed by high resolution MS, to determine the effect of sample preparation on data quality.

We have here shown how proteomic analysis of LCM enriched specimens not only improved peptide and protein quantification when compared to WTL samples, but also more clearly defined tumor protein expression. On the MS level, a significantly higher amount of peaks were repeatedly sequenced in WTL samples, which suggested a more complex matrix and interference at the LC

level. This would probably cause repeated triggering of MS/MS events, resulting in decreased peptide and protein identifications and increased missing data points. The lower identification rates in the WTL set - an aspect that has been extensively assessed in the analysis of plasma proteins [27,28] - could be due to interference from highly abundant proteins or blood contaminants present in tumor surrounding tissues that would alter the dynamic range and decrease chromatographic resolution in the detection of low abundance proteins. Furthermore, while alignment did not significantly increase identification rates in the LCM set, the “match-between-runs” feature greatly improved the total number of identified peptides and proteins in WTL specimens.

ECM and plasma proteins were found enriched in WTL compared to LCM samples, as demonstrated by subcellular localized protein distribution analysis. Due to the fact that surrounding tissues are co-analyzed with tumor epithelial cells, a relatively higher abundance of stromal and blood-derived proteins is expected in such samples. This assumption was further confirmed by 2D analysis of shared proteins between the two datasets after aligned search, which resulted in anti-correlating abundances of proteins belonging not only to the innate and adaptive immune systems and to lipid and cholesterol metabolism. In the light of these findings, LCM based enrichment provides a higher level of depth for downstream proteomic analysis due to reduced matrix complexity [13], resulting in higher identification rates and a lower number of missing values. We therefore argue that LCM enrichment of tumor cells prior to proteome analysis would improve success rates of identifying tumor-related biomarkers.

When assessing key breast cancer markers, a discrepancy between LCM and WTL sets was revealed: not only ESR1, GATA3 and ERBB2 proteins were observed in a relatively higher percentage of samples in the LCM set, but a significant difference was also observed between poor and good outcome patients in this set for clinically relevant proteins such as ESR1 and ERBB2 [29]. This remarkable finding implies that key markers of tamoxifen resistance previously identified in LCM samples [19] would have likely been missed if WTL samples had been used for analysis. This phenomenon likely also holds true for many other heterogeneous tissues being investigated for biomarker discovery.

### **Concluding remarks**

We have here presented a comparison between proteomic analyses of whole tissue and tumor enriched specimens. Tumor enrichment by LCM coupled to high resolution MS constitutes a robust approach capable of detecting a higher amount of proteins compared to whole tissue lysate analysis.

In this perspective LCM offers a more reliable tool to delve into the tumor proteome with much more sensitivity when compared to WTL analysis. Furthermore, as stromal, epithelial, and immunologic factors possess a specific role in cancer [30], their comprehensive analysis would be more likely enabled by LCM based enrichment and comparative analysis of these segregated components rather than whole tissue analysis.

## References:

- [1] Wiśniewski, J.R., Ostasiewicz, P., Mann, M., High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J. Proteome Res.* 2011, 10, 3040–9.
- [2] Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., et al., A draft map of the human proteome. *Nature* 2014, 509, 575–81.
- [3] Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., et al., Mass-spectrometry-based draft of the human proteome. *Nature* 2014, 509, 582–7.
- [4] Zhang, B., Wang, J., Wang, X., Zhu, J., et al., Proteogenomic characterization of human colon and rectal cancer. *Nature* 2014, 513, 382–387.
- [5] Jr., D.J.J., Rodriguez-Canales, J., Mukherjee, S., Prieto, D.A., et al., Approaching solid tumor heterogeneity on a cellular basis by tissue proteomics using laser capture microdissection and biological mass spectrometry. *J. Proteome Res.* 2010, 8, 2310–18.
- [6] Longo, D., Tumor Heterogeneity and Personalized Medicine. *NEJM* 2012, 366, 956–7.
- [7] Kondo, T., Inconvenient truth: cancer biomarker development by using proteomics. *Biochim. Biophys. Acta* 2014, 1844, 861–5.
- [8] Emmert-buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., et al., Laser Capture Microdissection. *Science* (80-. ). 1996, 274, 8–11.
- [9] Vogel, A., Lorenz, K., Horneffer, V., Hüttmann, G., et al., Mechanisms of laser-induced dissection and transport of histologic specimens. *Biophys. J.* 2007, 93, 4481–500.
- [10] Yang, F., Foekens, J. a, Yu, J., Sieuwerts, a M., et al., Laser microdissection and microarray analysis of breast tumors reveal ER-alpha related genes and pathways. *Oncogene* 2006, 25, 1413–9.
- [11] Umar, A., Luider, T.M., Foekens, J.A., Pasa-Tolić, L., NanoLC-FT-ICR MS improves proteome coverage attainable for approximately 3000 laser-microdissected breast carcinoma cells. *Proteomics* 2007, 7, 323–9.
- [12] Braakman, R.B.H., Tilanus-Linthorst, M.M.A., Liu, N.Q., Stingl, C., et al., Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. *J. Proteomics* 2012, 75, 2844–54.
- [13] Liu, N.Q., Braakman, R.B.H., Stingl, C., Luider, T.M., et al., Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J. Mammary Gland Biol. Neoplasia* 2012, 17, 155–64.
- [14] Zhu, W., Smith, J.W., Huang, C.M., Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* 2010, 2010.
- [15] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 2008, 26, 1367–72.

- [16] Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. a, et al., Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 2011, 10, 1794–805.
- [17] Cox, J., Hein, M.Y., Lubner, C. a., Paron, I., et al., Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* 2014, 13, 2513–2526.
- [18] Liu, N.Q., Stingl, C., Look, M.P., Smid, M., et al., Comparative Proteome Analysis Revealing an 11-Protein Signature for Aggressive Triple-Negative Breast Cancer. *J. Natl. Cancer Inst.* 2014, 106, Epub 2014.
- [19] De Marchi, T., Liu, N.Q., Stingl, C., Timmermans, M.A., et al., 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Mol. Oncol.* 2016, 10, 24–39.
- [20] De Marchi, T., Liu, N.Q., Sting, C., Smid, M., et al., Global proteomic characterization of microdissected estrogen receptor positive breast tumors. *Data Br.* 2015, 5, 399–402.
- [21] Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J. a, et al., The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013, 41, D1063–9.
- [22] Ashburner, M., Ball, C. a, Blake, J. a, Botstein, D., et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 2000, 25, 25–29.
- [23] Apweiler, R., The universal protein resource (UniProt) in 2010. *Nucleic Acids Res.* 2009, 38, 142–148.
- [24] de Hoon, M.J.L., Imoto, S., Nolan, J., Miyano, S., Open source clustering software. *Bioinformatics* 2004, 20, 1453–4.
- [25] Cox, J., Mann, M., 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* 2012, 13, S12.
- [26] Sandin, M., Chawade, A., Levander, F., Is label-free LC-MS/MS ready for biomarker discovery? *Proteomics - Clin. Appl.* 2015, 9, 289–94.
- [27] Anderson, N.L., Anderson, N.G., The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 2002, 1, 845–67.
- [28] Anderson, N.L., Polanski, M., Pieper, R., Gatlin, T., et al., The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* 2004, 3, 311–326.
- [29] Burki, T.K., Adjuvant treatment for HER2-positive breast cancer. *Lancet. Oncol.* 2015, 16, e59.
- [30] Hanahan, D., Weinberg, R.A., Hallmarks of cancer: the next generation. *Cell* 2011, 144, 646–74.

## **Chapter 4**

### **4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer**

Tommaso De Marchi, Ning Qing. Liu, Cristoph Stingl, Mieke A. Timmermans, Marcel Smid, Maxime P. Look, Mila Tjoa, Rene B. Braakman, Mark Opdam, Sabine C. Linn, Fred C. Sweep, Paul N. Span, Mike Kliffen, Theo M. Luiders, John A. Foekens, John W. Martens, Arzu Umar.

*Mol Oncol* 2016; 10, 1: 24-39.

## Abstract

Estrogen receptor (ER) positive tumors represent the majority of breast malignancies, and are effectively treated with hormonal therapies, such as tamoxifen. However, in the recurrent disease resistance to tamoxifen therapy is common and a major cause of death. In recent years, in-depth proteome analyses have enabled identification of clinically useful biomarkers, particularly, when heterogeneity in complex tumor tissue was reduced using laser capture microdissection (LCM). In the current study, we performed high resolution proteomic analysis on two cohorts of ER positive breast tumors derived from patients who either manifested good or poor outcome to tamoxifen treatment upon recurrence. A total of 112 fresh frozen tumors were collected from multiple medical centers and divided into two sets: an in-house training and a multi-center test set. Epithelial tumor cells were enriched with LCM and analyzed by nano-LC Orbitrap mass spectrometry (MS), which yielded > 3,000 and > 4,000 quantified proteins in the training and test sets, respectively. Raw data are available via ProteomeXchange with identifiers PXD000484 and PXD000485. Statistical analysis showed differential abundance of 99 proteins, of which a subset of 4 proteins was selected through a multivariate step-down to develop a predictor for tamoxifen treatment outcome. The 4-protein signature significantly predicted poor outcome patients in the test set, independent of predictive histopathological characteristics (hazard ratio [HR] = 2.17; 95% confidence interval [CI] = 1.15 to 4.17; multivariate Cox regression *p value* = 0.017). Immunohistochemical (IHC) staining of PDCD4, one of the signature proteins, on an independent set of formalin-fixed paraffin-embedded tumor tissues provided independent technical validation (HR = 0.72; 95% CI = 0.57 to 0.92; multivariate Cox regression *p value* = 0.009). We hereby report the first validated protein predictor for tamoxifen treatment outcome in recurrent ER-positive breast cancer. IHC further showed that PDCD4 is an independent marker.

## Introduction

ER positive tumors constitute the majority of all breast malignancies. Tamoxifen therapy has been shown to significantly improve survival and cure of patients with primary ER positive breast tumors, but upon recurrence about half of the patients show intrinsic resistance, while those initially responding will ultimately develop acquired resistance (Cardoso et al., 2012; Milani, 2014). The need for biomarkers capable of determining mechanisms of resistance has led to the development of several predictive signatures, though none has been introduced in the clinic so far (Beelen et al., 2012). With the recent advancements in MS techniques, in-depth quantification of the human proteome has become possible and the ability of measuring protein abundance over a broad dynamic range has established proteomics as a robust tool for biomarker discovery (Drabovich et al., 2014; Kim et al., 2014; Wilhelm et al., 2014). The proteomic analysis of tissue specimens is, however, hindered by their heterogeneity, which alters protein abundance dynamic range. Furthermore, the presence of stromal and infiltrating cells adds another layer of complexity by hampering accurate protein quantitation of target epithelial tumor cells (Kondo, 2014). To address this issue, LCM offers a robust cell sub-population enrichment technique, allowing accurate downstream analysis of morphologically heterogeneous specimens (Emmert-buck et al., 1996; Vogel et al., 2007). Genomic and proteomic analyses of LCM derived material showed the feasibility of this technique in molecular profiling studies and pointed out its efficacy in studying disease associated signaling pathways when compared to whole tissue analyses (Cheng and Zhang, 2013; Sereni et al., 2015; Xu, 2010; F. Yang et al., 2006). LCM yields sub-microgram protein amounts due to the fact that only a limited number of cells can be dissected from each sample. In the light of this, coupling LCM enrichment to chemical labeling methods would require extensive sample preparation and workflow optimization, which would be unsuitable in the analysis of large sample sets. Label-free quantification (LFQ) software algorithms have demonstrated to be accurate tools in the quantitation of proteins, allowing high yield identification and reliable quantitation of measured peptides even from minute amount of analyzed specimens (Cox and Mann, 2008; Megger et al., 2013). We have optimized a tissue proteomic pipeline for biomarker discovery coupling LCM cell enrichment to high resolution LC-MS and LFQ, capable of quantifying more than 3,000 proteins from only 4,000 dissected epithelial cells (Braakman et al., 2012; Liu et al., 2012). Using this workflow, we recently developed and validated a prognostic protein signature for triple negative breast cancer (Liu et al., 2014). Despite our workflow has demonstrated to be a robust methodology for the discovery of cancer biomarkers, application of shotgun proteomics in clinical diagnostics remains problematic due to the extensive and time consuming sample preparation required. In this perspective, IHC or selected reaction monitoring/multiple reaction monitoring

(SRM/MRM) MS may be more suitable biomarker verification techniques that do not require extensive method optimization or sample preparation (Whiteaker et al., 2011). Although antibody specificity and lack of accurate quantitation remain important issues, IHC still remains a major technique in clinical diagnostics and significantly requires less amount of optimization time in comparison to ELISA or even SRM/MRM MS.

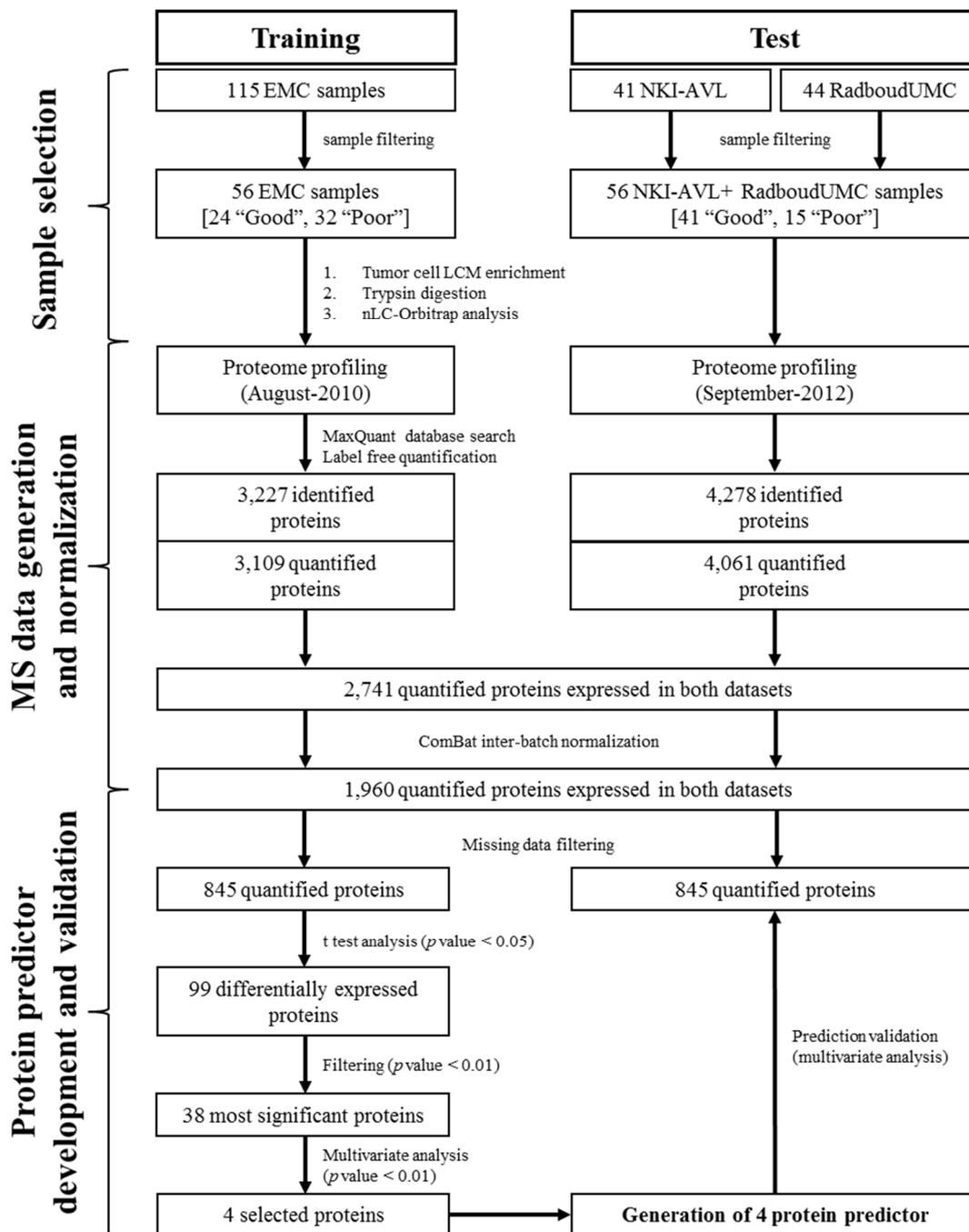
In this study we describe the development of a predictive protein signature for tamoxifen resistance in ER positive breast cancer by coupling LCM tumor cell enrichment and high resolution LC-MS in the analysis of independent training and test patient cohorts. We also provide further validation by IHC analysis of signature proteins on an independent panel of paraffin-embedded tissues captured in a tissue micro-array (TMA).

## **Materials and Methods**

### *Sample sets*

From an initial selection of 200 tissues collected from patients that received tamoxifen as first line therapy we excluded tissues with a low percentage of tumor cells (i.e. < 40%; n = 88; Figure 4.1). A total of 112 ER positive fresh frozen primary breast cancer tissue samples were then included in our sets: 56 from Erasmus MC University Medical Center (EMC), Rotterdam (years of surgery: 1981-1994), 41 from the National Cancer Institute – Antoni van Leeuwenhoek hospital (NKI-AVL), Amsterdam (1980-1996), and 15 from Radboud University Medical Center (RadboudUMC), Nijmegen (1991-1996; Table 4.1). EMC derived samples constituted the training set, while NKI-AVL and RadboudUMC provided an independent external test set. ER positivity in tumor cytosols was assessed by quantitative biochemical assays (EMC), reverse-transcriptase quantitative polymerase chain reaction (RadboudUMC), or IHC (NKI-AVL). All patients underwent surgery of their primary tumor (conservative or non-conservative), developed recurrent disease, and were treated with tamoxifen as first line therapy. Due to lack of response data for a subset of specimens, treatment outcome was defined based on time to progression (TTP): disease progression  $\leq$  6 months and  $>$  6 months after start of first line tamoxifen administration were defined as poor and good outcome, respectively. The training set comprised 24 and 32 patients who showed good and poor outcome upon tamoxifen treatment, respectively. The test set included tumors of 41 good and 15 poor outcome patients. The NKI-AVL cohort did not contain stage IV tumors, while such specimens were found in the EMC and RadboudUMC sets. In addition, 2 tumor tissues of which clinical follow up information was not available were used as LCM and whole tissue lysate (WTL)

controls. For biological replicates, both tumor tissues were subjected to 4 rounds of LCM. Of one of these, a WTL was prepared from one sample and digested in triplicate.



**Figure 4.1. Data analysis flow-chart and development of predictor for tamoxifen treatment outcome.**

Patients were divided into two independent cohorts and separately measured by LC-MS. Proteomic data from training and test sets were analyzed separately in MaxQuant. Identified proteins were filtered for reversed sequences and for Posterior Error Probability score (PEP < 0.05), intensities of commonly expressed proteins were normalized using ComBat algorithm to minimize batch effects, and filtered for missing data (10 minimum observations for global proteomic analysis and allowing 30% and 0% missing data in training and test set respectively for predictor generation).

Student t test ( $p$  value < 0.05) was then used to assess differences in protein expression levels between good and poor outcome patients. A multivariate regression model was used to obtain an optimal list of 4 proteins to be tested as a predictor of tamoxifen treatment outcome: CGN, G3BP2, PDCD4 and OCIAD1. The 4-protein signature was confirmed in an external test set.

**Table 4.1. Patient and tumor characteristics.**

	Training*		Test*	
	EMC	NKI-AVL	RadboudUMC	
All patients	56 (100)	41 (100)	15 (100)	
Age				
≤ 55 years	15 (27)	12 (29)	1 (7)	
> 55 years	41 (73)	29 (71)	14 (93)	
Menopausal status				
Premenopausal	10 (18)	11 (27)	0 (0)	
Postmenopausal	46 (82)	30 (63)	15 (100)	
Tumor size				
T1 (≤ 2cm)	12 (21)	20 (49)	5 (33)	
T2 (2-5cm) + Tx	40 (72)	19 (46)	9 (60)	
T3 (> 5cm) + T4	4 (7)	2 (5)	1 (7)	
Tumor differentiation**				
Good/Moderate	13 (59)	29 (71)	8 (53)	
Poor	33 (23)	12 (29)	4 (27)	
Unknown	10 (18)	0 (0)	3 (20)	
Disease free interval				
≤ 12 months	24 (43)	4 (10)	5 (33)	
> 12 months	32 (57)	37 (90)	10 (67)	
PgR†				
Negative	9 (16)	17 (41)	11 (73)	
Positive	44 (79)	24 (59)	4 (27)	
Involved lymph nodes				
0	31 (55)	24 (58)	6 (40)	
≥ 1	20 (36)	16 (39)	7 (47)	
unknown	5 (9)	1 (3)	2 (13)	
Dominant site of relapse				
Loco-regional	8 (14)	4 (10)	0 (0)	
Bone	26 (46)	12 (29)	6 (40)	
Visceral	13 (24)	6 (15)	9 (60)	
Bone and other	9 (16)	14 (34)	0 (0)	
Unknown	0 (0)	5 (12)	0 (0)	

\* Data are reported as number (percentage).

\*\* Histopathological characteristics were evaluated by local pathologists, according to standard clinical practice at time of sample collection.

† Missing data not reported.

In addition, a total of 447 formalin-fixed and paraffin-embedded tissues collected from EMC and regional hospitals were comprised in a tissue micro-array (Supplemental Material). For further analyses, we included only ER positive tumors and patients who did not receive hormonal adjuvant therapy. Patients with a revised histology that showed no tumor, or patients with a progression within 3 weeks were excluded as well, leading to a total of 408 ER positive tissues from patients treated with tamoxifen as first-line therapy for recurrent disease. Response data were collected according to the standard International Union Against Cancer criteria (Hayward and Carbone, 1977). In this set, 11 (2.7%) and 51 (12.5%) patients respectively showed complete (CR) and partial remission (PR). Two hundred and five (50.3%) patients showed no change (NC) of disease, of whom 170 (41.7%) showed NC > 6 months (defined as stable disease, SD) while 35 (8.6%) showed NC ≤ 6 months after start of therapy. Progressive disease (PD) was observed in 141 (34.6%) patients. Clinical benefit was defined as CR + PR + SD patients (n = 232; 57%), while objective response was defined as CR + PR only (n = 62; 15%). This retrospective study used coded primary tumor tissues, in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (<http://www.federa.org/codes-conduct>). Reporting Recommendations for Tumor Marker Prognostic Studies were followed where possible (Altman et al., 2012).

#### *Laser capture microdissection*

All tissue samples were cut into 8 µm cryo-sections, and collected on UV-sterilized polyethylene naphthalate (PEN) coated glass slides (Carl Zeiss Microsystems GmbH, Göttingen, Germany) for downstream laser capture microdissection. In addition, 5 µm sections were collected on regular glass slides and stained with hematoxylin and eosin dyes for histological evaluation. Sections on PEN slides were dehydrated with 95% ethanol and immediately stored at -80°C, until further processing. Prior to LCM, PEN slides were thawed at room temperature and subsequently stained with hematoxylin as follows: distilled water, hematoxylin, distilled water, 50% ethanol, 70% ethanol, 95% ethanol, 100% ethanol, 100% ethanol. During dehydration steps Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific Inc, Rockford, IL, USA) at a 1:100 v/v concentration was added in order to prevent proteolytic degradation of proteins. An area of approximately 500,000 µm<sup>2</sup> (~ 4,000 tumor cells) was collected from each tissue using a photo-activated localization microscopy Micro Beam device and gathered in an opaque adhesive cap (Carl Zeiss Microsystems GmbH, Göttingen, Germany). A volume of 20 µl of 0.1% w/v Rapigest surfactant (Waters Corporation, Milford, MA, USA) in 50 mM ammonium bicarbonate solution was used to

transfer the collected LCM samples into LoBind™ Eppendorf tubes (Eppendorf AG, Hamburg, Germany). Tissue containing buffer was immediately frozen after collection and stored at -80°C.

### *Protein digestion*

LCM collected material was disrupted in a horn sonifier bath using an Ultrasonic Disruptor Sonifier II (Bransons Ultrasonics, Danbury, CT, USA) at 70% amplitude. Proteins were denatured at 95°C, reduced with 100 mM DTT for 30 min at room temperature, and alkylated in the dark with 300 mM iodoacetamide for 30 min at room temperature. Samples were then digested for 4 h at 37°C after addition of MS grade trypsin at a 1:4 enzyme-protein ratio (i.e. 100 ng/μl). Samples were acidified with TFA, and spun down at 14,000 RPM. Supernatants were collected and transferred to HPLC vials (Sigma-Aldrich Corporation, St. Louis, MO, USA).

### *High resolution MS*

Mass spectrometry measurements were performed with a nano liquid chromatography system (Ultimate 3000, Dionex, Amsterdam, The Netherlands) coupled online to a linear Ion Trap – Orbitrap XL™ mass spectrometer (Thermo Electron, Bremen, Germany). Samples were first loaded on a trap column (PepMap C18, 300 μm ID × 5 mm length, 5 μm particle size, 100 Å pore size; Dionex), then washed and desalted in 0.1% TFA acidified water. Trap column and analytical column (PepMap C18, 75 μm ID × 50 cm, 3 μm particle size and 100 Å pore size; Dionex) were then coupled and peptides were eluted in a 3 h binary gradient (flow: 300 nl/min; mobile phase A: 2% acetonitrile and 0.1% formic acid in H<sub>2</sub>O; mobile phase B: 80% acetonitrile and 0.08% formic acid). Gradient was run as follows: 0% to 25% mobile phase B for 2 h, increase to 50% mobile phase B in 1 h. For ESI, metal-coated nano ESI emitters (New Objective, Woburn, MA) were used and a spray voltage of 1.6 kV was applied. High-resolution scan was acquired from 400 to 1,800 Th and was used for MS detection. Automatic gain was set at 106 ions and lock mass was set at 445.120025 u protonated with (Si(CH<sub>3</sub>)<sub>2</sub>O)<sub>6</sub>. The 5 most intense peaks in full scan were selected and fragmented by collision induced dissociation (CID) applying 35% normalized collision energy and detected in the ion trap. Ions falling out of the ±5 ppm window or for which precursor intensity fell below 1.5 signal-to-noise ratio during 10 scans were excluded.

### *Protein identification and quantification*

A total of 112 samples were analyzed by LTQ-Orbitrap XL™ MS, together with 4 biological LCM replicates of control samples, and of which one was measured with a triplicate of its matching WTL. MS spectra of the training and test cohorts were generated and analyzed separately with a time interval of two years. Orbitrap .RAW files derived from MS analyses were imported and analyzed in MaxQuant (version 1.2.2.5) (Cox and Mann, 2008), using Andromeda peptide search engine (Cox et al., 2011). Analysis of spectra was performed using the following options: acetylation of the N-terminus and oxidation of methionine were selected as variable modifications, multiplicity was set to 1. FASTA file used for protein search was UniProt-SwissProt human canonical database (version 2012-09, human canonical proteome; 20.243 identifiers). Minimal peptide length was set to 7 amino acids, match between runs and LFQ options were selected and kept as default. Other options were kept as default (e.g. fixed peptide modifications: carbamidomethylation; false discovery rate = 0.01). For further data analysis, “ProteinGroups.txt” file was imported into Microsoft Excel and protein identifiers were filtered based on posterior error probability score (cutoff < 0.05). Contaminants and reversed sequences were excluded. LFQ intensities for each sample were selected and each value was Log10 transformed. Protein intensities from training and test sets were then normalized using ComBat (Johnson et al., 2007) algorithm in R free software, allowing 10 minimum observations for whole dataset analysis. A second protein list was generated allowing 30% missing data points in the training set and none in the test set for predictor development. LCM and WTL control samples were not included in the ComBat normalization procedure due to the lower amount of identified and quantified proteins. Coefficients of variations of Log10 transformed MS data were calculated according to the following formula:

$$CV=10^{(\text{Standard deviation})} - 1 \text{ (Bland and Altman, 1996).}$$

Pearson correlation coefficients between measurements of LCM and WTL replicates were calculated in Perseus (Max Planck Institute for Biochemistry, Muenchen, Germany). The MS proteomic data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (Vizcaíno et al., 2013) with dataset identifiers PXD000484 and PXD000485.

### *Tissue micro-array*

TMA was prepared using an ATA 27 (Beecher Instruments, Sun Prairie, WI, USA). 408 paraffin-embedded primary, ER positive breast cancer tissues derived from patients treated with first line tamoxifen upon recurrence were used to prepare the TMA. Tissue cores of 0.6 mm were taken from each tissue paraffin block and transferred in triplicate into a TMA recipient block. For each tumor tissue sample, three different areas of the tumor were taken as biological replicates. TMA slides were digitalized and analyzed using Slidepath software (Leica Microsystems, Solms, Germany).

### *Immunohistochemistry*

Paraffin-embedded tissues on glass slides were de-paraffinized at 60°C, and remnants of paraffin were removed by sequential washings in xylene (3 x 5 min). Re-hydration was performed by washings through decreasing concentrations of ethanol following with distilled water as follows: 100% ethanol (1 x 5 min, 2 x 2 min), 70% ethanol (1 x 2 min), 50% ethanol (1 x 2 min), distilled water (1 x 2 min). Slides were then incubated at 95°C for 40 min in DAKO (Agilent Technologies Inc, Santa Clara, CA, USA) antigen retrieval solution (pH 6) diluted 1:10 in MilliQ water, cooled down to room temperature and washed with PBS buffer 3 times for 5 min. Blocking solution consisting of 5% BSA in PBS was added to the slides and incubated for 30 min. Primary antibodies were diluted in DAKO Antibody Diluent, added to each slide and incubated for 1 h at room temperature. Slides were then washed with PBS, and DAKO Envision® secondary antibody (Goat anti-Mouse-HRP and G anti-R-HRP, 100 µl per slide) solution was added to each slide and incubated for 45 min at room temperature. A washing cycle with PBS was performed for 5 min and a 1:15 solution of DAB+ chromogen in antibody diluent was added, following incubation in the dark for 10 min. Slides were then washed in tap water for 5 min, stained with hematoxylin/eosin for 1 min each and dehydrated again through sequential washings in 50%-70%-100% ethanol and xylene of 5 min each. Cover glasses were mounted with Pertex and slides were left to dry. TMA slides were stained for Programmed Cell Death 4 (PDCD4) protein (1:200), OCIAD1 (1:800), G3BP2 (1:50), and CGN (1:25). Anti-PDCD4 mouse monoclonal (id: LS-B2949; clone K4C1) and anti-OCIAD1 rabbit polyclonal (id: LS-B5046) antibodies were purchased from Lifespan Technologies (Lifespan technologies Inc, Seattle, WA, USA), anti-G3BP2 rabbit polyclonal (id: NBP1-82976) antibody was purchased from Novus Biologicals (Novus Biologicals LLC, 8100 Littleton, CO, USA), and anti-CGN rabbit polyclonal (HPA027657) antibody was purchased from Sigma.

### *IHC staining analysis*

Data from scored tissues were filtered for missing data and adjuvant endocrine therapy, leading to a final list of 294 tissue samples. PDCD4 antibody stained tissues were separately scored for nuclear and cytoplasmic staining intensity (categories: negative, weak, moderate, strong) and percentage of stained tumor cells (categories: 0%, 1-10%, 11-20%, 21-30%, 31-40%, 41-50%, 51-60%, 71-80%, 81-90%, 91-100%). CGN, and OCIAD1 stained tissues were scored based on intensity parameters only, while G3BP2 scoring included quantity levels as well. TMA was scored by two independent researchers, and the average, consolidated scores of triplicate cores were used for statistical analysis. Due to the fact that PDCD4 cytoplasmic and nuclear stainings were co-expressed in the evaluated TMA cores, these were merged in order to assess total protein levels. PDCD4 nuclear and cytoplasmic scores were numerically transformed and merged into a histo-score (Supplemental Material) according to formula:

Histo-score = (nuclear quantity x nuclear intensity) + cytoplasmic intensity

Histo-score cutoff (i.e. 30) reflective of weak vs. strong protein expression was used to stratify patient groups: low and high PDCD4 protein expressing tumors displayed a histo-score below (<) or above ( $\geq$ ) the cutoff (Supplemental Material). PDCD4 cytoplasmic quantity was ranging only from 80% to 90% so it was not included in the histo-score calculation formula.

### *Statistical analysis*

Differences in clinical parameters between training and test sets were evaluated by Mann-Whitney U and Pearson  $\chi^2$  tests (two sided tests). Commonly expressed proteins between the two LCM sets and proteins quantified in WTL sample replicates were annotated through DAVID (Huang et al., 2009a, 2009b) for organelle distribution using Swissprot keyword database. Average abundance levels of these proteins in all 112 measured samples were used to generate a waterfall plot of protein abundance distribution.

Protein list used for predictor development was tested for protein differential abundance between patient groups through Student's t-test (two sided, unequal variances assumed). Hierarchical clustering was performed on all quantified and differentially expressed proteins (t test *p value* < 0.05), respectively (complete linkage; distance metric: correlation-uncentered). Significant proteins in the training set were submitted along with their fold changes and t test *p values* to Ingenuity Pathway Analysis (IPA) network analysis with the following settings: Data sources: all;

Confidence: high (predicted) and experimentally observed; species: human. Network was plotted using path designer (Ingenuity Systems, Redwood City, CA, USA).

In order to rule out possible indiscriminative identifiers, the protein predictor was developed selecting the 38 most significant proteins (univariate  $p$  value  $< 0.01$ ) in the training set and a Cox regression multivariate analysis was performed with a step-down procedure, which involved iteratively removing the least significant proteins (multivariate  $p$  value  $\geq 0.01$ ) until all remaining proteins in the model showed a multivariate  $p$  value  $< 0.01$ . Each protein score (t value) was then multiplied by its abundance, and values were then summed for all proteins to obtain a patient score, which was then coupled to outcome data. Each patient score was plotted in a receiver operating characteristic (ROC) curve. Youden index (max of  $J = \text{Sensitivity} + \text{Specificity} - 1$ ) was set as cutoff in the training set and used to categorize patients in the test set. Log-rank tests on the survival curves of predicted groups were performed to assess significance of prediction. Association of predictor proteins to TTP was assessed through Cox regression, correcting for patient and tumor characteristics. IHC stainings were used to test for association with TTP, clinical benefit and objective response in combination with clinical parameters by Cox and logistic regression analyses, respectively. Co-variables that were found not significant in univariate regression analyses were excluded from multivariate models. Cox regression and logistic regression analyses, hazard ratios, odds ratios and confidence intervals were calculated in Stata (version 13.1; Stata Corp, College Station, TX, USA).

## Results

### *Analysis of patient cohorts*

One hundred and twelve ER positive primary breast tumor tissues, of which 56 comprised the training set and another 56 the test set, were processed according to our tissue proteomics workflow (Braakman et al., 2012; Liu et al., 2012) and analyzed through high resolution MS.

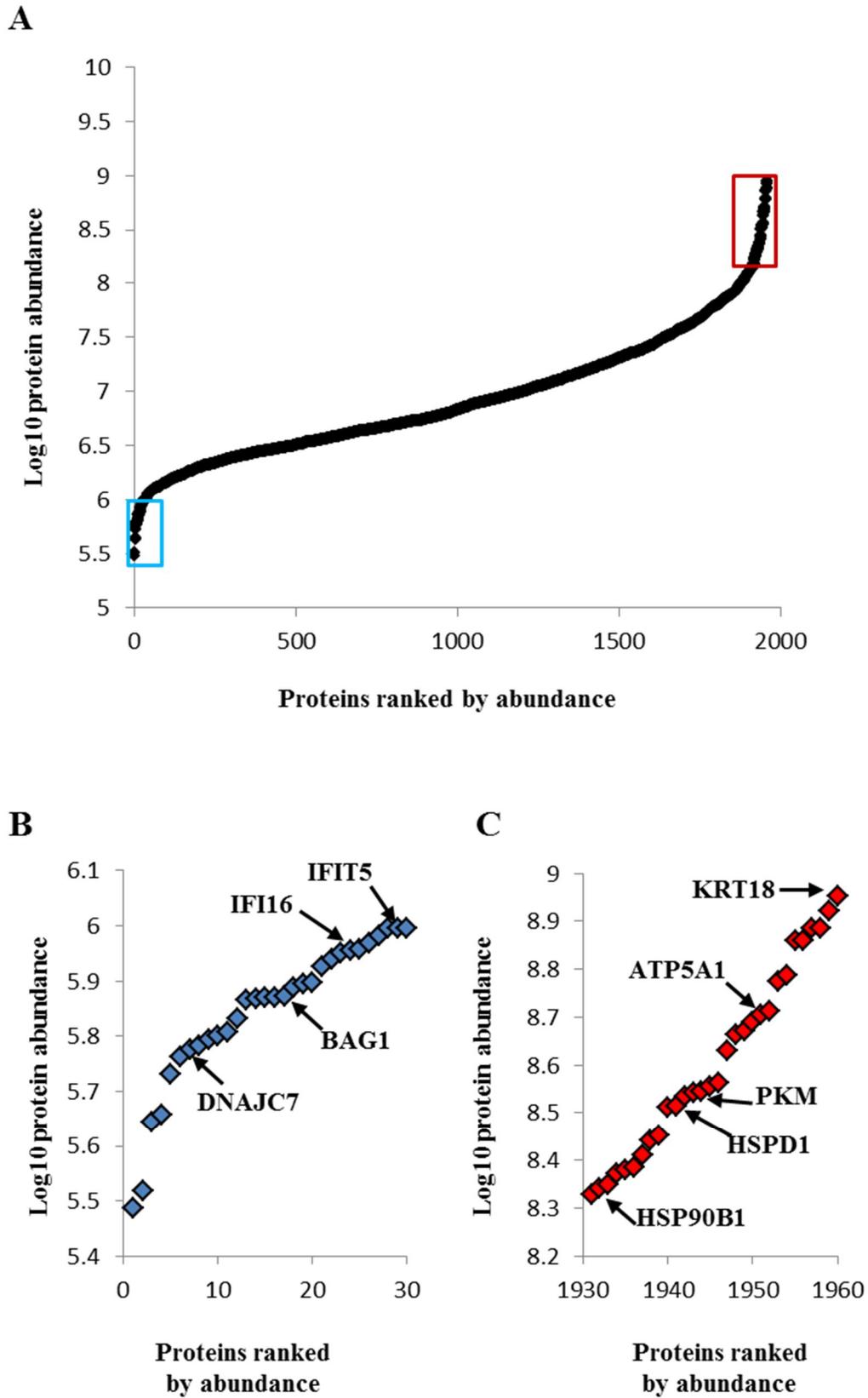
Analysis of tumor and patient characteristics between the training and test sets showed that age and menopausal status at start of tamoxifen therapy, lymph node status, and tumor size were not significantly different. The test set contained a higher proportion of poorly differentiated tumors (Pearson's  $\chi^2 = 21.19$ ,  $p$  value  $< 0.001$ ) compared to the training set. Furthermore, patients in the test cohort had a median disease free interval (DFI) of 51.4 months (range: 0 to 195 months), which was significantly longer (Mann-Whitney U = -3.814,  $p$  value  $< 0.001$ ) than for patients in the

training set (median: 16.4 months, range: 0 to 90.8 months). This can be attributed to the lack of stage IV tumors in the NKI-AVL cohort, which possibly contributed to the difference in DFI and grade between training and test set.

#### *MS analysis of ER positive breast cancer*

LCM discovery and test samples were analyzed along with 8 LCM replicates from 2 separate control tissues, and 3 technical replicates of a control WTL. A total of 2,215 proteins were quantified in LCM control samples, and 1,320 proteins in the WTL sample replicates, with only 852 proteins quantified in both LCM and WTL controls. Pearson correlation coefficients between LCM samples ranged from 0.92 to 0.97 while it ranged from 0.96 to 0.97 between WTL measurements (Supplemental Material). Hierarchical clustering of LCM and WTL controls showed grouping according to sample origin without miss-classifications (Supplemental Material). Median coefficients of variation of biological and technical replicates were 16.05% (interquartile range, IQR: 10.77-24.56) and 20.35% (IQR: 11.55-34.28), respectively. Reproducibility of MS measurements was defined as acceptable given the low number of control samples replicate measurements.

A total of 3,227 proteins were identified in the training set, of which 3,109 were quantified by LFQ. In the test set, 4,278 proteins were identified and 4,061 proteins were quantified. LFQ intensity values of 2,741 proteins commonly expressed between the training and test set were normalized for batch differences and filtered for missing data to generate two protein lists: a 1,960 protein list (10 minimum observations; Supplemental Material) for general proteome analysis and an 845 protein list for predictor development (30% missing data in training set and 0% missing data in test set; Supplemental Material). From the analysis of 1,960 expressed proteins, a wide distribution of protein abundances was observed over 3 orders of magnitude (Figure 4.2A). Interferon signaling related (e.g. IFI16, IFIT5) and chaperone associated proteins (e.g. DNAJC7, BAG1) displayed low overall abundance (Figure 4.2B), while luminal epithelial specific (e.g. KRT18), metabolism related (e.g. PKM, ATP5A1), and heat-shock (e.g. HSPD1, HSPB1) proteins were found to be highly abundant (Figure 4.2C). In the training set, CV was 14.10% (IQR: 10.22-18.78), whereas it was 13.86% (IQR: 10.33-18.57) in the test set.



**Figure 4.2. Protein abundance levels in 112 ER positive breast cancer samples.**

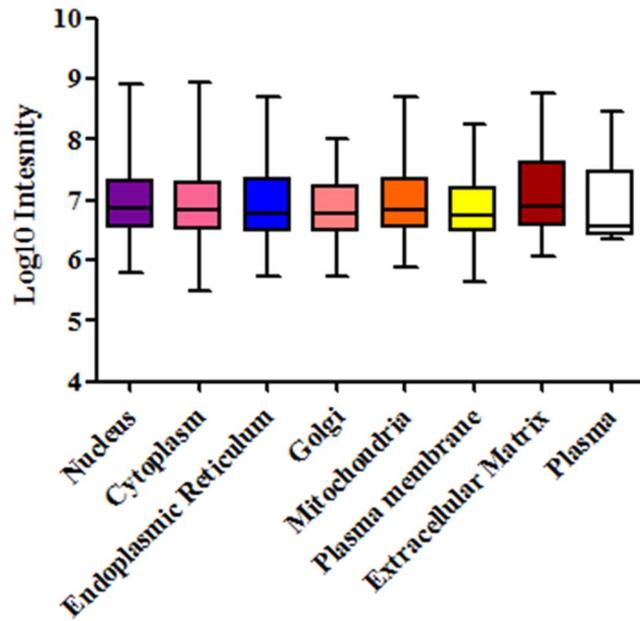
The waterfall plot shows mean protein abundance distribution of 1,960 commonly expressed proteins. The mean abundance of each quantified protein was calculated and plotted. The 30 least (blue) and most (red) abundant proteins are boxed in panel (A) and enlarged in panel (B) and (C), respectively.

DAVID based annotation for subcellular compartment showed that in the 112 breast cancer tissues the majority of expressed proteins belonged to the nuclear (25.76%) and cytoplasmic (56.38%) compartments while the endoplasmic reticulum (9.54%), Golgi apparatus (6.43%), mitochondria (12.65%), plasma membrane (7.50%), and the extracellular matrix (1.84%) comprised a lower amount of proteins. The smallest group consisted of plasma proteins (0.46%; Figure 4.3A). The distribution of intensity levels of the 1,320 proteins quantified in the WTL control sample showed a similar dynamic range but with increased variation, probably due to exclusion from the normalization procedure (Figure 4.3B). Annotation for cellular compartments showed a similar distribution of the 1,320 identified proteins into subcellular compartments compared to the 112 tissue set but with a notable enrichment of extracellular matrix (e.g. COL1A1) and plasma proteins (e.g. APOA1), which represented 7.19% and 6.89% of all quantified proteins in this set, respectively. The minor contribution of extracellular matrix and plasma proteins in the LCM samples suggests that LCM indeed resulted in highly enriched epithelial tumor cell fractions.

Distribution of intensities of organelle specific proteins showed comparable average levels of expression, therefore showing that all cell compartments were quantified. In the LCM annotated set several proteins showed multiple organelle localization. The nuclear and cytoplasmic compartments showed the highest degree of overlap with 249 (12.70%) proteins, mostly represented by proteasome subunits (e.g. PSME3) and proteins involved in RNA binding (e.g. RBM3, HNRNPA1). A small number of multi-compartmentalized proteins was constituted by vesicular transport components between the endoplasmic reticulum and the Golgi apparatus (n = 39; 1.99%) such as SEC23A. A subset of proteins was also found co-localized in both Golgi and cytoplasm (e.g. SEC24D). The remaining compartments showed expression of locally specific proteins (median overlap: 0.59% of total) such as oxidative chain proteins in the mitochondrion (e.g. UQCRC1) or DNA replication and repair involved proteins in the nucleus (e.g. FEN1). These data indicate the capability of LC-MS coupled to LCM enrichment to assess protein abundances throughout all cellular compartments from minute amounts of epithelial tumor tissues.

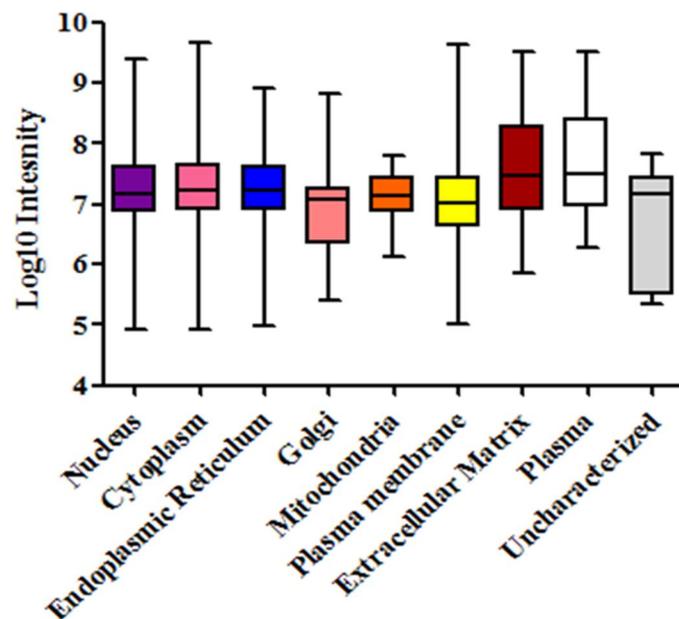
**A**

N proteins	505	1,105	187	126	248	147	36	9	Cumulative total	Total
%	25.76	56.38	9.54	6.43	12.65	7.50	1.84	0.46	120.56%	100.00%



**B**

N proteins	332	761	99	51	32	96	95	91	7	Cumulative total	Total
%	25.13	57.61	7.49	3.86	2.42	7.27	7.19	6.89	0.53	118.39%	100.00%



**Figure 4.3. Protein compartmentalization and abundance correlation analysis.**

Panel shows quantified protein abundance range per subcellular compartment in the LCM enriched 112 ER positive tumors (A) and in WTL control replicates (B). Number of proteins per compartment and percentages are displayed above the dot plot.

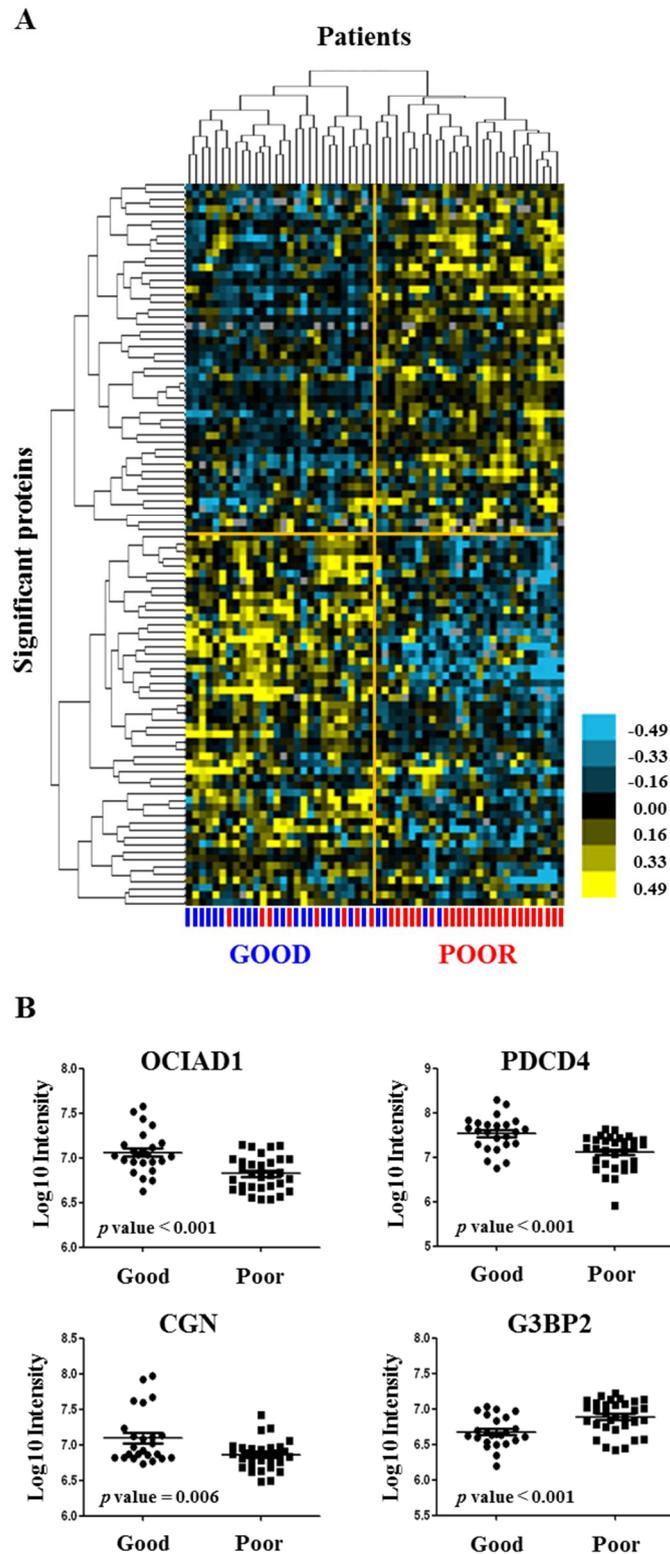
### *Analysis of differentially expressed proteins*

Due to the fact that the 1,960 proteins did not clearly discriminate patient groups (Supplemental Material), a more stringent filter for missing values was therefore applied and candidate proteins were selected based on their differential abundance between patient groups. On the panel of filtered 845 quantified proteins, a Student's t test was performed to identify 99 proteins that were differentially abundant between good and poor outcome patients in the training set ( $p$  value < 0.05). Of these, 50 proteins were found upregulated in the poor outcome group and 49 displayed higher expression in the good outcome group (Supplemental Material). In order to define molecular interaction networks between significant molecules, network analysis in IPA was performed. The network that displayed the most hits comprised proteins involved in cell growth and proliferation and cell death and survival, such as CDC37 (upregulated in poor outcome) and PDCD4 (upregulated in good outcome; Supplemental Material). Several molecules included in the network that were found upregulated in the poor outcome patient group were involved in integrin-linked kinase signaling (e.g. ITGB1, CFL1), a key pathway in cell migration and proliferation, protein translation (e.g. EIF4G1), and DNA mismatch repair (e.g. MSH2). The proteins found upregulated in the good outcome group and comprised in this network were involved in cell cycle (e.g. KRT18) and cell growth (e.g. NOP58). Although not present among the significant proteins, Akt and MAPK pathways constituted the focal point of the network, suggesting their activation based on their interactors expression levels. IPA analysis showed that differentially expressed proteins were involved in cell growth and proliferation and suggested that actors involved in such pathways may have a key role in tamoxifen resistance.

### Development of a protein signature predictive of tamoxifen therapy outcome

Based on the 99 differentially abundant proteins, hierarchical clustering separated the two patient groups (Figure 4.4A): 20 out of 28 predicted good outcome patients were correctly classified as "Good", while 24 out of 28 predicted poor outcome patients correctly grouped in the "Poor" cluster. After more stringent filtering ( $p$  value < 0.01) 38 proteins remained, which were included in a multivariate Cox regression model. Using a step-down approach, we identified a 4-protein signature that best predicted outcome to tamoxifen treatment. The signature comprised the following proteins: programmed cell death 4 (PDCD4; t test  $p$  value < 0.001), Cingulin (CGN; t test  $p$  value = 0.006), ovarian carcinoma immuno-reactive antigen domain containing protein 1 (OCIAD1; t test  $p$  value <

0.001) and Ras-GTPase activating protein-binding protein 2 (G3BP2; t test  $p$  value < 0.001; Table 4.2 and Table 4.3).



**Figure 4.4. Hierarchical clustering and differential protein abundance of 4-protein predictor.**

Samples in the training set ( $n = 56$ ) were hierarchically clustered based on 99 differentially abundant proteins (t test  $p$  value < 0.05). Log<sub>10</sub> intensities of differentially abundant proteins constituting the predictor for tamoxifen treatment

outcome are shown in scatter dot plots. Eight poor and four good outcome patients were misclassified (A). Three out of four proteins, CGN (Uniprot accession number: Q9P2M7; *p value* = 0.006), OCIAD1 (Uniprot accession number: Q9NX40; *p value* < 0.001) and PDCD4 (Uniprot accession number: Q53EL6; *p value* < 0.001), had higher abundance in patients with good outcome, whereas G3BP2 (Uniprot accession number Q9UN86; *p value* < 0.001) was found more highly expressed in the poor outcome patient group (B).

**Table 4.2. LFQ based identification of 4 proteins in discovery and validation sets.**

Protein ID	Gene name	Molecular weight (kDa)	Peptides/Unique peptides*		Sequence coverage/Unique Sequence coverage (%/%)**		PEP score†	
			Training	Test	Training	Test	Training	Test
Q9P2M7	CGN	136.380	22/22	40/40	22.8/22.8	39.5/39.5	3.63E-133	7.88E-268
Q9UN86	G3BP2	54.120	5/5	8/7	16.4/16.4	22.0/19.5	1.61E-34	5.85E-84
Q9NX40	OCIAD1	27.626	8/8	9/9	32.2/32.2	32.7/32.7	5.35E-146	2.59E-247
Q53EL6	PDCD4	51.735	19/19	18/18	49.3/49.3	47.8/47.8	1.24E-225	3.23E-281

\*Ratio between peptides and unique peptides associated to each predictor protein

\*\*Peptides/ Unique peptides sequence coverage of each protein sequence

†PEP: represents an estimation of a false identification.

**Table 4.3. Information on the 4 proteins constituting the predictor for tamoxifen therapy outcome.**

Gene name	GO cellular component	Protein name	Student t	<i>p value</i>
CGN	Cell junction	Cingulin	3.13	0.006
G3BP2	Cytoplasm	Ras GTPase-activating protein-binding protein 2	3.50	< 0.001
OCIAD1	Endosome, Mitochondrion	Ovarian carcinoma immunogenic antigen domain-containing protein 1	4.15	< 0.001
PDCD4	Cytoplasm, Nucleus	Programmed cell death protein 4	3.99	< 0.001

Based on LFQ intensity levels, OCIAD1, CGN and PDCD4 showed a relatively high abundance in good outcome patients, while G3BP2 was more highly abundant in the poor outcome group (Figure 4.4B). Next, patient scores of the 4-protein predictor were plotted in a ROC curve to select a cut-off with the highest sensitivity and specificity at predicting poor outcome (*J* = 0.740, area under the curve = 0.93, sensitivity = 90.6%, specificity = 83.3%; Figure 4.5A). The 4-protein predictor was then validated in the test cohort through Cox regression and Kaplan-Meier analyses. In both Cox univariate and multivariate regression analysis for TTP, the 4-protein signature was significantly correlated with outcome of tamoxifen therapy (HR = 2.44; 95% CI = 1.30 to 4.54; *p value* = 0.006) and multivariate (HR = 2.17; 95% CI = 1.15 to 4.17; *p value* = 0.017) regression analyses corrected for traditional predictive factors (Table 4.4). In Kaplan Meier analysis, patients with predicted poor

outcome had significantly shorter TTP compared to those with a predicted good outcome (HR = 2.32; 95% CI = 1.29 to 4.17; Log-rank *p* value = 0.004; Figure 4.5B). In the test set, sensitivity, specificity, positive predicted value (PPV), and negative predicted value (NPV) in predicting poor outcome patients were 86.7%, 41.5%, 35.1% and 89.5%, respectively.

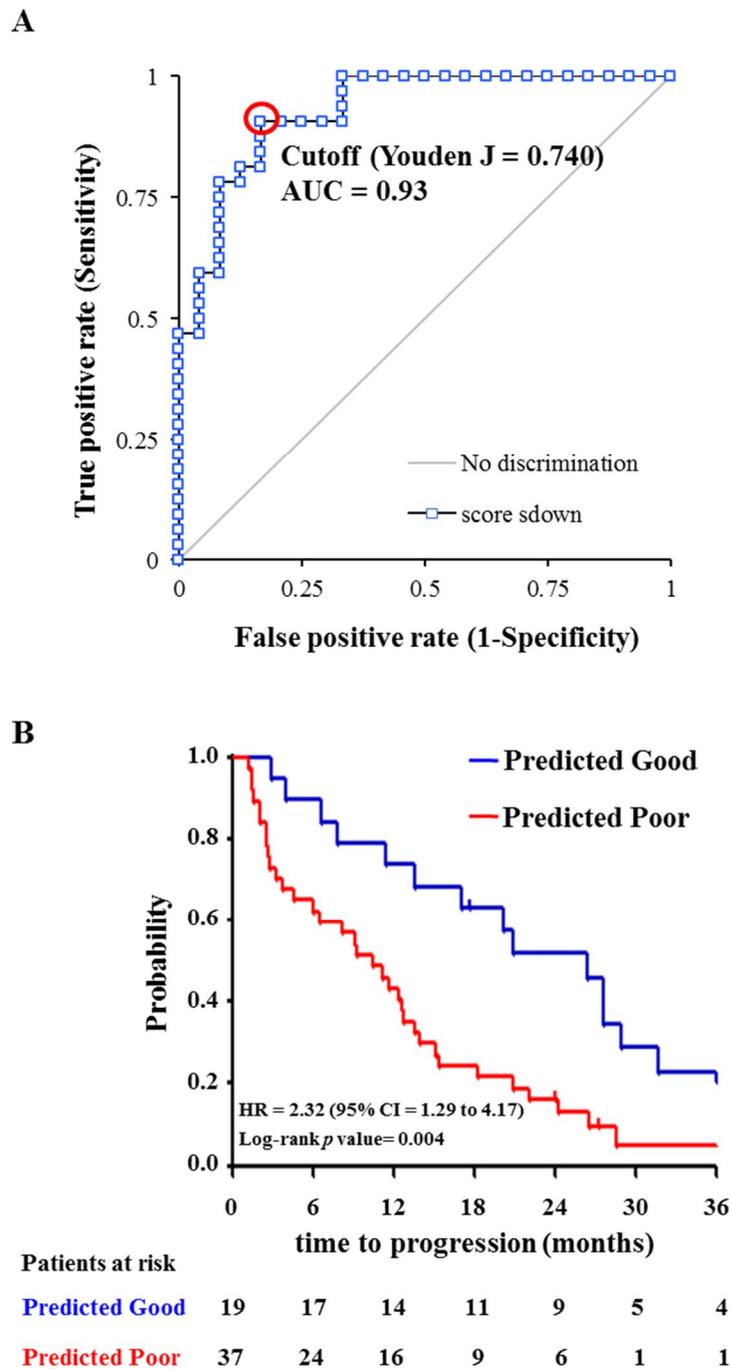


Figure 4.5. ROC curve of the training set and Kaplan-Meier curves for TTP as a function of predicted outcome in patients in the test set.

Patient outcome scores from the training set were calculated based on abundance levels of the 4 predictor proteins and protein weights (i.e. Student t value). The ROC curve was generated and Youden maximum ( $J = 0.740$ ) was chosen as the best discriminatory cutoff (A). Patient scores were subsequently calculated for patients in the test set, survival curves were generated for the predicted groups and differences were assessed with the Log-rank test (B).

**Table 4.4. Univariate and multivariate Cox regression analysis for time to progression.**

Factors	Hazard ratio	Univariate		Hazard ratio	Multivariate	
		95% CI	<i>p</i> value		95% CI	<i>p</i> value
4 protein predictor score						
High	1.00			1.00		
Low	2.44	1.30 to 4.54	0.006	2.17	1.15 to 4.17	0.017
Age						
≤55 years	1.00			1.00		
>55 years	0.44	0.23 to 0.86	0.017	0.55	0.28 to 1.08	0.083
Disease free interval						
≤12 months	1.00					
>12 months	0.63	0.30 to 1.31	0.213			
Dominant site of relapse						
Loco-regional	1.00					<i>Overall p</i>
Bone	0.89	0.30 to 2.68				0.270
Visceral	0.68	0.22 to 2.09				
Bone and other	0.45	0.14 to 1.40				
PgR						
Negative	1.00					
Positive	0.57	0.32 to 1.00	0.052			

Acronym: PgR: progesterone receptor.

#### *Immunohistochemical assessment of PDCD4 expression and correlation with TTP*

While our tissue proteomics pipeline proved to be successful in identifying and validating the 4-protein predictor, this technology is not yet readily available in a clinical setting. Therefore, we assessed protein expression of PDCD4, G3BP2, CNG, and OCIAD1 through IHC, a technology that is routinely used in diagnostic laboratories, in an independent set of formalin-fixed paraffin-embedded breast cancer tissues incorporated in a TMA. Normal breast epithelium (i.e. acini and ducts) and leukocytes displayed expression of all markers except for CGN, which stained the myoepithelial cell layer only. Blood vessels displayed expression of all 4 proteins, while overall low to negative staining was displayed in the stromal compartment. Examples of comparative IHC

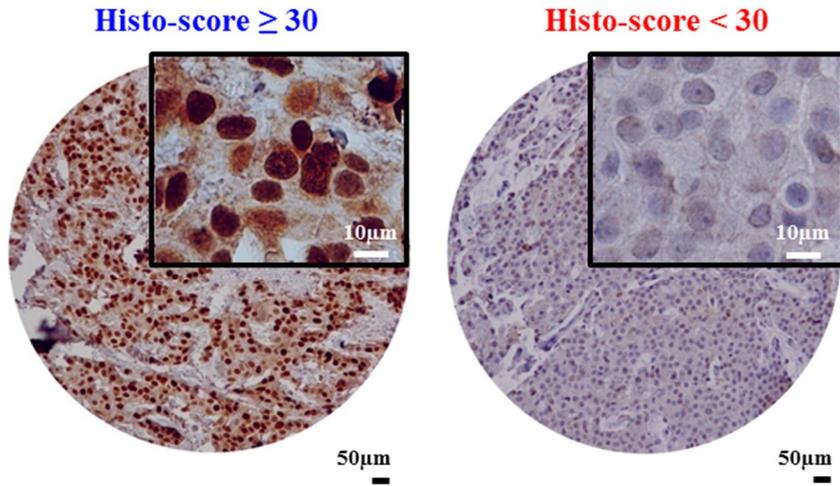
analysis of normal breast tissue, blood vessels, leucocytes and breast carcinoma cells are displayed in Supplemental Materials. Strong PDCD4 staining (histo-score  $\geq 30$ ) was found to be significantly associated with longer TTP in univariate (HR = 0.75; 95% CI = 0.59 to 0.96; *p* value = 0.020) and multivariate Cox regression analysis (HR = 0.72; 95% CI = 0.57 to 0.92; *p* value = 0.009) corrected for traditional predictive factors (Table 4.5). PDCD4 stained tissues showing both low or high protein expression and the Kaplan-Meier curve for TTP as a function of the PDCD4 histo-score are shown in Figure 4.6A and Figure 4.6B, respectively. In logistic regression analyses for clinical benefit or objective response, PDCD4 levels (histo-score  $\geq 30$  vs.  $< 30$ ) were not significantly associated with the type of response (data not shown). OCIAD1, CGN and G3BP2 stainings showed strong intensities and high quantities of stained tumor cells in the vast majority of specimens. The limited dynamic range in staining intensities proved insufficient to find a significant association of CGN, OCIAD1 and G3BP2 levels with TTP, clinical benefit or objective response (data not shown).

**Table 4.5. Univariate and multivariate Cox regression analysis for time to progression.**

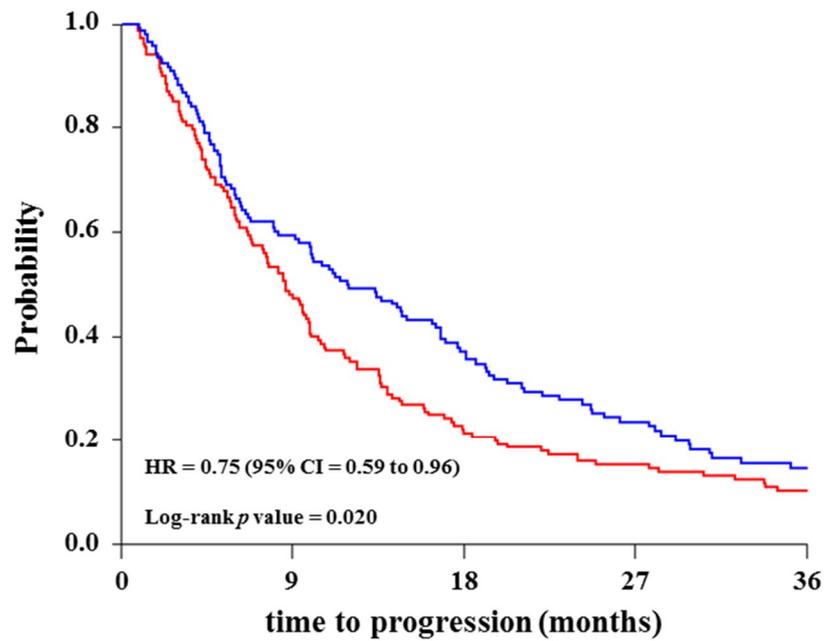
Factors	Univariate			Multivariate		
	Hazard ratio	95% CI	<i>p</i> value	Hazard ratio	95% CI	<i>p</i> value
PDCD4						
Low	1.00			1.00		
High	0.75	0.59 to 0.96	0.020	0.72	0.57 to 0.92	0.009
Age						
$\leq 55$	1.00			1.00		
$> 55$	0.58	0.45 to 0.70	$< 0.001$	0.52	0.40 to 0.67	$< 0.001$
Disease free interval						
$\leq 12$ months	1.00			1.00		
$> 12$ months	0.73	0.54 to 0.99	0.042	0.63	0.46 to 0.87	0.004
Dominant site of relapse						
Loco-regional	1.00					
Bone	1.39	0.91 to 2.10				
Visceral	1.15	0.73 to 1.81				
Bone and other	1.38	0.89 to 2.13				
			<i>Overall p</i>			
			0.310			
PgR						
Negative	1.00					
Positive	0.77	0.59 to 1.01	0.062			

Acronym: PgR: progesterone receptor.

A



B



Patients at risk

<b>Histo-score <math>&lt; 30</math></b>	152	72	32	22	14
<b>Histo-score <math>\geq 30</math></b>	142	72	47	26	15

**Figure 4.6. PDCD4 immunohistochemical staining of tissue micro-array.**

Tissue cores showed two different staining patterns that have been evaluated by histo-score (i.e. Histo-score  $< 30$  and  $\geq 30$ ), representing low and high PDCD4 protein expression (A). Patients were categorized according to histo-score cutoff and TTP was plotted as a Kaplan-Meier curve. The Log-rank test was used to test for differences in TTP between the two survival curves (B).

## Discussion

About half of the recurrent ER positive breast cancer patients treated with tamoxifen show intrinsic resistance to the drug. Despite many studies describing several mechanisms associated to tamoxifen resistance and a large amount of markers associated to patient hormonal treatment outcome, there is no molecular predictor available in the clinic (Chung and Baxter, 2012; Droog et al., 2013). Furthermore, the search for biomarkers in the analysis of clinical specimen is often hindered by tissue heterogeneity, which complicates accurate measurement of tumor protein abundance. In the light of this, tissue enrichment technologies offer an invaluable tool to quantify the proteome of specific cell subpopulations. Though mechanisms of resistance encompass not only a plethora of molecular mechanisms, but also different cell types as stromal ones (den Boon et al., 2015; Jung et al., 2015), analysis of whole tissue specimens would suffer from “signal dilution” derived from protein differential expression in heterogeneous tissues. Furthermore, analysis of microdissected stroma is hindered by the presence of high-abundance proteins (e.g. collagen family) and often needs additional protein separation methods. In this perspective, we have focused only on the epithelial tumor markers involved in tamoxifen resistance. Having successfully coupled LCM tissue enrichment with high resolution MS in a biomarker discovery pipeline (Braakman et al., 2012; Liu et al., 2014, 2012), we have here developed and validated a 4-protein signature predicting outcome to tamoxifen treatment in an independent set of ER positive recurrent breast cancer.

Despite the low amount of material derived from tissue enrichment compared to whole tissue specimens, a higher number of proteins was identified and quantified in our LCM samples (training and test sets, and controls) compared to the WTL control, suggesting interference from highly abundant proteins (e.g. collagen family) in the latter. Furthermore, from our global proteomic analysis of our combined training and test sets we showed that plasma and stromal proteins contamination was minimized in the LCM derived material while proteins expressed in subcellular compartments were enriched. This allowed us to take a unique snapshot of protein abundance of breast cancer epithelial tissue and to derive markers specifically involved in tumor cell treatment resistance pathways. From a subset of commonly expressed proteins in our 112 ER-positive breast cancer tissues we developed and validated a protein signature comprising PDCD4, CGN, OCIAD1 and G3BP2, which was capable of predicting tamoxifen treatment outcome in the test set with 86.7% sensitivity, 41.4% specificity, 35.1% PPV and 89.5% NPV and independently from traditional predictive parameters.

The selection of a large cohort of hormonal-treatment naive patients allowed us to assess tumor protein abundance directly related to first line tamoxifen treatment without any expression changes

derived from previous therapies. Furthermore, the availability of an in-house training and a multi-center test set enabled us to test the robustness of our predictor in a heterogeneous set of samples, reflective of differences in pathological evaluation and standard of care among medical centers. While our in-house training set showed almost equal distribution of patient groups, the multi-center cohort comprised a high number of good outcome patients, which could be explained by different grading systems used in local hospitals. To transfer our findings more easily to a clinical setting, we also performed IHC staining on an independent cohort of ER-positive breast cancer tissues, which confirmed PDCD4 to be an independent predictive marker of tamoxifen sensitivity. Nevertheless, the MS based 4-protein signature was a stronger predictor than the single marker PDCD4, emphasizing the potential of proteomic technologies in the dissection of tumor molecular pathways. Still, introduction of high resolution MS in routine clinical diagnostics remains problematic due to extensive and laborious sample preparation and relatively high costs. On the other side, targeted MS methods offer an accurate tool to detect and quantitate target analytes (i.e. peptides or metabolites) from biological specimens at a relatively lower cost, sample processing and measurement times (Grebe and Singh, 2011; Yassine et al., 2013), and would therefore constitute a more eligible technique for clinical introduction.

Pathways analysis on differentially expressed proteins showed how cell growth and proliferation pathways are key components in tamoxifen therapy response and resistance. Akt and MAPK, although not present among differentially expressed proteins, constituted the center of the molecular interaction network, showing how cell cycle progression through estrogen-independent mechanisms can overcome tamoxifen treatment. Activation of Akt signaling has been linked to tamoxifen resistance in previous studies (Clark et al., 2002; Klinge, 2015; Nass and Kalinski, 2015), but other molecular mechanisms may be involved. In the light of this, the 4 protein signature not only is capable of discriminating patients that manifested good and poor outcome to tamoxifen treatment, but may also pinpoint other molecular mechanisms of resistance. PDCD4 is an inhibitor of protein translation, which functions both in the nucleus and the cytoplasm (Lankat-Buttgereit and Göke, 2009). This protein has already been described as a tumor suppressor capable of inhibiting protein synthesis and gene expression by preventing the interaction of eukaryotic initiation factor (eIF) 4A1 and eIF4G, and by binding to target gene transcripts (e.g. MAP4K) in the nucleus, respectively (Biyanee et al., 2014; H. Yang et al., 2006). The nuclear localization of PDCD4 is attributed to Akt phosphorylation in a PI3K-dependent manner (Palamarchuk et al., 2005). PDCD4 levels have also been negatively correlated to increased expression of miR-21 in MCF-7 cells after tamoxifen treatment (Klinge et al., 2010; Manavalan et al., 2011). CGN is involved in tight junction formation and it has been described as a potential epithelial differentiation marker in human neoplasias (Citi et

al., 1991; Paschoud et al., 2007). Together with Paracingulin, CNG controls the expression of GATA-4, contributing to down-regulation of RhoA in cells, a key regulator of cell cycle progression that displays its function through cytoskeletal re-organization (Guillemot et al., 2013). OCIAD1 expression has been suggested as a thyroid cancer biomarker and has been correlated to distant metastasis formation, since it was found overexpressed in metastatic ovarian cancer by MS analysis (Sengupta et al., 2008; Yang et al., 2012). Recent studies have demonstrated that OCIAD1 directly interacts with STAT3 and aids in its activation, though whether this leads to activation of the tumor suppressor pathway or the oncogenic one still remains unclear (Lee et al., 2012; Musteanu et al., 2010; Sinha et al., 2013). G3BP2 has been shown to be involved in stress granule formation along with its relative G3BP1, as well as in mRNA binding and gene expression regulation. G3BP1 protein has been shown to have a distinct role in breast cancer cell proliferation by stabilizing mRNA molecules, but its homologue G3BP2 was not associated to any of these characteristics, keeping the function of this protein still ambiguous (Kociok et al., 1999; Matsuki et al., 2013; Winslow et al., 2013). With the exception of OCIAD1, no studies observed a correlation between levels of PDCD4, G3BP2, or CGN and patient survival or therapy response in clinical cancers; nonetheless these markers may play a role in the type of response to tamoxifen in breast cancer. The anti-proliferative effects of PDCD4 and CNG may have a synergistic role with the anti-estrogenic action of tamoxifen, which results in the block of cell proliferation. Due to its relatively high expression in good outcome patients, OCIAD1 may activate the tumor suppressor role of STAT3 in ER positive breast cancer patients, further inhibiting proliferation. On the other hand, expression of G3BP2 could actually counteract tamoxifen action by stabilizing mRNAs of estrogen-responsive elements as well as the ones of ER unrelated growth factors.

## **Conclusions**

We hereby demonstrate that LCM coupled to high resolution LC-MS not only enables the proteomic analysis of pure cell subpopulations, but it also provides a powerful tool for biomarker discovery studies. This allowed us to delve into the breast cancer proteome and to generate and validate a signature predictive of tamoxifen therapy outcome in recurrent ER-positive breast cancer. In addition, a technical validation through IHC verified that PDCD4 is an independent marker associated with good outcome patients, although it is difficult to distinguish small changes in protein expression by IHC. Despite the fact that shotgun LC-MS coupled to LCM based cell enrichment has shown to be a robust tool for biomarker discovery, time-consuming sample preparation and relatively high costs may hinder its introduction into a clinical setting. In the light

of this, targeted LC-MS methods such as multiple reaction monitoring would be suited to fill this gap, given the fact that accurate quantification of target analytes can be performed at lower costs with reasonable optimization times and in a multiplexed fashion.

## References:

- Altman, D.G., McShane, L.M., Sauerbrei, W., Taube, S.E., 2012. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med.* 9, e1001216. doi:10.1371/journal.pmed.1001216
- Beelen, K., Zwart, W., Linn, S.C., 2012. Can predictive biomarkers in breast cancer guide adjuvant endocrine therapy? *Nat. Rev. Clin. Oncol.* 9, 529–41. doi:10.1038/nrclinonc.2012.121
- Biyanee, a, Ohnheiser, J., Singh, P., Klempnauer, K.-H., 2014. A novel mechanism for the control of translation of specific mRNAs by tumor suppressor protein Pdc4: inhibition of translation elongation. *Oncogene* 1–9. doi:10.1038/onc.2014.83
- Bland, J., Altman, D., 1996. Measurement error proportional to the mean. *BMJ Br. Med. J.* 313, 1996.
- Braakman, R.B.H., Tilanus-Linthorst, M.M.A., Liu, N.Q., Stingl, C., Dekker, L.J.M., Luider, T.M., Martens, J.W.M., Foekens, J.A., Umar, A., 2012. Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. *J. Proteomics* 75, 2844–54. doi:10.1016/j.jprot.2012.01.022
- Cardoso, F., Harbeck, N., Fallowfield, L., Kyriakides, S., Senkus, E., 2012. Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 23 Suppl 7, vii11–9. doi:10.1093/annonc/mds232
- Cheng, L., Zhang, S., 2013. Laser-assisted microdissection in translational research: theory, technical considerations, and future applications. *Appl. ...* 21, 31–47.
- Chung, L., Baxter, R.C., 2012. Breast cancer biomarkers: proteomic discovery and translation to clinically relevant assays. *Expert Rev. Proteomics* 9, 599–614. doi:10.1586/epr.12.62
- Citi, S., Amorosi, A., Franconi, F., 1991. Cingulin, a specific protein component of tight junctions, is expressed in normal and neoplastic human epithelial tissues. *Am. J. Pathol.* 138, 781–789.
- Clark, A.S., West, K., Streicher, S., Dennis, P. a, 2002. Constitutive and inducible Akt activity promotes resistance to chemotherapy, trastuzumab, or tamoxifen in breast cancer cells. *Mol. Cancer Ther.* 1, 707–717.
- Cox, J., Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–72. doi:10.1038/nbt.1511
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J. V, Mann, M., 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–805. doi:10.1021/pr101065j
- den Boon, J. a., Pyeon, D., Wang, S.S., Horswill, M., Schiffman, M., Sherman, M., Zuna, R.E., Wang, Z., Hewitt, S.M., Pearson, R., Schott, M., Chung, L., He, Q., Lambert, P., Walker, J., Newton, M. a., Wentzensen, N., Ahlquist, P., 2015. Molecular transitions from papillomavirus infection to cervical precancer and cancer: Role of stromal estrogen receptor signaling. *Proc. Natl. Acad. Sci.* 201509322. doi:10.1073/pnas.1509322112

- Drabovich, A.P., Martínez-Morillo, E., Diamandis, E.P., 2014. Toward an integrated pipeline for protein biomarker development. *Biochim. Biophys. Acta.* doi:10.1016/j.bbapap.2014.09.006
- Droog, M., Beelen, K., Linn, S., Zwart, W., 2013. Tamoxifen resistance: From bench to bedside. *Eur. J. Pharmacol.* 717, 47–57. doi:10.1016/j.ejphar.2012.11.071
- Emmert-buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., Liotta, L.A., 1996. Laser Capture Microdissection. *Science* (80- ). 274, 8–11.
- Grebe, S., Singh, R., 2011. LC-MS/MS in the Clinical Laboratory—Where to From Here? *Clin. Biochem. Rev.* 32, 5–31.
- Guillemot, L., Spadaro, D., Citi, S., 2013. The junctional proteins cingulin and paracingulin modulate the expression of tight junction protein genes through GATA-4. *PLoS One* 8, e55873. doi:10.1371/journal.pone.0055873
- Hayward, J., Carbone, P., 1977. Assessment of response to therapy in advanced breast cancer. *Cancer* 39, 1289–94.
- Huang, D.W., Sherman, B.T., Lempicki, R., 2009a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.
- Huang, D.W., Sherman, B.T., Lempicki, R. a, 2009b. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi:10.1093/nar/gkn923
- Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–27. doi:10.1093/biostatistics/kxj037
- Jung, Y.Y., Lee, Y.K., Koo, J.S., 2015. Expression of cancer-associated fibroblast-related proteins in adipose stroma of breast cancer. *Tumor Biol.* doi:10.1007/s13277-015-3594-9
- Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., Thomas, J.K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabudhe, N. a, Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L.D.N., Patil, A.H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S.K., Marimuthu, A., Sathe, G.J., Chavan, S., Datta, K.K., Subbannayya, Y., Sahu, A., Yelamanchi, S.D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K.R., Syed, N., Goel, R., Khan, A. a, Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T.-C., Zhong, J., Wu, X., Shaw, P.G., Freed, D., Zahari, M.S., Mukherjee, K.K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C.J., Shankar, S.K., Satishchandra, P., Schroeder, J.T., Sirdeshmukh, R., Maitra, A., Leach, S.D., Drake, C.G., Halushka, M.K., Prasad, T.S.K., Hruban, R.H., Kerr, C.L., Bader, G.D., Iacobuzio-Donahue, C. a, Gowda, H., Pandey, A., 2014. A draft map of the human proteome. *Nature* 509, 575–81. doi:10.1038/nature13302
- Klinge, C.M., 2015. miRNAs regulated by estrogens, tamoxifen, and endocrine disruptors and their downstream gene targets. *Mol. Cell. Endocrinol.* doi:10.1016/j.mce.2015.01.035
- Klinge, C.M., Riggs, K. a, Wickramasinghe, N.S., Emberts, C.G., McConda, D.B., Barry, P.N., Magnusen, J.E., 2010. Estrogen receptor alpha 46 is reduced in tamoxifen resistant breast cancer cells and re-expression inhibits cell proliferation and estrogen receptor alpha 66-regulated target gene transcription. *Mol. Cell. Endocrinol.* 323, 268–76. doi:10.1016/j.mce.2010.03.013

- Kociok, N., Esser, P., Unfried, K., Parker, F., Schraermeyer, U., Grisanti, S., Toqué, B., Heimann, K., 1999. Upregulation of the RAS-GTPase activating protein (GAP)-binding protein (G3BP) in proliferating RPE cells. *J. Cell. Biochem.* 74, 194–201.
- Kondo, T., 2014. Inconvenient truth: cancer biomarker development by using proteomics. *Biochim. Biophys. Acta* 1844, 861–5. doi:10.1016/j.bbapap.2013.07.009
- Lankat-Buttgereit, B., Göke, R., 2009. The tumour suppressor Pcd4: recent advances in the elucidation of function and regulation. *Biol. Cell* 101, 309–17. doi:10.1042/BC20080191
- Lee, J., Kim, J.C.K., Lee, S.-E., Quinley, C., Kim, H., Herdman, S., Corr, M., Raz, E., 2012. Signal transducer and activator of transcription 3 (STAT3) protein suppresses adenoma-to-carcinoma transition in *Apcmin/+* mice via regulation of Snail-1 (SNAIL) protein stability. *J. Biol. Chem.* 287, 18182–9. doi:10.1074/jbc.M111.328831
- Liu, N.Q., Braakman, R.B.H., Stingl, C., Luider, T.M., Martens, J.W.M., Foekens, J.A., Umar, A., 2012. Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J. Mammary Gland Biol. Neoplasia* 17, 155–64. doi:10.1007/s10911-012-9252-6
- Liu, N.Q., Stingl, C., Look, M.P., Smid, M., Braakman, R.B.H., De Marchi, T., Sieuwerts, A.M., Span, P.N., Sweep, F.C.G.J., Linderholm, B.K., Mangia, A., Paradiso, A., Dirix, L.Y., Van Laere, S.J., Luider, T.M., Martens, J.W.M., Foekens, J.A., Umar, A., 2014. Comparative Proteome Analysis Revealing an 11-Protein Signature for Aggressive Triple-Negative Breast Cancer. *J. Natl. Cancer Inst.* 106, Epub 2014. doi:10.1093/jnci/djt376
- Manavalan, T.T., Teng, Y., Appana, S.N., Datta, S., Kalbfleisch, T.S., Li, Y., Klinge, C.M., 2011. Differential expression of microRNA expression in tamoxifen-sensitive MCF-7 versus tamoxifen-resistant LY2 human breast cancer cells. *Cancer Lett.* 313, 26–43. doi:10.1016/j.canlet.2011.08.018
- Matsuki, H., Takahashi, M., Higuchi, M., Makokha, G.N., Oie, M., Fujii, M., 2013. Both G3BP1 and G3BP2 contribute to stress granule formation. *Genes Cells* 18, 135–46. doi:10.1111/gtc.12023
- Megger, D., Bracht, T., Meyer, H., Sitek, B., 2013. Label-free quantification in clinical proteomics. ... (BBA)-Proteins Proteomics 1834, 1581–1590.
- Milani, A., 2014. Overcoming endocrine resistance in metastatic breast cancer: Current evidence and future directions. *World J. Clin. Oncol.* 5, 990. doi:10.5306/wjco.v5.i5.990
- Musteanu, M., Blaas, L., Mair, M., Schleder, M., Bilban, M., Tauber, S., Esterbauer, H., Mueller, M., Casanova, E., Kenner, L., Poli, V., Eferl, R., 2010. Stat3 is a negative regulator of intestinal tumor progression in *Apc(Min)* mice. *Gastroenterology* 138, 1003–11.e1–5. doi:10.1053/j.gastro.2009.11.049
- Nass, N., Kalinski, T., 2015. Tamoxifen resistance: From cell culture experiments towards novel biomarkers. *Pathol. Res. Pract.* 211, 189–197. doi:10.1016/j.prp.2015.01.004
- Palamarchuk, A., Efanov, A., Maximov, V., Aqeilan, R.I., Croce, C.M., Pekarsky, Y., 2005. Akt phosphorylates and regulates Pcd4 tumor suppressor protein. *Cancer Res.* 65, 11282–6. doi:10.1158/0008-5472.CAN-05-3469

- Paschoud, S., Bongiovanni, M., Pache, J.-C., Citi, S., 2007. Claudin-1 and claudin-5 expression patterns differentiate lung squamous cell carcinomas from adenocarcinomas. *Mod. Pathol.* 20, 947–54. doi:10.1038/modpathol.3800835
- Sengupta, S., Michener, C.M., Escobar, P., Belinson, J., Ganapathi, R., 2008. Ovarian cancer immuno-reactive antigen domain containing 1 (OCIAD1), a key player in ovarian cancer cell adhesion. *Gynecol. Oncol.* 109, 226–33. doi:10.1016/j.ygyno.2007.12.024
- Sereni, M.I., Baldelli, E., Gambarà, G., Deng, J., Zanotti, L., Bandiera, E., Bignotti, E., Ragnoli, M., Tognon, G., Ravaggi, A., Meani, F., Memo, M., Angioli, R., Liotta, L. a, Pecorelli, S.L., Petricoin, E., Pierobon, M., 2015. Functional characterization of epithelial ovarian cancer histotypes by drug target based protein signaling activation mapping: Implications for personalized cancer therapy. *Proteomics* 15, 365–73. doi:10.1002/pmic.201400214
- Sinha, A., Khadilkar, R.J., S, V.K., Roychowdhury Sinha, A., Inamdar, M.S., 2013. Conserved regulation of the Jak/STAT pathway by the endosomal protein asrij maintains stem cell potency. *Cell Rep.* 4, 649–58. doi:10.1016/j.celrep.2013.07.029
- Vizcaíno, J.A., Côté, R.G., Csordas, A., Dienes, J. a, Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., O’Kelly, G., Schoenegger, A., Ovelheiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R., Hermjakob, H., 2013. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41, D1063–9. doi:10.1093/nar/gks1262
- Vogel, A., Lorenz, K., Horneffer, V., Hüttmann, G., von Smolinski, D., Gebert, A., 2007. Mechanisms of laser-induced dissection and transport of histologic specimens. *Biophys. J.* 93, 4481–500. doi:10.1529/biophysj.106.102277
- Whiteaker, J.R., Lin, C., Kennedy, J., Hou, L., Trute, M., Sokal, I., Yan, P., Schoenherr, R.M., Zhao, L., Voytovich, U.J., Kelly-Spratt, K.S., Krasnoselsky, A., Gafken, P.R., Hogan, J.M., Jones, L. a, Wang, P., Amon, L., Chodosh, L. a, Nelson, P.S., McIntosh, M.W., Kemp, C.J., Paulovich, A.G., 2011. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* 29, 625–34. doi:10.1038/nbt.1900
- Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F., Kuster, B., 2014. Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582–7. doi:10.1038/nature13319
- Winslow, S., Leandersson, K., Larsson, C., 2013. Regulation of PMP22 mRNA by G3BP1 affects cell proliferation in breast cancer cells. *Mol. Cancer* 12, 156. doi:10.1186/1476-4598-12-156
- Xu, B.J., 2010. Combining laser capture microdissection and proteomics: methodologies and clinical applications. *Proteomics. Clin. Appl.* 4, 116–23. doi:10.1002/prca.200900138
- Yang, A.-H., Chen, J.-Y., Lee, C.-H., Chen, J.-Y., 2012. Expression of NCAM and OCIAD1 in well-differentiated thyroid carcinoma: correlation with the risk of distant metastasis. *J. Clin. Pathol.* 65, 206–12. doi:10.1136/jclinpath-2011-200416

Yang, F., Foekens, J.A., Yu, J., Sieuwerts, a M., Timmermans, M., Klijn, J.G.M., Atkins, D., Wang, Y., Jiang, Y., 2006. Laser microdissection and microarray analysis of breast tumors reveal ER-alpha related genes and pathways. *Oncogene* 25, 1413–9. doi:10.1038/sj.onc.1209165

Yang, H., Matthews, C., Clair, T., 2006. Tumorigenesis suppressor Pcd4 down-regulates mitogen-activated protein kinase kinase kinase kinase 1 expression to suppress colon carcinoma cell invasion. *Mol. Cell. Biol.* doi:10.1128/MCB.26.4.1297

Yassine, H., Borges, C.R., Schaab, M.R., Billheimer, D., Stump, C., Reaven, P., Lau, S.S., Nelson, R., 2013. Mass spectrometric immunoassay and MRM as targeted MS-based quantitative approaches in biomarker development: Potential applications to cardiovascular disease and diabetes. *Proteomics - Clin. Appl.* 7, 528–540. doi:10.1002/prca.201200028

## **Chapter 5**

### **Antibody-Based Capture of Target Peptides in Multiple Reaction Monitoring Experiments**

Tommaso De Marchi, Eric Kuhn, Steven A. Carr, Arzu Umar

*Methods Mol Biol.* 2015; 1293:123-35.

## **Abstract**

Targeted quantitative mass spectrometry of immunoaffinity-enriched peptides, termed immuno-multiple reaction monitoring (iMRM), is a powerful method for determining the relative abundance of proteins in complex mixtures, like plasma or whole tissue. This technique combines 1,000-fold enrichment potential of antibodies for target peptides with the selectivity of multiple reaction monitoring mass spectrometry (MRM-MS). Using heavy isotope-labeled peptide counterparts as internal standards ensures high levels of precision. Further, LC-MRM-MS selectivity allows for multiplexing; antibodies recognizing different peptides can be added directly to a single mixture without subjecting to interferences common to other multiple antibody protein assays. Integrated extracted ion chromatograms (XIC) of product ions from endogenous unlabeled “light” peptide and stable isotope-labeled internal standard “heavy” peptides are used to generate a light/heavy peak area ratio. This ratio is proportional to the amount of peptide in the digestion mixture and can be used to estimate the concentration of protein in the sample.

## Introduction

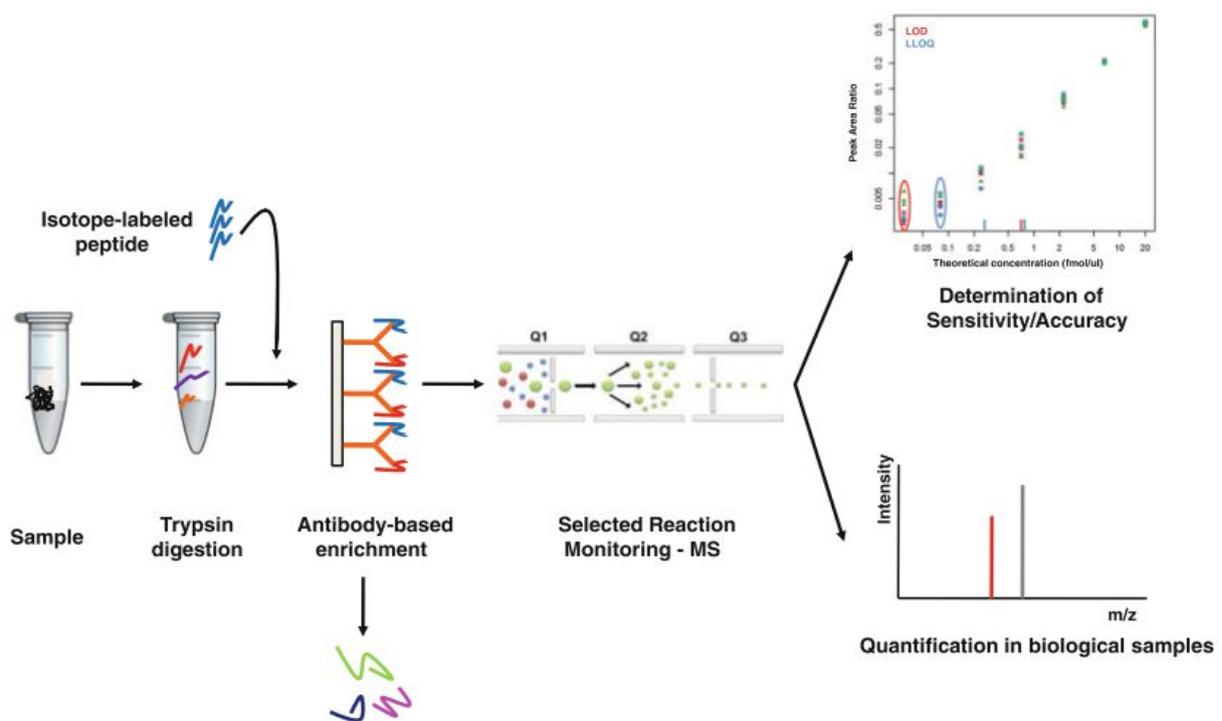
Targeted mass spectrometry (MS) has made rapid advancements during the last decade and has begun to demonstrate the sensitivity and selectivity of highly developed immunoassays. These advancements make it possible to simultaneously measure concentration of 10's of proteins or more in a single biological sample (e.g. body fluid, tissue lysate) reproducibly (1–3). Using a scan type called multiple reaction monitoring (MRM; also referred to as selected reaction monitoring [SRM]) on a triple quadrupole mass spectrometer (QqQ), mixtures of digested peptides derived from biological samples separated by nanoflow reversed phase liquid chromatography (nano-RPLC) are detected and analyzed. Eluted peptides are ionized from liquid to the gas phase by a combination of applied voltage (typically 2000V in nanospray) and drying gases (N<sub>2</sub>). Electrospray ionization generates gas phase ions that are introduced into the mass spectrometer. In the first quadrupole (Q1) target ions are filtered and accelerated toward quadrupole two (Q2) which is set to a higher pressure and acts as a collision cell. Peptides are fragmented in Q2 by collision-induced dissociation (CID) and accelerated toward quadrupole three (Q3) which scans for the pre-determined corresponding product ions. This transition ion, a Q3 detected mass of a peptide fragment that corresponds to a selected mass of a peptide precursor, is acquired in approximately 10 ms. Depending on the number of transition ions in a method and the chromatographic peak width, between six and twenty scans are collected for each transition. To confirm identification, typically 3-5 transition ions are monitored per peptide. To increase precision of quantification, stable isotope labeled peptides, peptides containing an amino acid or acids with <sup>13</sup>C or <sup>15</sup>N (<sup>2</sup>H is not preferred as chromatographic retention time is affected), can be synthesized and added to the mixture prior to nano-RPLC-MRM-MS analysis. This method, termed stable isotope dilution mass spectrometry (SID-MS), can account for variability in chromatographic retention time and ionization, which can be affected by variable amounts of background (peptidic and otherwise) in a mixture from run to run and sample to sample (4–8).

To determine the assay performance of each peptide, serial dilutions of concentration standard peptides are added to a background mixture of peptides prepared to mimic the expected matrix of the patient or study-derived sample. So called “reverse curves” are curves comprised of a variable concentration of stable isotope standard peptide and a fixed or singular concentration of corresponding unlabeled peptide are prepared to determine the lower limit of quantitation (LLOQ) and limit of detection (LOD) (9, 10). This configuration, while generating measurements with a peak area ratio inverse to that of the intended “forward” measurement, reduces the interference that endogenous peptide can introduce. Depending on the analytes selected for MRM, background

matrix of cell lysates or plasma may contain sufficient quantity of endogenous peptide to interfere with the determination of lower limits for the assay.

LC-MRM-MS based techniques have proven useful for determining the relative amounts of proteins in a complex sample (1), but the sensitivity for analyte peptides present in lower abundance is also affected by high amounts of the background proteome derived from a complex sample. Therefore enrichment of target analytes is necessary to determine quantities of peptides present at lower concentrations (e.g.  $< 1 \mu\text{g/mL}$  in plasma) (5, 8, 9). Immunoaffinity enrichment of proteins from cell lysates for example, is a well-established method for extracting the protein out of the mixture (11). Recently, this methodology has been applied to the peptide level enrichment using antibodies generated against the analyte peptides from proteins of interest directly without a pre-enrichment step such as fractionation or depletion (12). Anderson et al. were the first to describe the use of this approach coupled with SID-MRM-MS, and called it Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA (12)). Here we refer to the method by the more generic name of immuno-MRM or iMRM. The antibodies used for peptide immunoaffinity enrichment are typically polyclonal (2, 3, 12, 13) but can also be generated as monoclonals (14). Polyclonal antibodies are purified by peptide affinity chromatography from anti-sera generated by immunizing rabbits with KLH-conjugated peptides unique for a target protein that have been selected from MRM public repositories (e.g. SRM atlas, GPM, etc.) (15). The most reliable method to obtain a peptide list to quantitate proteins in biological samples is to generate an *in silico* library from previous MS experiments using software packages such as Skyline (16). Skyline is frequently used to generate a peptide library which can be derived from spectra that have been searched using Mascot (17) or Spectrum Mill (Agilent Technologies Inc., Santa Clara CA). Peptide ion intensities can also be ranked *in silico* using the ESP algorithm (18). Further selection of peptides with fragment ions best suited for MRM-MS can be made by matching the *in silico* digest of a target protein with previously generated MS/MS spectra. Generating an antibody of sufficient titer and affinity is also dependent on amino acid sequence of the peptide. Algorithms for determining the hydrophilicity of a section of peptide sequence within the context of the full length protein sequence exist; however, specialized algorithms to rank the immunogenicity of single peptides are not currently available. To increase the success rate of antibody generation, rabbits are immunized in pairs with three to five peptides per protein (13). To evaluate antibody affinity, stable isotope labeled counterparts for peptides against which an antibody was generated are synthesized and purified from commercial sources (New England Peptide, Thermo, 21st Century Biochemicals). Antibody performance is then assessed by capture efficiency (percent of available peptide in a mixture) and LOD (from calibration curves). Immunoaffinity enrichment combined with high sensitivity MRM-MS makes

iMRM particularly well suited for quantifying analyte peptides (surrogates for relative abundance of protein) in complex samples, such as plasma, tissues, or cell lysates. Due to the selective properties of MRM-MS, antibodies can be used individually or within a multiplex without diminishing their performance (2, 19). Linking antibodies onto Protein G magnetic beads makes the assay more amenable to robotic automation and the potential to increase throughput and robustness. We herein describe a method for relative precise quantification of analyte peptides from breast cancer whole tissue lysate samples through iMRM-MS (Figure 5.1).



**Figure 5.1.**

Outline of an immuno-MRM experiment. Once target peptides have been selected, they are captured along with a known amount of their isotope-labeled counterpart through anti-peptide antibodies. Heavy and endogenous peptides are eluted and analyzed on a triple quadrupole mass spectrometer in MRM mode. Capturing a fixed amount of endogenous peptide with decreasing concentrations of its isotope-labeled version enables determination of the limit of detection (LOD) and lower limit of quantitation (LLOQ) of the method. Quantification of target peptide/protein in biological samples is derived from Peak Area Ratio (PAR) determination between endogenous peptide level and its heavy counterpart.

## Materials

### *Selection of target peptides*

From either previous tandem MS experiments or public repositories select a minimum of 2 unique tryptic peptides per target protein.

### *Peptides and antibody buffers*

1. Antibody wash solution 1: 1X PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, 2 mM KH<sub>2</sub>PO<sub>4</sub>) - 0.03% w/v CHAPS. Dissolve 300 mg of CHAPS in 1 L of 1X PBS.
2. Antibody wash solution 2: 0.1 X PBS - 0.03% w/v CHAPS. Dilute 1X PBS 1:10 in HPLC-grade water and dissolve 300 mg of CHAPS in 1 L of 0.1X PBS.
3. Stable isotope labeled (heavy) peptide high concentration diluent: 30% v/v ACN/0.1% v/v FA. Dissolve 300 mL acetonitrile (ACN) and 10 mL formic acid (FA) in 1 L HPLC-grade water.
4. Stable isotope labeled (heavy) peptide low concentration diluent – MS blank buffer – Antibody elution buffer: 3% v/v ACN/5% v/v acetic acid (AcOH). Dissolve 30 mL ACN and 50 mL AcOH in 1 L HPLC grade water.

### *Heavy peptide standard stock solutions and mixtures*

1. Heavy peptide high concentration stocks: prepare serial dilutions of heavy peptide stocks to a final concentration of 100 pmol/  $\mu$ L using 30% ACN/0.1% FA as diluent.
2. Combine equivalent volumes of heavy peptide high concentration stocks in one tube. Do not dilute. Concentration depends on number of peptides in the final mixture
3. Heavy peptide low concentration mixtures: dilute the high concentration stocks such that the final concentration is 100 fmol/  $\mu$ L, use 3% ACN/5% AcOH as diluent.

### *Anti-peptide antibody master-mix solutions*

1. Antibody high concentration mixture: combine equivalent amounts of antibody stock solutions to a final concentration 20 µg/mL for each antibody is. Use PBS 0.03% CHAPS as diluent.
2. Antibody titration solutions: prepare a serial dilution from the high concentration antibody mixture into 1X PBS 0.03% CHAPS to generate a series of final concentrations for each antibody per capture of 4 , 2, 1, 0.5, 0.25, and 0.125 µg per 50 µL

### *Cross-linking solutions*

1. Crosslinking solution: 200 mM triethanolamine (TEA) pH 8.5. Dissolve 10 mL of neat triethanolamine into 400 mL HPLC-grade water; adjust pH with 5 M HCl.
2. Crosslinking solution: 20mM Dimethyl pimelimidate (DMP) in 200 mM TEA. Dissolve 1.03 g of DMP in 200 mL of 200 mM TEA.
3. Crosslinking quenching solution: 150 mM monoethanolamine (MEA) pH 9.0: dissolve 3.15 mL of pure monoethanolamine in 400 mL HPLC-grade water; adjust pH with 5 M HCl.
4. Crosslinking washing solution: 5% v/v AcOH/0.03% w/v CHAPS: Dissolve 50 mL acetic acid and 30 mg of CHAPS in 1 L HPLC-grade water.
5. Re-suspension buffer: 1X PBS/0.03% w/v CHAPS/0.1% w/v NaN<sub>3</sub>. Dissolve 30 mg CHAPS and 1 g of NaN<sub>3</sub> in 1 L PBS.
6. Magnetic rack and Kingfisher Flex Magnetic Particle Processor (Thermo Scientific, Rockford, IL USA).
7. MyOne Protein G 1µm magnetic beads (Dyna/Invitrogen/Life Technologies/Thermo)

## **Methods**

### *Antibody crosslinking using KingFisher (optional)*

1. Add volumes of Protein G (ProG) magnetic beads into Antibody mixes to a final bead : antibody ratio of 2:1 (µL to µg). Tumble mix solutions overnight at 4°C.
2. Add 900 µL/well of cross-linking solution onto a 1 mL Kingfisher plate.

3. Set Kingfisher method as follows:
  - o Plate 1: antibody-beads solutions.
  - o Plate 2: cross-linking solution (900  $\mu\text{L}$ /well), rinse 30'.
  - o Plate 3: crosslinking quenching solution (900  $\mu\text{L}$  150 mM MEA), rinse 30'.
  - o Plate 4: crosslinking washing solution (900  $\mu\text{L}$  5% AcOH/0.03% CHAPS), rinse 5'.
  - o Plate 5: crosslinking washing solution (900  $\mu\text{L}$  5% AcOH/0.03% CHAPS), rinse 5'.
  - o Plate 6: PBS washing solution (900  $\mu\text{L}$  PBS 0.03% CHAPS), rinse 5'.
  - o Plate 7: re-suspension buffer (900  $\mu\text{L}$  PBS/0.03% CHAPS/0.1% NaN<sub>3</sub>), rinse 5'.
  - o Plate 8: magnetic tip cover plate.

#### *Antibody capture efficiency evaluation*

1. Prepare a 1X PBS 0.03% CHAPS stock solution. Re-suspend amount of digested lyophilized protein to target concentration per capture well as desired (e.g. cell lysates 100-500  $\mu\text{g}/200$   $\mu\text{L}$ , plasma 10  $\mu\text{L}$  equivalent = 600  $\mu\text{g}/200$   $\mu\text{L}$ ) in 1X PBS-0.03% CHAPS-0.5 fmol/  $\mu\text{L}$  heavy mix.
2. Prepare a 1X PBS 0.03% CHAPS stock solution. Re-suspend amount of digested lyophilized protein to target concentration per capture well as desired (e.g. cell lysates 100-500  $\mu\text{g}/200$   $\mu\text{L}$ , plasma 10  $\mu\text{L}$  equivalent = 600  $\mu\text{g}/200$   $\mu\text{L}$ ) in 1X PBS-0.03% CHAPS-0.5 fmol/  $\mu\text{L}$  heavy mix.
3. Add 100 fmol heavy peptide (e.g. to final concentration of 0.5 fmol/ $\mu\text{L}$  for 200  $\mu\text{L}$ ). Transfer 200  $\mu\text{L}$  of reconstituted background or sample in each well of a King Fisher 250 96-wellplate. Add 50  $\mu\text{L}$  of antibody mix to each well. Seal plate with aluminum adhesive foil and tumble mix plate overnight at 4°C.
4. Set Kingfisher magnetic particle processor method as follows:
  - a. Plate 1: antibody capture plate.
  - b. Plate 2: washing (250  $\mu\text{L}$ /well), 1.5' rinse.

- c. Plate 3: washing (250  $\mu$ L 1X PBS/0.03% CHAPS), 1.5' rinse.
  - d. Plate 4 washing (250  $\mu$ L 0.1X PBS/0.03% CHAPS), 1.5' rinse.
  - e. Plate 5: elution (25  $\mu$ L 3% ACN/5% AcOH), 5' rinse.
  - f. Plate 6: bead collection, (200  $\mu$ L 1X PBS/0.03% CHAPS/0.1%NaN<sub>3</sub>), 5' rinse.
  - g. Plate 7: magnetic tip comb.
- 
5. Once KingFisher method has completed, remove Elution Plate 5 and place on a magnetic plate holder on wet ice.
  6. Place a fresh 96-well PCR plate on wet ice and add 5  $\mu$ L of 3% ACN/5% AcOH per well.
  7. Using a multi-channel pipet, transfer 25  $\mu$ L of the supernatant from Elution Plate 5 into the fresh PCR plate.
  8. Using a multi-channel pipet, transfer 25  $\mu$ L of the supernatant from Elution Plate 5 into the fresh PCR plate.
  9. Centrifuge plate briefly (30 s) to eliminate air bubbles. and place plate onto autosampler for analysis on TQMS.
  10. Centrifuge plate briefly (30 s) to eliminate air bubbles. and place plate onto autosampler for analysis on TQMS.
  11. Analyze by MRM for pre-selected optimized transitions for light and heavy peptides masses by nano-RPLC-MRM-MS.
  12. Export results file to Skyline to integrate data. Select optimal capture concentrations per each antibody (i.e. maximum heavy peptide signal, least noise).

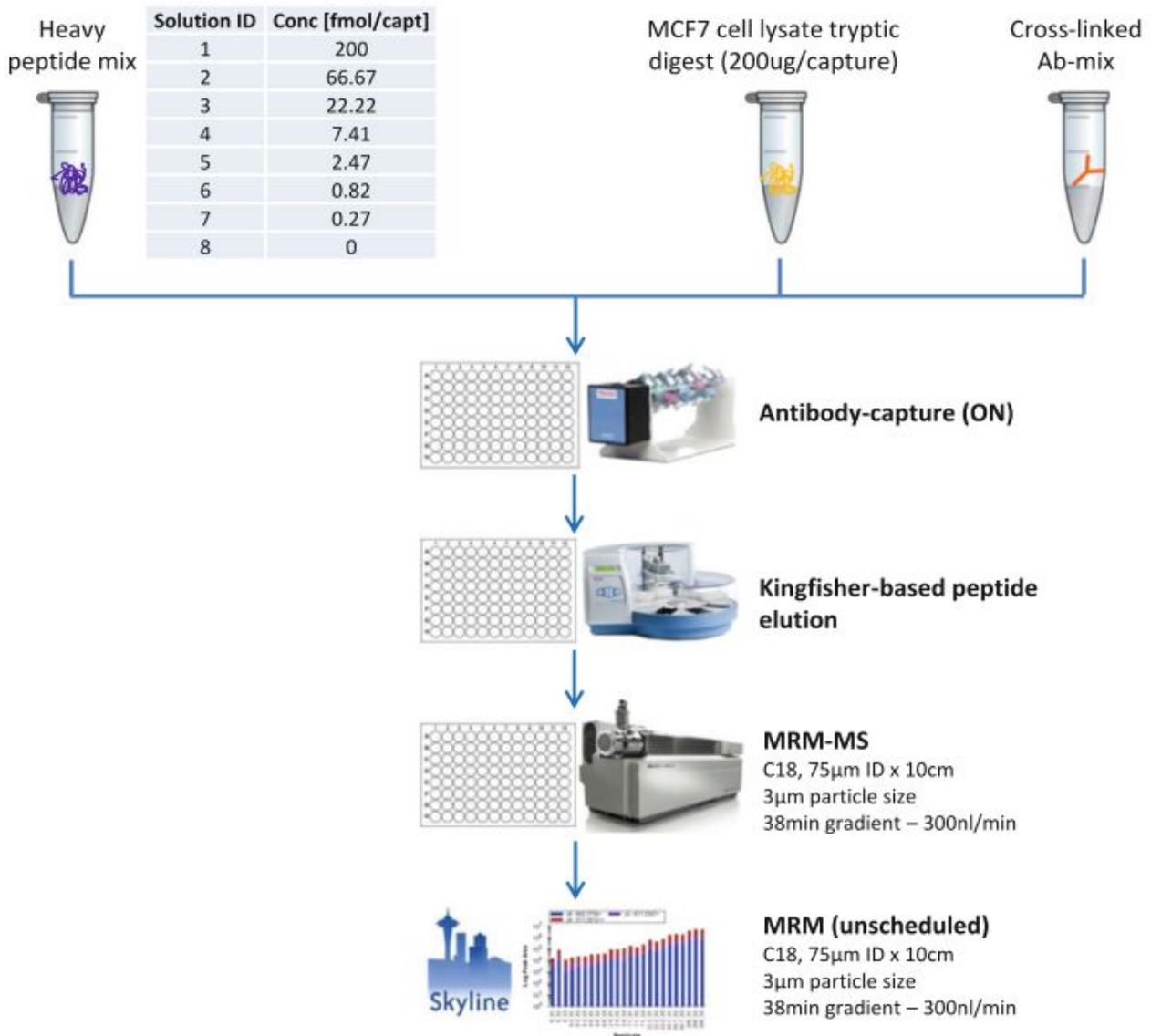
#### *Evaluation of passenger peptide*

1. Prepare 3 replicates of antibody capture at a fixed antibody mix concentration. Add equivalent amounts of heavy peptide mix per each replicate capture. Depending on type of background matrix prepared (digested cell lysate, tissue or plasma) add 1:200 to 1:500 diluted background per each replicate capture.

2. Perform overnight capture and peptide elution as in 3.2.3.
3. Analyze on a triple quadrupole MS instrument configured with nanoflow liquid chromatograph and autosampler.
4. Export results in Skyline and evaluate presence of light peptide in captures with undiluted vs diluted background.

#### *Generation of reverse curve*

Prepare serial dilutions of heavy peptide mixtures for a 7 or 8 point curve with concentrations from 0.3 to 200fmol in 200  $\mu$ L volume of 1X PBS/0.03% CHAPS/100-500  $\mu$ g background digested protein per capture. Include a comparable solution without heavy peptide. Mixtures may be made in bulk for total number of replicates (e.g. 600  $\mu$ L for 3 replicates; Figure 5.2).



**Figure 5.2.**

Figure represents typical reverse curve experiment for determination of LOD/LLOQ of an iMRM assay. Heavy peptides are captured at different concentrations by antibodies in presence of digested protein background (plasma or digested cell line proteins). Peptides are immunoaffinity enriched by tumble mixing 96-well plates overnight at 4°C. Plates are transferred to the KingFisher magnetic particle handler the next day and peptides are eluted from the beads. MRM-MS is performed after transfer of captured peptides onto a new plate and results are imported and analyzed in Skyline.

1. Transfer 200 µL into wells on a King Fisher 250 well plate.
2. Based on capture efficiency data (see Antibody Capture Efficiency Evaluation) prepare the antibody mix (crosslinked or non-crosslinked as desired) and add 50 µL of antibody mix to each well.
3. Cover with aluminum adhesive foil and tumble mix overnight at 4°C.

4. Set Kingfisher magnetic particle processor method as follows:
  - o Plate 1: antibody capture plate.
  - o Plate 2: washing (250  $\mu$ L/well), 1.5' rinse.
  - o Plate 3: washing (250  $\mu$ L 1X PBS/0.03% CHAPS), 1.5' rinse.
  - o Plate 4 washing (250  $\mu$ L 0.1X PBS/0.03% CHAPS), 1.5' rinse.
  - o Plate 5: elution (25  $\mu$ L 3% ACN/5% AcOH), 5' rinse.
  - o Plate 6: bead collection, (200  $\mu$ L 1X PBS/0.03% CHAPS/0.1%NaN<sub>3</sub>), 5' rinse.
  - o Plate 7: magnetic tip comb.
5. Analyze on a triple quadrupole MS instrument configured with nanoflow liquid chromatograph and autosampler.
6. Import results in a Skyline version containing QuaSAR. In the Result Grid tab fill in "SampleGroup" and "Concentration", which refer respectively to the sample replicate ID and to the concentration of the analyte. The "IS spike" file refers to the concentration of light peptide present in each capture. If a light version of each peptide is used then fill light peptide concentration in "IS spike", while fill in "1" if light peptide is not detected or not being used (area only curve, not based on peak area ratio). (<https://skyline.gs.washington.edu/labkey/announcements/home/software/Skyline/tools/thread.view?rowId=5436>). QuaSAR analysis in point 3.4.7 refers to an experiment in which only heavy peptides are used.
7. Perform a QuaSAR analysis, setting "Analyte" and "Standard" fields as heavy area and light area, respectively. Un-tick "Standard present" option if not using light peptides. Tick all options in the "Generate" sub-menu. LOD and LOQ of the method will be generated.
8. Export linear and log plots of concentration curves, tables of LOD and LOQ, and plots of CVs for all peptides in the group.

## Notes

1. Target peptides to be analyzed in an iMRM experiment should be derived from trypsin digestion without missed-cleavages, be unique to target protein and not contain cysteine or methionine residues. Sequences containing serial arginine (R) or lysine (K) residues (e.g. RR or KK) would be cleaved randomly at one of the basic residues and the whole peptide would change in mass, making its detection and quantification problematic. Non-unique peptides would make the quantitation less accurate, since their intensities would derive from different proteins, while sulphur-containing amino acids are targets for covalent modifications that could change the total molecular weight of the peptide. Therefore it is advised to select sequences, in which these are absent.
2. An aqueous solution containing a low percentage of organic solvent and a small percentage of acid is well suited to be used as a blank in MS experiments. Running blank solutions before and after each sample (duplicate/triplicate) would assess the presence of carry-over but would not solve it. A saw tooth gradient with multiple ramps of high organic solvent concentration is advised to remove/minimize any carry over. Antibody affinity is optimal at neutral pH, based on a combination of non-covalent interactions. These interactions are removed or inhibited under acidic conditions ( $\text{pH} < 2.5$ ) and the bound peptide is released from the antibody. A small amount of organic solvent (3% ACN) aids in peptide solubility in the absence of matrix and antibody.
3. Preparation and dilution of heavy peptide mixtures reduces the need for potential additional freeze-thaw cycles of original stocks, thereby reducing the possibility of degradation of peptides. With these solutions it is possible to evaluate the capture efficiency of the anti-peptide antibodies, comparing a captured heavy peptide with its spiked-in counterpart.
4. CHAPS is a non-denaturing, non-ionic detergent which makes it more amenable to downstream mass spectrometry. It is added primarily to keep magnetic beads from settling on the KingFisher magnetic bead handler. It also helps re-solubilize protein digests prepared under denaturing conditions (e.g. Urea). Other detergents may be used to solubilize proteins prior to digestion, but additional detergent removal steps may be necessary to avoid

irreversibly binding onto C18 packing material as well as contaminating the mass spectrometer.

5. Typical concentration range for antibodies in immuno-purification experiments varies between 0.5 and 2.0  $\mu\text{g}$  per capture for polyclonal antibodies (lower for monoclonals), but in order to determine the optimal concentration for each antibody, a titration curve above and below that range should be prepared. Seven concentration levels from 0.125  $\mu\text{g}$  to 4  $\mu\text{g}$  were chosen, but further considerations at the extremes are advised. To maintain antibody concentration at the low end (0.125  $\mu\text{g}$ ), use immediately after preparing dilution series and do not store for long periods of time. At the high end, depending on the number of antibodies in the multiplex, significant amounts of magnetic beads may be needed which could be non-specifically lost during wash and elution steps on KingFisher. The target volume for each capture is typically 250  $\mu\text{L}$  per well – which consists of 200  $\mu\text{L}$  1X PBS, 0.03% CHAPS, 2  $\mu\text{L}$  100 fmol/ $\mu\text{L}$  heavy mix and 50  $\mu\text{L}$  of antibody mix. Try not to exceed 300  $\mu\text{L}$  per well in the Kingfisher 96-well plates.
6. TEA is viscous and care should be used to accurately pipet volumes ensure mixing is complete once added to HPLC grade-water. Prepare DMP fresh and use immediately. pH of TEA and MEA solutions is critical (8.5 and 9.0 respectively). Use a pH meter when adjusting pH by addition of 5 M HCl.
7. Sodium azide ( $\text{NaN}_3$ ) is a preservative and a bacteriostatic agent. Re-suspending cross-linked antibody beads into a  $\text{NaN}_3$  containing solution allows storage at 4°C for several months or longer.
8. Protein G is the recommended ligand for binding antibodies derived from rabbit. It is a bacterial-derived protein binds the Fc of antibodies leaving the variable regions accessible to bind and release peptide epitopes. It is commercially available conjugated to agarose (Agarose Bead; ABT Technologies, Tampa, FL USA), sepharose (Sephacrose 4B®; Sigma-Aldrich Corporation, St. Louis, MO, USA), POROS (POROS®; Life Technologies, Foster

City, CA USA), and magnetic beads (Dynabeads®, Life Technologies, Foster City, CA USA). Each bead has a different size. Magnetic beads were chosen to make the process more amenable to automation. Smaller bead sizes yield more surface area per volume of bead. In addition to analyte peptides, intact IgG will also elute from Protein G under acidic conditions which can overload and foul the LC column used in nanoflow conditions (1 µg total capacity). Therefore, it is recommended to crosslink antibodies with DMP (in addition to the benefit of removing passenger peptides (see section 3.3)) to reduce non-specific background for reducing the overall signal in the mass spectrometer.

9. Optimal ratio of antibody to magnetic-bead ranges from 1:1 to 1:10 (according to manufacturer specifications) depending on the antibody. Crosslinking after capture at different ratios to determine the optimal one is advised.
  
10. The maximum volume per well in a deep well plate (max volume 2 mL/well) is limited on the KingFisher to 0.9 mL. Volumes above 0.9 mL will exceed the capacity of the well with the magnetic head and tip comb are inserted during mixing and transfer of magnetic beads. For these cases, use a magnet to readjust the final volume below 0.9 mL prior to crosslinking on KingFisher.
  
11. The choice of background material depends on the nature of the samples that have to be analyzed, while the amount to be added to each capture has to be determined experimentally. The aforementioned capture was performed adding 500 µg of digested proteins from MCF7 breast cancer cell line to each capture. Generating a reverse curve in the presence of a background that best simulates the biological sample matrix and complexity will help determine the most representative LOD and LLOQ for the samples being analyzed. Each capture should be performed at least in triplicate. The concentration of non-specific background peptides is proportional to the amount of antibodies used in each capture. These levels may vary intentionally (as in the titration experiment) or systematically (between a mock sample background and the real samples themselves). To minimize these effects on the analytical process, it is advised to cross-link antibodies onto magnetic beads.

12. Elution into small volumes (less than 50  $\mu\text{L}$ ) into the 250 King Fisher plates is not advised, Therefore Elution plate 5 is a 150  $\mu\text{L}$  96-well PCR plate to increase the recovery from magnetic beads.
13. Small amounts of magnetic beads may elute with the captured peptides. A magnetic plate holder would allow transferring the supernatants from the PCR elution plate onto a fresh plate without transferring the beads as well. Putting the plate on ice decreases evaporation of organic solvents and preserves captured peptides during handling.
14. QuaSAR is a Skyline add-on which can be downloaded at: <https://brendanx-uw1.gs.washington.edu/labkey/announcements/home/software/Skyline/tools/thread.view?rowId=5436>. And may not be available directly through Skyline without add-on.

## References:

1. Addona, T. a, Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., Spiegelman, C. H., Zimmerman, L. J., Ham, A.-J. L., Keshishian, H., Hall, S. C., Allen, S., Blackman, R. K., Borchers, C. H., Buck, C., Cardasis, H. L., Cusack, M. P., Dodder, N. G., Gibson, B. W., Held, J. M., Hiltke, T., Jackson, A., Johansen, E. B., Kinsinger, C. R., Li, J., Mesri, M., Neubert, T. a, Niles, R. K., Pulsipher, T. C., Ransohoff, D., Rodriguez, H., Rudnick, P. a, Smith, D., Tabb, D. L., Tegeler, T. J., Variyath, A. M., Vega-Montoto, L. J., Wahlander, A., Waldemarson, S., Wang, M., Whiteaker, J. R., Zhao, L., Anderson, N. L., Fisher, S. J., Liebler, D. C., Paulovich, A. G., Regnier, F. E., Tempst, P., and Carr, S. a (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* 27, 633–641
2. Kuhn, E., Addona, T., Keshishian, H., Burgess, M., Mani, D. R., Lee, R. T., Sabatine, M. S., Gerszten, R. E., and Carr, S. a (2009) Developing multiplexed assays for troponin I and interleukin-33 in plasma by peptide immunoaffinity enrichment and targeted mass spectrometry. *Clin. Chem.* 55, 1108–17
3. Kuhn, E., Whiteaker, J. R., Mani, D. R., Jackson, A. M., Zhao, L., Pope, M. E., Smith, D., Rivera, K. D., Anderson, N. L., Skates, S. J., Pearson, T. W., Paulovich, A. G., and Carr, S. a (2012) Interlaboratory evaluation of automated, multiplexed peptide immunoaffinity enrichment coupled to multiple reaction monitoring mass spectrometry for quantifying proteins in plasma. *Mol. Cell. Proteomics* 11, M111.013854
4. Ossola, R., Schiess, R., Picotti, P., Rinner, O., Reiter, L., and Aebersold, R. (2011) in *Serum/Plasma Proteomics SE - 11, Methods in Molecular Biology.*, eds Simpson RJ, Greening DW (Humana Press), pp 179–194.
5. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* 9, 555–566
6. Liebler, D. C., and Zimmerman, L. J. (2013) Targeted quantitation of proteins by mass spectrometry. *Biochemistry* 52, 3797–3806
7. Ehardt, H. A. (2014) in *Plant Proteomics SE - 16, Methods in Molecular Biology.*, eds Jorin-Novo J V, Komatsu S, Weckwerth W, Wienkoop S (Humana Press), pp 209–222.
8. Keshishian, H., Addona, T., Burgess, M., Kuhn, E., and Carr, S. a (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* 6, 2212–2229
9. Keshishian, H., Addona, T., Burgess, M., Mani, D. R., Shi, X., Kuhn, E., Sabatine, M. S., Gerszten, R. E., and Carr, S. a (2009) Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* 8, 2339–49
10. Mani, D. R., Abbatiello, S. E., and Carr, S. a (2012) Statistical characterization of multiple-reaction monitoring mass spectrometry (MRM-MS) assays for quantitative proteomics. *BMC Bioinformatics* 13 Suppl 1, S9
11. Harlow, E., and Lane, D. (1999) *Using antibodies: a laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor). 1st Ed.

12. Anderson, N., and Anderson, N. (2004) Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* 3, 235–44
13. Whiteaker, J. R., Zhao, L., Abbatiello, S. E., Burgess, M., Kuhn, E., Lin, C., Pope, M. E., Razavi, M., Anderson, N. L., Pearson, T. W., Carr, S. a, and Paulovich, A. G. (2011) Evaluation of large scale quantitative proteomic assay development using peptide affinity-based mass spectrometry. *Mol. Cell. Proteomics* 10, M110.005645
14. Schoenherr, R. M., Zhao, L., Whiteaker, J. R., Feng, L.-C., Li, L., Liu, L., Liu, X., and Paulovich, A. G. (2010) Automated screening of monoclonal antibodies for SISCAPA assays using a magnetic bead processor and liquid chromatography-selected reaction monitoring-mass spectrometry. *J. Immunol. Methods* 353, 49–61
15. Deutsch, E. W., Lam, H., and Aebersold, R. (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* 9, 429–434
16. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–8
17. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567
18. Fusaro, V. a, Mani, D. R., Mesirov, J. P., and Carr, S. a (2009) Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat. Biotechnol.* 27, 190–198
19. Whiteaker, J. R., Zhao, L., Lin, C., Yan, P., Wang, P., and Paulovich, a. G. (2012) Sequential Multiplexed Analyte Quantification Using Peptide Immunoaffinity Enrichment Coupled to Mass Spectrometry. *Mol. Cell. Proteomics* 11, M111.015347–M111.015347

## Chapter 6

### **Targeted MS assay quantifies proteins predicting tamoxifen resistance in estrogen receptor positive breast cancer**

Tommaso De Marchi, Erik Kuhn, Lennard Dekker, Christoph Stingl, Rene B. Braakman, Mark Opdam, Sabine C. Linn, Fred C. Sweep, Paul N. Span, John A. Foekens, Theo M. Luider, John W. Martens, Steven A. Carr, Arzu Umar.

*J Proteome Res*, 2016 [epub ahead of print]

## Abstract

We recently reported on the development of a 4-protein-based classifier (PDCD4, CGN, G3BP2 and OCIAD1) capable of predicting outcome to tamoxifen treatment in recurrent, estrogen receptor positive breast cancer, based on high resolution MS data. In order to move these findings towards a clinical setting, a precise and high throughput assay to measure these proteins in a multiplexed, targeted fashion would be favorable to measure large numbers of patient samples. By coupling immuno-precipitation to multiple reaction monitoring (MRM) MS and stable isotope dilution, we developed a high precision assay to measure the 4-protein signature in 38 primary breast cancer whole tissue lysates (WTL). Furthermore, we evaluated the presence and patient stratification capabilities of our signature in an independent set of 24 matched (pre- and post-therapy) sera. We compared the performance of iMRM with direct MRM in the absence of fractionation and shotgun proteomics in combination with label-free quantification (LFQ) on both WTL and laser capture microdissected (LCM) tissues. Measurement of the 4-proteins by iMRM not only showed higher accuracy in measuring proteotypic peptides (Spearman  $r$ : 0.74 to 0.93) when compared to MRM (Spearman  $r$ : 0.0 to 0.76), but also significantly discriminated patient groups based on treatment outcome (hazard ratio [HR]: 10.96; 95% confidence interval [CI]: 4.33 to 27.76; Log-rank  $P < 0.001$ ) when compared to LCM (HR: 2.85; 95% CI: 1.24 to 6.54; Log-rank  $P = 0.013$ ) and WTL (HR: 1.16; 95% CI: 0.57 to 2.33; Log-rank  $P = 0.680$ ) LFQ-based predictors. Serum sample analysis by iMRM confirmed detection of the 4 proteins in these samples. We hereby report that iMRM outperformed regular MRM, confirmed our previous high resolution MS results in tumor tissues and have shown that the 4-protein signature is measurable in serum samples.

## Introduction

Estrogen receptor (ER) positive breast cancer constitutes three quarters of all breast malignancies, resulting in more than 200,000 cases each year in the United States <sup>1</sup>. Treatment options for these patients include anti-hormonal drugs such as tamoxifen, which effectively reduces yearly breast cancer death by 38% in a 5-year adjuvant regimen <sup>2</sup>. In the recurrent setting, tamoxifen's therapeutic efficiency is drastically reduced by either acquired or intrinsic resistance <sup>3,4</sup>. Through high resolution MS analysis on LCM derived tumor material, we have recently developed a protein signature that is able to stratify tamoxifen treated patients according to treatment outcome independent of traditional predictive parameters <sup>5</sup>. The coupling of LCM-based tumor cell enrichment with high resolution MS has demonstrated to be a robust and sensitive platform for biomarker discovery, being able to assess protein abundance in specific cell populations while minimizing interference from surrounding tissues <sup>6-8</sup>. Although global proteome analysis of LCM derived material offers a sensitive analysis of relatively pure population of cells, extensive sample preparation and relative long measurement times thwart their introduction into the clinical setting. Traditional immuno-assays constitute a viable option for biomarker verification due to the high sensitivity and specificity provided by high quality antibodies, if available, either suffer from time consuming optimization (e.g. ELISA) or non-linear protein quantitation (e.g. immunohistochemistry [IHC]).

To validate the clinical relevance of our 4-protein predictor, we used IHC as an independent technology. For PDCD4, we were able to show that it was an independent marker for tamoxifen therapy resistance, although protein quantification by IHC appeared to be difficult due to the lack of linearity of chromogenic signals. In addition, the single marker PDCD4 as measured by IHC had inferior performance (HR = 0.75; 95% CI: 0.59 to 0.96, Log-rank  $P = 0.020$ ) compared to the MS-based 4-protein predictor (HR = 2.32; 95% CI: 1.29 to 4.17; Log-rank  $P = 0.004$ ). In the light of this a quantitative targeted biomarker verification assay is needed in order to validate the utility of the candidate marker panel in a large number of patient samples. MRM coupled to stable isotope dilution provides a precise method to quantify proteins in biological specimens, has significantly shorter development and optimization times, and can approach sensitivity and reproducibility of immune-assays <sup>9,10</sup>. Moreover, MRM MS enables measurement of up to 100's of candidate markers in a single biological sample <sup>11-14</sup>. However, protein measurement can still be hindered by the extreme dynamic range of total protein content (e.g. plasma) <sup>15,16</sup>, and ultimately requires the use of a protein reference standard. In order to address this issue, several studies showed that enrichment

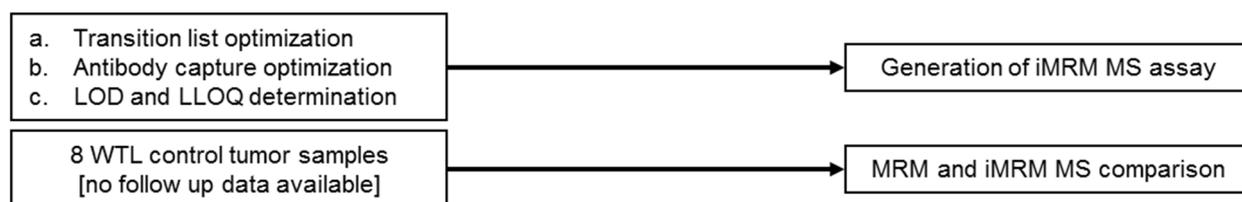
of target analytes by immuno-precipitation (or immuno-capture) allows protein detection and quantification with minimization of matrix effects in complex biological specimens<sup>17,18</sup>.

The main aim of the current study was to develop a targeted immuno-MRM (iMRM, also known as SISCAPA: Stable Isotope Standards Capture with Anti-Peptide Antibodies)<sup>17</sup> assay for our predictive 4-protein signature, that would be applicable for both breast cancer tissues and patient-derived serum samples. Here we demonstrate that target protein enrichment with anti-peptide antibodies coupled to MRM MS is capable of measuring and quantifying the 4 signature proteins in a precise and reproducible way. Furthermore, both tissue and serum-derived classifiers enabled prediction of tamoxifen treatment outcome and showed significant correlation with time to progression. The iMRM assay we developed establishes a technical verification of our 4-protein predictor and provides a platform for its potential introduction into clinical diagnostics.

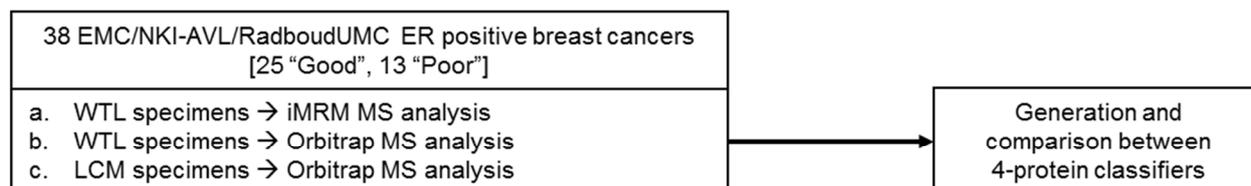
## **Experimental section**

Extensive descriptions of serum sample collection, MRM analysis of control tissues and serum samples, as well as high resolution MS analysis, are provided in supporting information (available online). Protein identification procedures were previously described<sup>5</sup>. iMRM assay development workflow, comparison between regular MRM and iMRM, and evaluation of tissue enrichment influence on protein predictor is displayed in Figure 6.1.

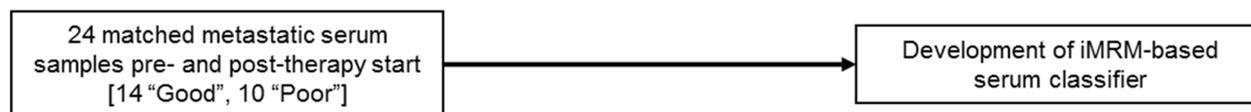
## 1. iMRM assay optimization and quality control:



## 2. Performance of 4-protein signature:



## 4. Performance of 4-protein signature in breast cancer sera:



**Figure 6.1. Schematic representation of iMRM assay development workflow.**

This study was performed in three steps. First, transition and immunoaffinity enrichment were optimized, LOD and LLOQ were determined, and transition measurement was compared between iMRM and direct MRM. Second, high resolution MS analysis of LCM and WTL specimen was compared with iMRM analysis on a cohort of 38 matched breast cancer specimens through classifier generation. Third, a second independent serum cohort was analyzed to build a serum-based classifier.

Acronyms: ER: estrogen receptor; LOD: limit of detection; LLOQ: lower limit of quantitation; MRM: multiple reaction monitoring; WTL: whole tissue lysate.

### *Patient cohorts*

In order to derive a confirmatory cohort (i.e. confirmation dataset), we selected breast cancer samples from our previously described patient cohorts<sup>5</sup>. From a total of 112 tumor tissues (training set N = 56; test set N = 56), we selected samples of which enough material (i.e. at least 100 µg of protein) was available and furthermore displayed at least 50% tumor area based on hematoxylin and eosin stained slides derived from the same tumor specimen at the moment of LCM and WTL collection. Outcome to tamoxifen therapy was defined as patients manifesting disease progression before ( $\leq$ ; poor) or after ( $>$ ; good) 6 months. This subset comprised 25 good and 13 poor patients, respectively (Table 6.1). After iMRM analysis, whole tissue lysates (WTL) of these samples were prepared and subjected to Orbitrap MS analysis. Extraction of high resolution proteomic data from matching LCM samples was performed. LCM and WTL mass spectrometry data have been

deposited to the ProteomeXchange Consortium <sup>19</sup> via the PRIDE partner repository with the dataset identifier PXD002381.

An independent set of 24 breast cancer patient-derived sera was selected to verify the detectability and therapy outcome prediction of the 4-protein signature in blood (collection dates: 1989-2000), and to assess whether the expression of the 4 protein markers would change based on tamoxifen therapy. Sera were collected both prior to start of first line tamoxifen therapy for recurrent disease (defined as “PRE”; Table S-1), and at the end of first line treatment (defined as “POST”), which coincided with disease progression. This set comprised 14 good and 10 poor outcome patients. Both cohorts were analyzed by iMRM as biological replicates.

In order to assess the efficacy of immuno-capture coupled to MRM MS, a set of 8 ER positive control tumor specimens ( $\geq 50\%$  tumor area) was analyzed in triplicate both by direct MRM (no fractionation) and by iMRM.

This study was approved by the local institutional medical ethics committee (MEC 02.953).

**Table 6.1. Clinical information of breast cancer patients included in the confirmation cohort.**

	Outcome group		
	Good*	Poor*	Total*
Patients	25 (100.0)	13 (100.0)	38 (100.0)
Age			
≤ 55 years	3 (12.0)	4 (30.8)	7 (18.4)
> 55 years	22 (88.0)	9 (69.2)	31 (81.6)
Menopausal status			
Premenopausal	2 (8.0)	3 (23.1)	5 (13.2)
Postmenopausal	23 (92.0)	10 (76.9)	33 (86.8)
Tumor size			
T1 (≤ 2cm)	7 (28.0)	3 (23.1)	10 (26.3)
T2 (2-5cm) + Tx	17 (68.0)	9 (69.2)	26 (68.4)
T3 (> 5cm) + T4	1 (4.0)	1 (7.7)	2 (5.3)
Tumor differentiation**			
Good/Moderate	12 (48.0)	4 (30.8)	16 (42.1)
Poor	9 (36.0)	8 (61.5)	17 (44.7)
Unknown	4 (16.0)	1 (7.7)	5 (13.2)
Disease free interval			
≤ 12 months	4 (16.0)	4 (30.8)	8 (21.1)
> 12 months	21 (84.0)	9 (69.2)	30 (78.9)
PgR†			
Negative	11 (44.0)	8 (61.5)	19 (50.0)
Positive	13 (52.0)	5 (38.5)	18 (47.4)
Involved lymph nodes			
0	11 (44.0)	6 (46.2)	17 (44.7)
≥ 1	13 (52.0)	7 (53.8)	20 (52.7)
unknown	1 (4.0)	0 (0.0)	1 (2.6)
Dominant site of relapse			
Loco-regional	2 (8.0)	3 (23.1)	5 (13.2)
Bone	5 (20.0)	4 (30.8)	9 (23.7)
Visceral	7 (28.0)	4 (30.8)	11 (28.9)
Bone and other	8 (32.0)	2 (15.3)	10 (26.3)
Unknown	3 (12.0)	0 (0.0)	3 (7.9)

\* Data are reported as number (percentage).

\*\* Histopathological characteristics were evaluated by local pathologists, according to standard clinical practice at time of sample collection.

† Missing data not reported.

Acronym: PgR: progesterone receptor

### *Cell culture*

MCF-7 breast cancer cell line was cultured in RPMI medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) and antibiotic agents (100 µg/mL Penicillin, 20 ng/mL Streptomycin and 80 µg/mL Gentamycin) in a humidified atmosphere with 95% air and 5% CO<sub>2</sub>. Cells were gathered from culture dishes using a scraper, suspended in PBS + 0.03% CHAPS (Sigma), sonicated at 70% amplitude (Bransons Ultrasonics, Danbury, CT, USA) using a horn sonifier, and collected supernatants were transferred into a new tube and stored at -80°C for further processing.

### *Clinical samples preparation*

A total of 10 sections of 10 µm were cut from all ER+ frozen tumor specimens, and processed according to our tissue proteomic workflow, as previously described <sup>7</sup>. Tissue sections were collected in pre-frozen protein Lo-bind (Eppendorf) tubes and stored at -80°C for further processing. Prior protein digestion, tissue sections were suspended in 100 µL 0.1% w/v Rapigest surfactant (Waters) in 50 mM ammonium bicarbonate solution and sonicated using an Ultrasonic Disruptor Sonifier II (Bransons Ultrasonics). Amplitude was set at 70%. Sonicated tissues were then centrifuged for 15 min at 14,000 g and supernatants were collected and transferred to a new tube.

### *Protein digestion*

Total protein concentration was measured by bi-cinchoninic acid assay in extracts of breast cancer tissues and MCF7 cells in order to determine protein concentration. A fixed volume of serum and plasma (i.e. 50 µL) samples was instead digested. Breast cancer sera, MCF-7 cells, and healthy donor plasma and serum proteins were denatured and reduced in 10 M urea and 5 mM DTT, respectively, and subsequently alkylated in the dark with a 15 mM of iodoacetamide solution for 30 min at room temperature. Samples were then digested overnight at 37°C by adding trypsin in a 1:50 enzyme-protein ratio after dilution of Urea with Trizma® (Sigma-Aldrich, Steinheim, Germany).

Samples were acidified with formic acid, spun down at 14,000 RPM, de-salted through OASIS® (Waters Corporation, Milford, MA, USA) cartridges and vacuum-dried.

#### *Antibodies and isotope labeled peptide standards*

Anti-peptide antibodies against unique peptides of PDCD4 (SGLTVPTSPK and DLPELALDTPR), Cingulin (VQGIAGQPFFVVLNSGEK and LGQEQQTLNR), OCIAD1 (GILSSHPK and LENSPLGEALR) and G3BP2 (VEAKPEVQSQPPR and LPNFGFVVFDDSEPVR) proteins and isotope-labeled versions of unique peptides (labeled with <sup>15</sup>N/<sup>13</sup>C at Arg and Lys C-terminal residues) were purchased from New England Peptide (Gardner, MA, USA).

#### *Selection of predictor protein specific transitions*

Based on high resolution MS data (previously deposited to the ProteomeXchange consortium with identifiers PXD000484 and PXD000485) and in silico digestion of PDCD4, OCIAD1, G3BP2 and CGN canonical isoforms through SRM Collider software in the Skyline environment <sup>20</sup> (version 3.1.0; <https://brendanx-uw1.gs.washington.edu/labkey/project/home/software/Skyline/begin.view>), a peptide library for the 4 predictor proteins was generated. Peptides were selected based on the following features: uniqueness of their amino acid sequence in proteome (i.e., being “proteotypic”), full tryptic peptides (no missed cleavage), absence of methionine residues, exclusion of ragged ends (cleavage between R-R, R-K, K-K and K-R), and exclusion of low abundance peptides present in full scan of previous Orbitrap MS runs. These rules were applied except when too few eligible peptides per protein (e.g. 1) were derived. Verification of selected transitions was performed by injection of 10 µL of a 5 fmol/µL heavy peptide mix solution in triplicate. Optimization for collision energy was calculated in Skyline software for every filtered transition.

#### *Antibody cross-linking*

Each anti-peptide antibody was incubated overnight with Protein G Dynabeads® (Life Technologies, Foster City, CA, USA; Ab-bead ratio: 1:2) in a PBS/0.03% CHAPS solution at 4°C. After antibody capture by Protein G, antibody-beads solutions were aliquoted on a Kingfisher magnetic particle handler plate (Thermo-Fisher, Waltham, MA). PBS based buffer was removed and a cross-linking solution containing 20 mM dimethyl pimelimidate (DMP) and 200mM

triethanolamine (Sigma-Aldrich) was added, and antibody-beads were incubated for 30 minutes under mild shaking. DMP containing solution was then removed after cross-linking reaction was quenched by adding a 150 mM monoethanolamine solution. Antibody-beads were then washed twice with a 3% Acetic Acid/0.03% CHAPS solution and once with a PBS/0.03% CHAPS solution. Antibodies cross-linked to beads were re-suspended in a PBS/0.03% CHAPS + 0.01% Sodium Azide, and stored at 4°C for further processing.

#### *Immunoaffinity enrichment of target peptides*

Immunoaffinity enrichment was performed according to our previously described protocol <sup>21</sup>. Optimal antibody concentrations were determined by incubating 10 µL of digested healthy donor plasma and heavy peptide mix (200 fmol per peptide) with increasing concentrations of antibody-beads solution. Concentration of each antibody per solution was: 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg, 2.0 µg, 4.0 µg. Triplicate enrichments for each antibody concentration were performed. In order to determine the overall sensitivity in detection (i.e. limit of detection [LOD]) and quantitation (i.e. lower limit of quantitation [LLOQ]) of iMRM and direct MRM assays, heavy peptide response curves were prepared in a background of 100 µg lysed digested MCF-7 breast cancer cells (for tissue measurements) or 50µL healthy donor serum (for serum measurements; derived from our local serum bank) in triplicate, generating a 8 concentration-point reverse curve: 200.00, 66.67, 22.22, 7.41, 2.47, 0.82, 0.27, and 0.00 fmol. The iMRM reverse curves were immunoaffinity enriched using the optimized antibody amount determined by the titration experiments, while no enrichment was performed for the direct MRM curve. When measuring breast cancer tissues, an amount of heavy peptide was added to keep the ratio of light endogenous signal and heavy peptide signal close to 1:1 as possible to reduce dynamic range effects <sup>18</sup> (tissue: 200 fmol; plasma: 10 fmol). Trypsin digested breast cancer whole tissue lysates (~100µg) and sera (~50µL) were immunoaffinity enriched by overnight incubation with cross-linked antibody magnetic beads at 4C. Bound peptides were washed with PBS/0.03% CHAPS for 3 x 5 minutes, eluted in 25 µL of 3% acetonitrile/5% AcOH, and stored at 4°C until targeted MS analysis. In order to compare sensitivity, a reverse curve was also generated for direct MRM using MCF-7 lysates.

### *iMRM MS analysis*

Healthy donor plasma, the 38 tumor tissue samples, and MCF7 cell line samples were analyzed on a 4000 Q-Trap MS system, which was coupled online to a Tempo liquid chromatography (LC) system (Applied Biosystems, Foster City, CA). Peptides were eluted with a binary gradient (flow: 300 nl/min; mobile phase A: 0.1% formic acid in H<sub>2</sub>O; mobile phase B: 90% acetonitrile and 0.1% formic acid). Sample injection was performed on PicoFrit columns (inner diameter: 75 µm; New Objective, Woburn, MA) packed in-house with ReproSil reversed phase resin (C18-AQ; diameter: 3 µm; Dr. Maisch, GmbH) for a total column length of 10-12 cm. Gradient was run as follows: 3 to 20% of solvent B for 3 minutes, 20 to 55% solvent B for 35 minutes, and 55 to 80% solvent B for 3 minutes. Ion spray voltage was set at 2200 V, curtain and nebulizer gasses were set at respectively 20 and 3 psi. Optimized declustering potential (DP) and collision energy (CE) were used for each light/heavy peptide pair. Of each target peptides, three transitions were monitored and analyzed by MRM-MS in unscheduled mode.

### *Analysis of Multiple Reaction Monitoring data*

Analyst .wiff (tissues) and MassLynx .RAW (sera) derived MRM data files were imported and analyzed in Skyline free software. Transition lists of proteotypic peptides as measured in breast cancer primary tumor tissues and sera, along with .wiff and .RAW files have been deposited in PASSEL with dataset identifier: PASS00710<sup>22</sup>. Chromatographic peak areas of extracted ion chromatograms (XIC) of each light and heavy peptide transition were used to assess peptide abundances in antibody-beads titration experiments. Peak area values for each transition were plotted in Microsoft Excel for each antibody-beads concentration point. Peak area ratio (PAR) between the area of light and heavy peptides were used for quantitation in the analysis of breast cancer samples. Isotope dilution measurements were analyzed in QuaSAR loaded as an external tool function and performed with Skyline. LOD, LLOQ and measurement precision (coefficient of variation; CV) were calculated for each measured transition in QuaSAR (v 1.1; <https://skyline.gs.washington.edu/labkey/skyts/home/software/Skyline/tools/details.view?name=QuaSAR>).

### *Statistical analysis*

iMRM-derived PARs and high resolution MS label free quantitation (LFQ) intensities were imported in Microsoft Excel and Log<sub>10</sub> transformed. Differences in peptide/protein abundance between good and poor outcome patients were assessed by t test (unequal variances assumed), while correlation between proteotypic peptide levels were assessed by Spearman correlation. Every peptide score (t value) was multiplied by its quantitative value (i.e. Log<sub>10</sub> LFQ intensity or Log<sub>10</sub> PAR), and values were summed for all proteins to obtain a patient score. The so derived scores were then transformed into a Z distribution (i.e. Z-score) and plotted in a receiver operating characteristic (ROC) curve. Youden maximum (i.e.  $J = \text{sensitivity} + \text{specificity} - 1$ ) was chosen as cutoff for group prediction. Survival curves were plotted for each predicted group. Time to progression (TTP) was taken as endpoint and differences in survival curves were assessed by Log-rank and Gehan-Breslow-Wilcoxon tests. T test (paired and unpaired), Spearman correlation, ROC and survival curves were generated in GraphPad (v 5.1).

## **Results**

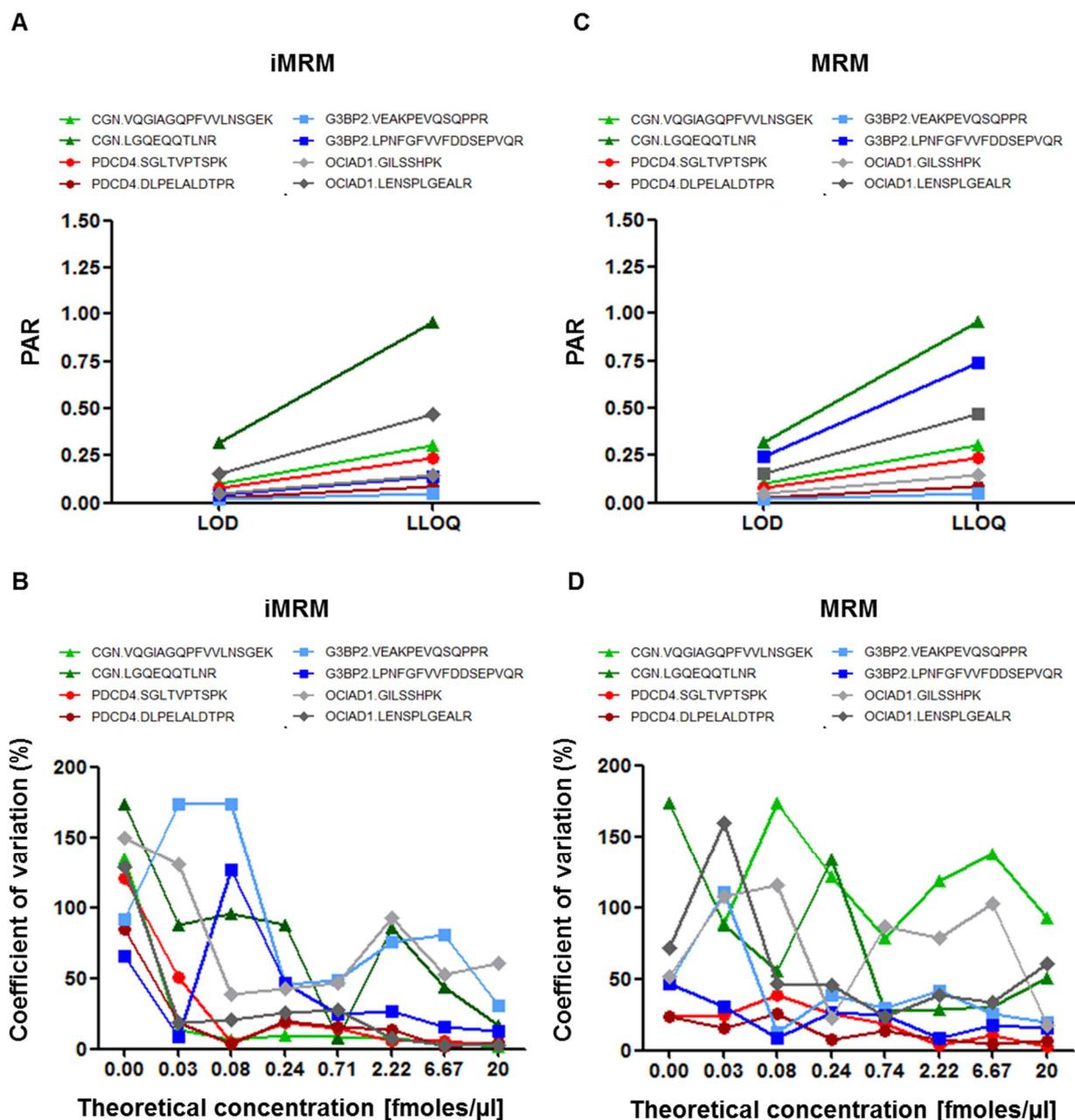
The outline of this study is schematically represented in Figure 6.1 and encompassed the development and optimization of a targeted iMRM assay for our 4-protein signature predicting tamoxifen treatment outcome of breast cancer patients, both in tissue and serum.

### *Selection of proteotypic peptides and antibody-capture optimization*

In order to maximize quantitation precision, the 3 most intense transitions for each peptide were monitored, which consisted of y and b ions. Next, in order to determine the best antibody concentration for purification of endogenous and isotope-labeled proteotypic peptides, a series of immunoaffinity enrichments was performed on heavy peptide mixes added to a background of healthy donor plasma using antibodies (scaling concentrations: 0.125 to 4 µg/capture) cross-linked to magnetic beads. The MS intensity maximum for every heavy peptide transition at the lowest antibody amount per capture was selected as optimal concentration for further immuno-captures (Table S-2).

### *Determination of LOD and LLOQ*

We subsequently investigated the performance of our iMRM assay compared to direct MRM (no fractionation, no enrichment). Standard dilutions of heavy peptides were directly injected (MRM) or captured (iMRM) in a constant background of MCF-7 digested lysates (i.e. 1  $\mu$ g [MRM]; 100  $\mu$ g [iMRM]). From the analysis of the MCF-7-based reverse curves, XIC of LENSPLGEALR, SGLTVPTSPK, DLPELALDTPR, VEAKEVQSQPPR and VQGIAGQPFVVLNSGEK peptides displayed high intensities. Lower intensity was displayed by LGQEQQTLNR, LPNFGFVVFDDSEPVQR and GILSSHPK peptides, probably due to early elution times, poor antibody capture, or poor ionization efficiency (data not shown). The ratio of heavy and light peptide peak area (i.e. MCF-7 derived) was plotted against the theoretical heavy peptide concentration for each of the proteotypic peptides in QuaSAR. Response curves for each peptide measured by iMRM in MCF-7 background are displayed in Figure S-1A-H. Median LOD and LLOQ for the 8 proteotypic peptides were 0.104 and 0.131 PAR, respectively (Figure 6.2A; Table 6.2). Coefficients of variation for every peptide were calculated at each concentration point and peptide CVs for each concentration point were plotted (Figure 6.2B). Median CVs across the 8 concentration points for at least 4 peptides across the reverse curve was below 20%, suggesting acceptable reproducibility of the assay (Table S-3). The fact that the remaining peptides showed higher CVs may derive from very low endogenous abundance levels, which may have hampered quantitation.



**Figure 6.2. Limit of detection and quantitation of iMRM and MRM assays.**

Reverse concentration curve out of iMRM (panel A and B) and direct MRM (panel C and D) measurements were analyzed in QuaSAR. LOD and LLOQ were calculated for every peptide transition, and plotted next to each other (A and C). Coefficients of variation were also calculated for each peptide transition and their distribution was plotted per each concentration point in the reverse curve (B and D). Displayed peptides: VQGIAGQPFVVLNSGEK (CGN, green), LGQEQQTLNR (CGN, dark green), SGLTVPTSPK (PDCD4, red), DLPELALDTPR (PDCD4, dark red), VEAKPEVQSQPPR (G3BP2, light blue), LPNFGFVVFDDSEPVQR (G3BP2, dark green), GILSSHPK (OCIAD1, grey), LENSPLGEALR (OCIAD1, dark grey).

**Table 6.2. Limit of detection and lower limit of quantitation for every transition in iMRM experiments in a background of MCF-7 lysate.**

Uniprot ID	Protein Name	Gene name	Peptide sequence*	AA	Light precursor ion m/z	Light product ion m/z	Heavy precursor ion m/z	Heavy product ion m/z	Tissue		Serum	
									LOD [PAR]	LLOQ [PAR]	LOD [PAR]	LLOQ [PAR]
Q9P2M7	Cingulin	CGN	VQGIAGQPF	50-66	871.975++	654.357+	875.982++	654.357+	0.248	0.744	6.912	20.737
						1345.711+		1353.725+	0.127	0.381	6.990	20.971
						1274.673+		1282.688+	0.106	0.319	9.616	28.848
						1089.593+		1097.608+	0.102	0.307	2.672	8.018
						746.404+		754.418+	0.138	0.416	6.166	18.497
Q9P2M7	Cingulin	CGN	LGQEQQTL	805-14	593.812++	898.462+	598.816++	898.462+	n/a	n/a	n/a	n/a
						888.453+		898.461+	0.319	0.957	n/a	n/a
						759.410+		769.419+	0.464	1.393	n/a	n/a
						631.352+		641.360+	0.556	1.670	n/a	n/a
						503.293+		513.301+	0.592	1.778	2.083	6.249
						508.759++		513.763+	0.504	1.514	82.063	246.191
Q53EL6	Programmed	PDCD	SGLTVPTSP	86-95	493.779++	729.414+	497.786++	737.428+	0.078	0.235	1.015	3.047
						628.366+		636.380+	0.101	0.304	1.764	5.292
						529.298+		537.312+	0.099	0.299	1.326	3.979
Q53EL6	Programmed	PDCD	DLPELALDT	245-55	620.332++	672.367+	625.336++	682.375+	0.031	0.093	4.065	12.195
						601.330+		611.338+	0.028	0.086	6.133	18.401
						506.277++		511.281+	0.033	0.100	3.035	9.106
Q9UN86	Ras GTPase-	G3BP	VEAKPEVQ	277-89	488.931+++	753.414+	492.267+++	753.414+	0.076	0.230	n/a	n/a
						584.315+		594.323+	0.024	0.074	47.439	142.316
						683.359++		688.363+	0.017	0.051	n/a	n/a
						618.838++		623.842+	0.130	0.392	n/a	n/a
Q9UN86	Ras GTPase-	G3BP	LPNFGFVVF	370-86	983.488++	1290.632+	988.493++	1300.640+	0.075	0.227	n/a	n/a
						1191.564+		1201.572+	0.054	0.164	3.882	11.647
						1092.495+		1102.503+	0.057	0.171	5.389	16.168
						499.298+		509.307+	0.045	0.137	17.986	53.959
Q9NX40	OCIA	OCIA	GILSSHPK	64-71	419.742++	668.372+	423.749++	676.386+	0.365	1.097	n/a	n/a
						555.288+		563.302+	0.049	0.148	n/a	n/a
						334.689++		338.697+	0.150	0.452	n/a	n/a

Q9NX40	OCIA	OCIA	LENSPLGEA	104-11	599.825++	956.516+	604.829++	966.524+	0.241	0.723	n/a	n/a
						842.473+		852.481+	0.155	0.467	n/a	n/a
						755.441+		765.449+	0.229	0.687	169.0123	507.037

\*Aminoacids labeled with 13C and 15N are highlighted in bold.

Acronyms: AA: aminoacid sequence; LOD: limit of detection; LLOQ: lower limit of quantitation

Using the same method, we assessed performance of direct MRM (response curves are displayed in Figure S-2A-H). Direct MRM median LOD and LLOQ for the 8 peptides were 0.233 and 0.699 PAR, respectively (Figure 6.2C; Table S-4), showing a slight decrease in sensitivity when compared to iMRM. Upon assessing variability of direct MRM measurements, 5 out of 8 peptides showed a higher median CV compared to their counterparts analyzed by iMRM, while the 3 remaining peptides showed comparable variation (Figure 6.2D; Table S-5). These data suggest that iMRM suffers less variability and achieves higher sensitivity in comparison to direct MRM.

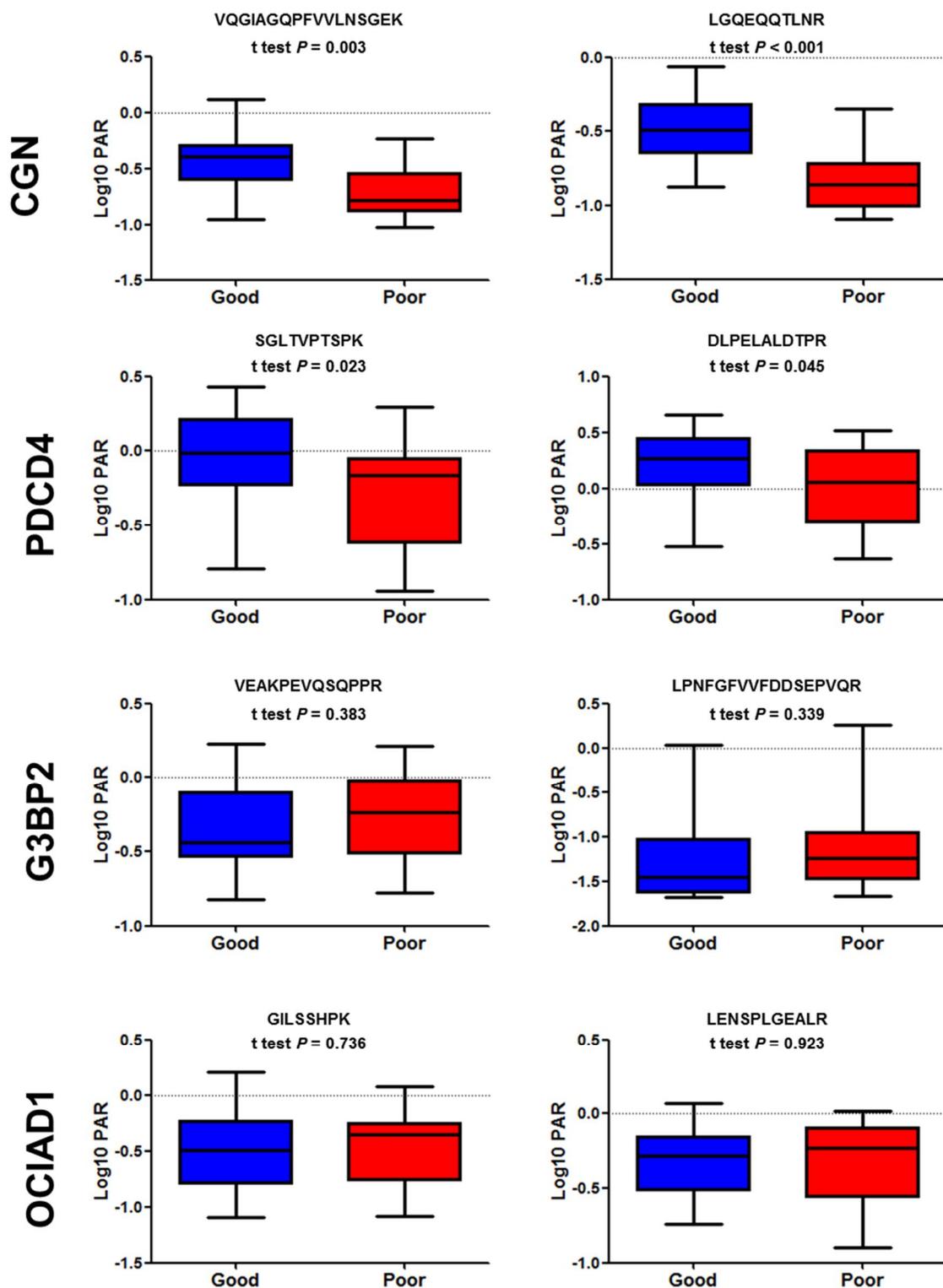
For serum analysis, a reverse curve in a background of healthy donor serum (i.e. 50  $\mu$ L; ~3 mg) was also generated and analyzed by iMRM. Response curves of proteotypic peptides measured in a background of healthy donor serum by iMRM are displayed in Figure S-3 (panels A to H). LOD and LLOQ in serum (Median [min – max] LOD: 3.035 [0.846 – 160.842] PAR; Median [min – max] LLOQ: 9.106 [2.539 – 482.525] PAR) were higher than in breast cancer tissues (Median [min – max] LOD: 0.064 [0.017 – 0.319] PAR; Median [min – max] LLOQ: 0.192 [0.051 – 0.957] PAR; Figure S-4A and Table 2). A higher CV was also observed in the serum-based reverse curve (Figure S-4B and Table S-3). Given the fact that we measured tissue proteins in serum, it is not surprising that LOD, LLOQ, and CV were higher than in cell line lysates.

#### *Comparison between MRM and iMRM*

In order to confirm whether immunoaffinity enrichment of proteotypic peptides would yield more reliable data we tested precision and concordance of measurement between MRM and iMRM approaches in a cohort of breast cancer tissues (analyzed in triplicate). Correlation analysis between iMRM and MRM analysis of proteotypic peptides showed that for 5 out of 8 peptides moderate to high correlation coefficients were observed, while only 3 (i.e. both OCIAD1 peptides and G3BP2 VEAQPEVQSQPPR) peptides showed poor or no correlation (Figure S-5). Furthermore, when assessing correlation between protein-specific peptides, iMRM measurements (Spearman  $r$  range: 0.74 to 0.93) proved to be more precise than standard MRM (Spearman  $r$  range: 0.00 to 0.76; Figure S-6). These data suggest that the iMRM assay for our 4-protein predictor is a more robust and reliable quantitation method than direct MRM.

### *iMRM analysis of breast cancer tissues*

After having optimized the iMRM assay for our 4-protein predictor and comparison with direct MRM, it was assessed whether proteotypic peptide levels measured by iMRM in breast cancer patient-derived tissues would show differential levels between patient groups in our confirmation dataset, and whether a classifier for patient outcome could be developed. Immuno-captures of the 8 proteotypic peptides were performed on a series of 38 ER positive breast cancer WTLs with a stable amount of isotope labeled peptides (i.e. 200 fmol). A significant difference in abundance was detected for CGN (VQGIAGQPFVVLNSGEK t test  $P = 0.003$ ; LGQEQQTLNR t test  $P < 0.001$ ), and PDCD4 (SGLTVPTSPK t test  $P = 0.023$ ; DLPELALDTPR t test  $P = 0.045$ ) between poor and good outcome patients (Figure 6.3). Furthermore, these peptides showed strong positive correlation, (Spearman  $r = 0.95$ , both peptides). Peptides measured for G3BP2 and OCIAD1 showed a moderate (Spearman  $r = 0.50$ ) and good (Spearman  $r = 0.79$ ) positive correlation respectively, but their abundance was not statistically different between patient groups (Figure 6.3 and Figure S-7). These data suggested that the 4-protein signature measured by iMRM MS may have predictive value in breast cancer tissues.



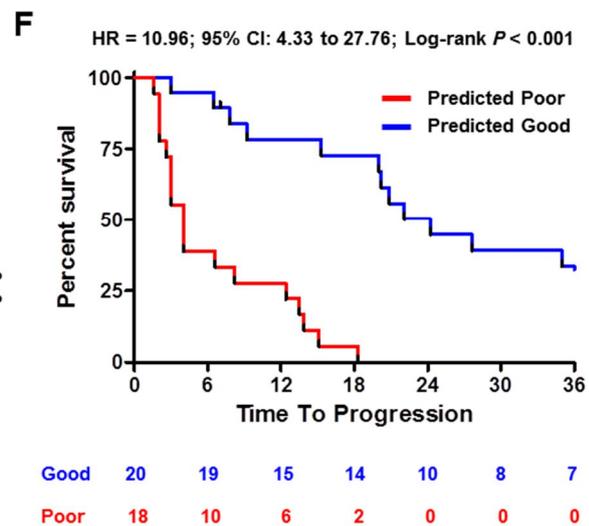
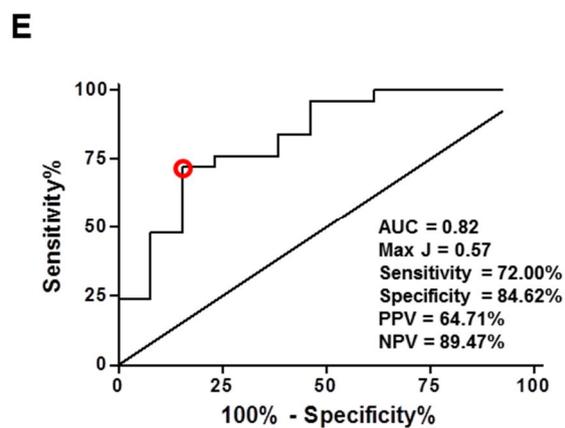
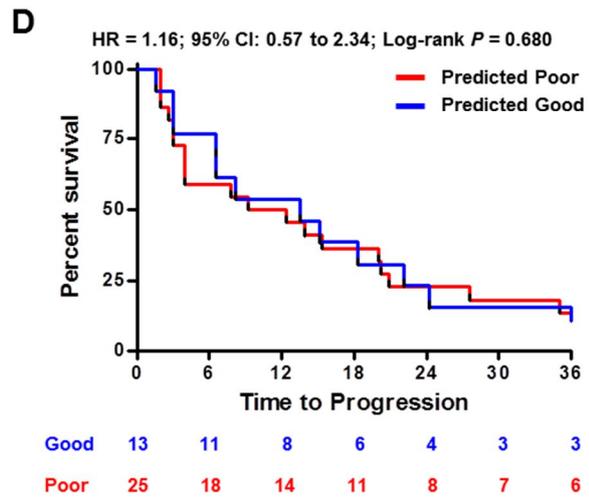
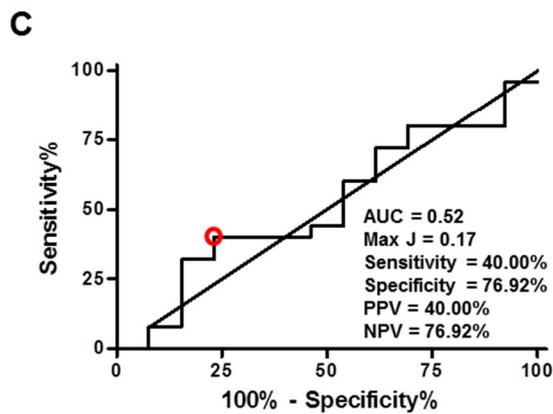
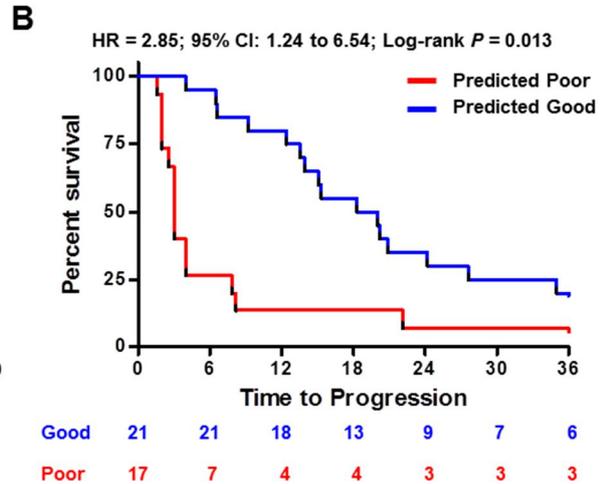
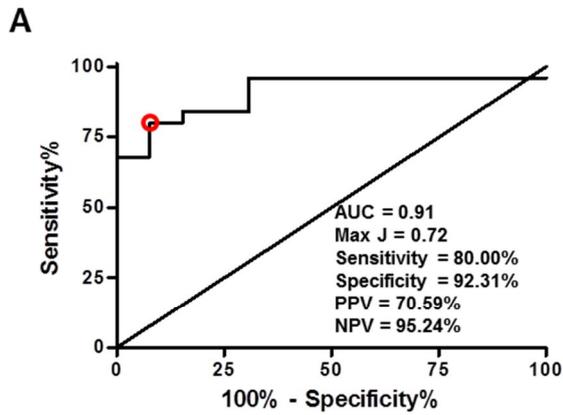
**Figure 6.3. Quantification of 4 protein signature in breast cancer tissue cohort.**

iMRM MS measurements were performed on 38 ER positive breast cancer tissues and Log<sub>10</sub> PARs were plotted for good and poor outcome patients. Difference in expression levels between the two patient groups was assessed by t test. Significant difference was found between the levels of VQGIAGQPFVVLNSGEK and LGQEQQTLNR CGN peptides and for the SGLTVPTSPK PDCD4 peptide between good and poor outcome patients.

### *Patient stratification according to global shotgun versus targeted iMRM analysis*

In order to compare the 4-protein predictor based on targeted iMRM with that developed by global shotgun MS data, we measured the 38 WTL samples by Orbitrap MS, and quantified protein abundance through LFQ (information on protein identification in LCM and WTL sets are provided in Table S-6). In addition, we extracted previously generated LCM Orbitrap MS data 5. We evaluated the 4-protein classifier prediction capability by plotting ROC curves of the protein levels obtained through global shotgun (LCM and WTL samples) and targeted iMRM MS (WTL). The area under the curve (AUC) of the ROC curve built based on LCM global shotgun MS data displayed the highest value (AUC = 0.91; Figure 6.4A), discriminating poor outcome patients with 92.31% sensitivity and 80.00% specificity (max J = 0.72). Predicted treatment outcome groups also showed significant difference in TTP (HR = 2.85; 95% CI: 1.24 to 6.54; Log-rank  $P = 0.013$ ; Figure 6.4B). The ROC curve based on WTL global shotgun MS data displayed no discriminating power (AUC = 0.52) and low Youden index (max J = 0.17; Sensitivity = 76.92%; Specificity = 40.00%; Figure 6.4C), with no significant difference between survival curves of predicted groups (HR = 1.16; 95% CI: 0.57 to 2.34; Log-rank  $P = 0.680$ ; Figure 6.4D). Thus, when applying global shotgun proteomics, tumor tissue enrichment through LCM appears to be essential to obtain enough discrimination and reliable outcome prediction based on the 4-proteins. On the other hand, discriminating power was retained by iMRM MS of signature peptides derived from WTLs. In fact, the ROC curve-based these data displayed a comparable AUC (0.82) and group prediction (max J = 0.57; sensitivity = 84.62%; specificity = 72.00%; Figure 6.4E) as the one derived from high resolution MS of LCM samples, reflective of the high sensitivity and interference reduction inherent to the technique. Also, a significant difference was observed between survival curves of the two groups (HR = 10.96; 95% CI: 4.33 to 27.76; Log-rank  $P < 0.001$ ; Figure 6.4F). Due to the fact that OCIAD1 and G3BP2 peptides did not show a significant difference between patient groups, we argued whether their removal from our predictive model based on iMRM data would benefit overall prediction. The ROC curve based solely on PDCD4 and CGN peptides showed a similar AUC (i.e. 0.85) when compared to the one derived from the 4 protein predictive model, though with notable differences in sensitivity (from 86.62% to 56.00%) and specificity (from 72.00% to 100.00%; Figure S-8A). Although significant, a decrease in survival differences between patient predicted groups was also observed (HR = 2.10; 95% CI: 1.04 to 4.23; Log-rank  $P = 0.037$ ; Figure S-8B). This suggests that, although

not retaining their t test significance, OCIAD1 and G3BP2 still contribute to the predictive model.



**Figure 6.4. Group prediction based on high resolution and iMRM MS data.**

Next to the iMRM analysis of 38 tumor tissues, high resolution MS data was generated for WTL and previously obtained for LCM samples from those same 38 tissues. ROC curves were built based on the levels of the 4 signature proteins (A, C, E), and group prediction was compared (B, D, F) for LCM shotgun (A, B), WTL shotgun (C, D), and WTL iMRM (E, F) data.

Acronyms: CI: confidence interval; HR: hazard ratio; ROC: receiver operating characteristic, PPV: percentage of positive predicted values; NPV: percentage of negative predicted values.

Overall, these data suggest that iMRM measurement of predictor proteins on breast cancer WTL allows discrimination of patient groups and prediction of treatment outcome comparable to high resolution MS on LCM enriched specimens.

*iMRM analysis of breast cancer patient sera*

Although our iMRM assay for the 4-protein signature performed well in tissues, blood specimens are more readily available in the clinic. In this perspective, successful prediction of tamoxifen outcome based on the 4-protein measurement in patient sera would constitute a more clinically applicable assay. We therefore assessed whether peptides from the 4-signature proteins were also detectable in breast cancer patient-derived serum. iMRM analysis was performed on patient sera obtained before (PRE) and after (POST) start of tamoxifen treatment for recurrent disease. All 8 signature peptides were detected in the patient serum cohort, although the amount of each protein was significantly reduced relative to the levels observed in whole breast cancer tissue lysates (data not shown) and for some samples few observations were missing. Upon assessment of differences in expression levels between good and poor outcome patients in the “PRE” serum cohort, trends were observed for CGN VQGIAGQPFVVLNSGEK (t test  $P = 0.087$ ) and LGQEQQLNR (t test  $P = 0.054$ ) peptides, while only G3BP2 peptide LPNFGFVVFDDSEPVQR showed high expression in good outcome patients (t test  $P < 0.001$ ; Figure S-9), which was the opposite of what was observed in breast cancer tissues. Despite of this, levels of the remaining 7 peptides showed similar expression pattern in patient groups when compared to breast cancer tissues. Correlation analysis between proteotypic peptides revealed that PDCD4 peptides showed the highest correlation (Spearman  $r = 0.70$ ), while CGN (Spearman  $r = 0.43$ ) and G3BP2 (Spearman  $r = 0.35$ ) showed moderate correlation, and OCIAD1 peptides displayed a negative one (Spearman  $r = -0.38$ ; Figure S-10). For proper prediction comparison with tissue data, we

used the 8 proteotypic peptides expression levels to build a serum-classifier, which predicted poor outcome patients with 83.33% sensitivity and 78.57% specificity (AUC: 0.83; max J = 0.62; Figure S-11A). Survival curves of predicted patient groups did not show any significant difference by Log-rank test (HR = 1.40; 95% CI: 0.58 to 3.35; Log-rank  $P = 0.449$ ), however a significant difference in survival was observed within the first 12 months after start of tamoxifen therapy (Gehan-Breslow-Wilcoxon  $P = 0.042$ ; Figure S-11B) 23,24. Upon evaluation of the survival curves within the first year after therapy start, i.e. censoring of progression events after [ $>$ ] 12 months after start of therapy, a significant difference in time to progression was observed between predicted outcome groups (HR = 4.17; 95% CI: 1.42 to 12.25; Log-rank  $P = 0.009$ ; Figure S-11C).

As the 4-protein signature significantly predicted patient groups after iMRM analysis of primary tumors, we expected that significant prediction would be achieved in sera after measurement of the PRE cohort, and not in the POST set. In the latter, no significant difference between patient groups was indeed observed (Figure S-12). Paired analysis for each peptide between PRE and POST sera revealed an increase in CGN VQGIAGQPFVVLNSGEK (paired t test  $P = 0.003$ ) and OCIAD1 GILSSHPK (paired t test  $P = 0.023$ ) peptides, while a decrease in expression of G3BP2 LPNFGFVVFDDSEPVQR (paired t test  $P < 0.001$ ) was observed. All other peptides showed no significant change in expression levels between the two time-points (Figure S-13). Furthermore, the survival curves of predicted tamoxifen outcome groups confirmed that no significant prediction could be made on treatment outcome (Figure S-14). Thus, our data suggest that significant prediction can be achieved only by analysis of serum specimen collected prior to start of therapy, and that, as expected, analysis of serum samples collected at the end of therapy cannot be used for therapy monitoring purposes.

## **Discussion**

Tamoxifen-based anti-hormonal therapy had proven successful in treating patients with primary ER positive breast cancer, but its efficacy is drastically reduced in the recurrent setting due to intrinsic and acquired resistance mechanisms<sup>4,25</sup>. We have recently developed a protein-based classifier for tamoxifen resistance in metastatic breast cancer by using LCM cell enrichment coupled to high resolution MS and LFQ<sup>5</sup>. The next step would be large-scale clinical validation in order to introduce the classifier into clinical practice. Ideally, such large

scale validation would be performed using a multiplexed, targeted, more quantitative, and relatively quick assay. Our initial workflow based on LCM, global shotgun MS and LFQ, however, requires extensive and time consuming sample preparation, provides only relative quantitation, and suffers from long measurement times. Furthermore, IHC analysis of our signature proteins provided verification of only one marker (i.e. PDCD4), probably due to poor antibody sensitivity and/or limited dynamic range inherent to the technique. We therefore developed a targeted MRM assay using stable isotope standards for quantitation, which offers robust quantitation of target analytes, however measurement can be thwarted by matrix effects that reduce sensitivity and reproducibility and can lead to false positives<sup>26,27</sup>. In order to address this issue, and to bypass the laborious LCM procedure of which the material yield is not sufficient for target analyte enrichment strategies, immuno-precipitation of target peptides from whole tissues with MRM MS were coupled to quantify our 4 signature proteins in both breast cancer tissue and serum cohorts, and iMRM-based tissue and serum classifiers were built.

Upon assessing LOD and LLOQ of our iMRM assay in breast cancer tissue and serum samples, a notable difference was detected in these two sets. While proteotypic peptides measured in breast cancer tissues displayed low LODs and LLOQs, sensitivity was decreased in serum. A similar trend was observed upon CV measurement, which was higher for peptides measured in serum. As CV generally increased at lower analyte concentrations in both the MCF-7 and healthy donor serum reverse curves, the lower concentration of endogenous peptides in the latter might have contributed to this observation. The addition of unlabeled synthetic peptides might have yielded better results in terms of precision, but would have not reflected the imprecision of measuring endogenous peptides in a biological matrix.

In order to evaluate what the added value the peptide immune-capture step has relative to direct MRM, we compared the results of the two methods in a set of control samples. Direct MRM quantification of proteotypic peptides showed to have higher detection and quantitation limits compared to iMRM, though variability varied from peptide to peptide. When comparing the two approaches in a cohort of breast cancer tissues, low correlation coefficients between proteotypic peptides measured by direct MRM were observed, especially between OCIAD1 and G3BP2 peptides. While high hydrophilicity (as observed for CGN LGQEQQTLNR, G3BP2 VEAQPEVQSQPPR, and OCIAD1 GILSSHYPK peptides) may impair quantitation due to early species elution, our data suggested possible interference from the sample matrix, like the presence of co-eluting peptides, as the responsible cause for such

discrepancy in the comparative measurement of OCIAD1 and G3BP2 by iMRM and MRM MS. This was further confirmed when higher concordance was found between proteotypic peptides measured by iMRM. In this perspective, iMRM would offer a valuable tool for laboratory and clinical assays due to the minimization of interfering compounds and proteins and the consequential precision in analyte measurement.

Upon assessing peptide abundance differences between patient groups in the primary breast cancer tissue confirmatory dataset, only PDCD4 and CGN peptides were found differentially expressed. In fact, correlation analysis between quantitative measurements of each proteotypic peptide couple (i.e. specific to each of the 4 proteins) showed a high measurement concordance and reproducibility between CGN and PDCD4 peptides, while G3BP2 and OCIAD1 peptides displayed low or negative correlation. Higher variation in the quantification of G3BP2 and OCIAD1 peptides may be explained by the influence of the complex serum matrix resulting in decreased peptide ionization efficiency, poor peptide recovery by capturing antibody or early elution during the chromatographic gradient <sup>28</sup>. This possibly is reflected by the lack of differential expression found for G3BP2 and OCIAD1 peptides in our breast cancer tissue cohort measured by iMRM MS. Furthermore, the low precision in measuring G3BP2 and OCIAD1 peptides seemingly reflects the need to develop better antibodies for immunoaffinity captures as well as better tools for selection of proteotypic peptides used to develop targeted MS assays. Yet, as on one hand the development of high sensitivity and specificity antibodies is restricted by the standard methods used (e.g. conjugation to immuno-stimulatory proteins), which only allow a posteriori assessment of antibody quality, on the other hand peptide predictive tools would unlikely be able to address the interference derived from biological samples matrix complexity.

In order to assess whether iMRM and LFQ, high resolution MS predictions would perform similarly in breast cancer tissues, and to assess whether analysis of WTL would impair protein quantitation, Orbitrap MS measurements of the 4 proteins in WTL and LCM material with iMRM-derived data were compared. In a cohort of 38 matching tumor tissues, prediction based on LCM LFQ and WTL iMRM measurements had comparable sensitivity and specificity. On the other hand, high resolution MS of the 4-protein signature in WTL was hindered by a higher amount of missing data, which could be ascribed to ion suppression or matrix effects. This likely contributed to the fact that no difference in the 4-protein expression was found between patient groups, which resulted in an indiscriminative ROC curve. Matrix effects due to specimen heterogeneity have been previously reported in proteomic studies,

where differences in cell population ratios, and possibly the interference from co-eluting species and the background proteome, hinder measurement reproducibility and quantitation<sup>29</sup>. In these cases highly abundant proteins tend to suppress the signal from less abundant ones, also causing chromatographic peak re-sampling, impairing not only quantitation but actual total protein identification<sup>30–32</sup>. Though such issue can be minimized through sample fractionation, this would result in a conspicuous increase of measurement time. In this perspective, iMRM analysis of WTLs is able to bypass sample matrix derived interference and enable precise quantitation, provided that target analytes are predominantly expressed in cell populations of interest and that high recovery purification techniques are used.

Upon building predictive models out of our MS data, removing OCIAD1 and G3BP2 from the predictive model did not impair the overall AUC of the ROC curve, though notable differences were observed in sensitivity when compared to the 4 protein model. However, because a notably lower HR was observed when differences in survival times between patient groups were assessed, it suggests a discriminatory role of the entire protein signature compared to a 2-protein based model.

Upon measuring proteotypic peptides in sera collected before start of tamoxifen therapy (PRE), CGN and PDCD4 peptides showed a trend in expression between good and poor outcome patients, while no differential expression was observed for G3BP2 and OCIAD1 peptides, with the sole exception of G3BP2 LPNFGFVVFDDSEPVQR. This G3BP2 peptide not only showed a highly significant differential expression, but also enrichment in good outcome patients. Though G3BP2 has been reported to be expressed in almost all of the human normal tissues (e.g. endometrium, colon, liver, brain, etc.), protein level alteration derived from other organs would likely impact the other G3BP2 proteotypic peptide. Previous reports had already shown that different peptides derived from the same protein (i.e. serum albumin) might indeed present anti-correlating abundance levels due to differential degradation from proteases, which appears to be sequence-dependent<sup>33</sup>. Furthermore, these findings may also be related to differential protease activity between good and poor outcome patients, which is a well-established hallmark of cancer invasion and progression, and has been shown to modify serum protein abundances both in vivo and ex vivo<sup>34–37</sup>. In this perspective, an increased protease activity in poor outcome patients, which generally display faster tumor progression, may possibly alter quantitation of G3BP2 proteotypic peptides. Despite these seemingly contradictory results, upon evaluating the effectiveness of our classifier, a significant difference between predicted groups was observed. As therapy

effectiveness is generally evaluated in the clinic every 6 months, a significant survival difference measured in the first year after start of tamoxifen treatment would likely be useful for clinicians in assessing which patients should receive an alternative systemic treatment. Upon assessing peptide levels in matched sera collected upon end of therapy (POST), no significant difference was detected between patient groups, nor could any significant difference be observed between predicted groups' survival curves. In this perspective, our 4 protein classifier would be better suited to predict patient outcome at primary tumor resection, or before start of first line tamoxifen therapy.

## **Conclusions**

We have here shown that iMRM MS is a valid assay for our 4-protein signature in clinical specimens due to immunoaffinity enrichment, which reduces sample matrix-derived interference. Furthermore, we have confirmed that an iMRM-based classifier can be used in both primary tumor tissues and serum specimens to predict tamoxifen therapy outcome for recurrent ER positive breast cancer.

## References:

- (1) Siegel, R.; Ma, J.; Zou, Z.; Jemal, A. Cancer statistics, 2014. *CA. Cancer J. Clin.* 2014, 64 (1), 9–29.
- (2) (EBCTCG), E. B. C. T. C. G. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005, 365 (9472), 1687–1717.
- (3) Schiavon, G.; Smith, I. E. Endocrine therapy for advanced/metastatic breast cancer. *Hematol. Oncol. Clin. North Am.* 2013, 27 (4), 715–736, viii.
- (4) Droog, M.; Beelen, K.; Linn, S.; Zwart, W. Tamoxifen resistance: From bench to bedside. *Eur. J. Pharmacol.* 2013, 717 (1-3), 47–57.
- (5) De Marchi, T.; Liu, N. Q.; Stingl, C.; Timmermans, M. A.; Smid, M.; Look, M. P.; Tjoa, M.; Braakman, R. B. H.; Opdam, M.; Linn, S. C.; et al. 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Mol. Oncol.* 2016, 10 (1), 24–39.
- (6) Braakman, R. B. H.; Tilanus-Linthorst, M. M. A.; Liu, N. Q.; Stingl, C.; Dekker, L. J. M.; Luider, T. M.; Martens, J. W. M.; Foekens, J. A.; Umar, A. Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. *J. Proteomics* 2012, 75 (10), 2844–2854.
- (7) Liu, N. Q.; Braakman, R. B. H.; Stingl, C.; Luider, T. M.; Martens, J. W. M.; Foekens, J. A.; Umar, A. Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J. Mammary Gland Biol. Neoplasia* 2012, 17 (2), 155–164.
- (8) Liu, N. Q.; Stingl, C.; Look, M. P.; Smid, M.; Braakman, R. B. H.; De Marchi, T.; Sieuwerts, A. M.; Span, P. N.; Sweep, F. C. G. J.; Linderholm, B. K.; et al. Comparative Proteome Analysis Revealing an 11-Protein Signature for Aggressive Triple-Negative Breast Cancer. *J. Natl. Cancer Inst.* 2014, 106 (2), Epub 2014.
- (9) Harlan, R.; Zhang, H. Targeted proteomics: a bridge between discovery and validation. *Expert Rev. Proteomics* 2014, 11 (6), 657–661.
- (10) Wang, P.; Whiteaker, J.; Paulovich, A. The evolving role of mass spectrometry in cancer biomarker discovery. *Cancer Biol. Ther.* 2009, 8 (12), 1083–1094.
- (11) Kuhn, E.; Addona, T.; Keshishian, H.; Burgess, M.; Mani, D. R.; Lee, R. T.; Sabatine, M. S.; Gerszten, R. E.; Carr, S. a. Developing multiplexed assays for troponin I and interleukin-33 in plasma by peptide immunoaffinity enrichment and targeted mass spectrometry. *Clin. Chem.* 2009, 55 (6), 1108–1117.
- (12) Boja, E. S.; Rodriguez, H. Mass spectrometry-based targeted quantitative proteomics: achieving sensitive and reproducible detection of proteins. *Proteomics* 2012, 12 (8), 1093–1110.
- (13) Whiteaker, J. R.; Zhao, L.; Zhang, H. Y.; Feng, L.-C.; Piening, B. D.; Anderson, L.; Paulovich, A. G. Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal. Biochem.* 2007, 362 (1), 44–54.

- (14) Burgess, M. W.; Keshishian, H.; Mani, D. R.; Gillette, M. a; Carr, S. a. Simplified and Efficient Quantification of Low-abundance Proteins at Very High Multiplex via Targeted Mass Spectrometry. *Mol. Cell. Proteomics* 2014, 13 (4), 1137–1149.
- (15) Jr., D. J. J.; Rodriguez-Canales, J.; Mukherjee, S.; Prieto, D. A.; C.Hanson, J.; Emmert-Buck, M.; Blonder, J. Approaching solid tumor heterogeneity on a cellular basis by tissue proteomics using laser capture microdissection and biological mass spectrometry. *J. Proteome Res.* 2010, 8 (5), 2310–2318.
- (16) Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* 2004, 3 (4), 311–326.
- (17) Anderson, N.; Anderson, N. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* 2004, 3 (2), 235–244.
- (18) Kuhn, E.; Whiteaker, J. R.; Mani, D. R.; Jackson, A. M.; Zhao, L.; Pope, M. E.; Smith, D.; Rivera, K. D.; Anderson, N. L.; Skates, S. J.; et al. Interlaboratory evaluation of automated, multiplexed peptide immunoaffinity enrichment coupled to multiple reaction monitoring mass spectrometry for quantifying proteins in plasma. *Mol. Cell. Proteomics* 2012, 11 (6), M111.013854.
- (19) Vizcaíno, J. A.; Côté, R. G.; Csordas, A.; Dianes, J. a; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013, 41 (Database issue), D1063–D1069.
- (20) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, 26 (7), 966–968.
- (21) De Marchi, T.; Kuhn, E.; Carr, S. A.; Umar, A. Antibody-Based Capture of Target Peptides in Multiple Reaction Monitoring Experiments. In *Methods Mol Biol*; Vivanco, M. del M., Ed.; Springer Science+Business Media: New York, 2015; Vol. 1293, pp 123–135.
- (22) Farrah, T.; Deutsch, E. W.; Kreisberg, R.; Sun, Z.; Campbell, D. S.; Mendoza, L.; Kusebauch, U.; Brusniak, M. Y.; Hüttenhain, R.; Schiess, R.; et al. PASSEL: The PeptideAtlas SRMexperiment library. *Proteomics* 2012, 12 (8), 1170–1175.
- (23) Martinez, R. L. M. C.; Naranjo, J. D. A pretest for choosing between logrank and wilcoxon tests in the two-sample problem. *Metron* 2010, 68 (2), 111–125.
- (24) Shen, W.; Le, C. T. Linear rank tests for censored survival data. *Commun. Stat. - Simul. Comput.* 2000, 29 (1), 21–36.
- (25) Beelen, K.; Zwart, W.; Linn, S. C. Can predictive biomarkers in breast cancer guide adjuvant endocrine therapy? *Nat. Rev. Clin. Oncol.* 2012, 9 (9), 529–541.

- (26) Duggan, J. X.; Vazvaei, F.; Jenkins, R. Bioanalytical method validation considerations for LC–MS/MS assays of therapeutic proteins. *Bioanalysis* 2015, 7 (11), 1389–1395.
- (27) Abbatiello, S. E.; Mani, D. R.; Keshishian, H.; Carr, S. a. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin. Chem.* 2010, 56 (2), 291–305.
- (28) Liebler, D. C.; Zimmerman, L. J. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* 2013, 52 (22), 3797–3806.
- (29) Kondo, T. Inconvenient truth: cancer biomarker development by using proteomics. *Biochim. Biophys. Acta* 2014, 1844 (5), 861–865.
- (30) Taylor, P. J. Matrix effects: The Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. *Clin. Biochem.* 2005, 38 (4), 328–334.
- (31) Annesley, T. M. Ion suppression in mass spectrometry. *Clin. Chem.* 2003, 49 (7), 1041–1044.
- (32) Van Eeckhaut, A.; Lanckmans, K.; Sarre, S.; Smolders, I.; Michotte, Y. Validation of bioanalytical LC-MS/MS assays: Evaluation of matrix effects. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 2009, 877 (23), 2198–2207.
- (33) Dekker, L. J. M.; Burgers, P. C.; Charif, H.; Van Rijswijk, A. L. C. T.; Titulaer, M. K.; Jenster, G.; Bischoff, R.; Bangma, C. H.; Luiders, T. M. Differential expression of protease activity in serum samples of prostate carcinoma patients with metastases. *Proteomics* 2010, 10 (12), 2348–2358.
- (34) Liotta, L. a.; Petricoin, E. F.; Martínez, J. M.; Prieto, I.; Ramírez, M. J. M.; Cueva, C.; Alba, F.; Ramírez, M. J. M.; Rao, J. S.; Zlotta, A. R. Serum peptidome for cancer detection: Spinning biologic trash into diagnostic gold. *Eur. Urol.* 2006, 3 (4), 26–30.
- (35) Villanueva, J.; Schaffer, D. R.; Philip, J.; Chaparro, C. A.; Erjument-Bromage, H.; Olshen, A. B.; Fleisher, M.; Lilja, H.; Brogi, E.; Boyd, J.; et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J. Clin. Invest.* 2006, 116 (1), 756–757.
- (36) Martínez, J. M.; Prieto, I.; Ramírez, M. J.; Cueva, C.; Alba, F.; Ramírez, M. Amino-peptidase activities in breast cancer tissue. *Clin. Chem.* 1999, 45 (10), 1797–1802.
- (37) Rao, J. S. Molecular mechanisms of glioma invasiveness: the role of proteases. *Nat. Rev. Cancer* 2003, 3 (7), 489–501.

## **Chapter 7**

### **Annexin-A1 and Caldesmon are associated with resistance to tamoxifen in estrogen receptor positive recurrent breast cancer**

Tommaso De Marchi, Anne M. Timmermans, Marcel Smid, Maxime P. Look, Christoph Stingl, Mark Opdam, Sabine C. Linn, Fred C.G.J. Sweep, Paul N. Span, Mike Kliffen, Carolien H.M. van Deurzen, Theo M. Luider, John A. Foekens, John W. Martens, Arzu Umar.

*Oncotarget* 2016; 17, 7: 3098-110.

## **Abstract**

Tamoxifen therapy resistance constitutes a major cause of death in patients with recurrent estrogen receptor (ER) positive breast cancer. Through high resolution mass spectrometry (MS), we previously generated a 4-protein predictive signature for tamoxifen therapy outcome in recurrent breast cancer. ANXA1 and CALD1, which were not included in the classifier, were however the most differentially expressed proteins. We first evaluated the clinical relevance of these markers in our MS cohort, followed by immunohistochemical (IHC) staining on an independent set of tumors incorporated in a tissue microarray (TMA) and regression analysis in relation to time to progression (TTP), clinical benefit and objective response. In order to assess which mechanisms ANXA1 and CALD1 might be involved in, we performed Ingenuity pathway analysis (IPA) on ANXA1 and CALD1 correlated proteins in our MS cohort. ANXA1 (Hazard ratio [HR] = 1.83; 95% confidence interval [CI]: 1.22–2.75;  $P = 0.003$ ) and CALD1 (HR = 1.57; 95% CI: 1.04–2.36;  $P = 0.039$ ) based patient stratification showed significant association to TTP, while IHC staining on TMA showed that both ANXA1 (HR = 1.82; 95% CI: 1.12–3.00;  $P = 0.016$ ) and CALD1 (HR = 2.29; 95% CI: 1.40–3.75;  $P = 0.001$ ) expression was associated with shorter TTP independently of traditional predictive factors. Pearson correlation analysis showed that the majority of proteins correlated to ANXA1 also correlated with CALD1. IPA indicated that ANXA1 and CALD1 were associated with ER-downregulation and NF $\kappa$ B signaling. We hereby report that ANXA1 and CALD1 proteins are independent markers for tamoxifen therapy outcome and are associated to fast tumor progression.

## Introduction

ER positive breast cancer constitutes three quarters of all breast malignancies. Treatment options of patients with such tumors include targeted anti-hormonal drugs, of which tamoxifen has been the first choice for decades, both in the adjuvant and in the recurrent setting (1). In the adjuvant setting, tamoxifen significantly increases patient survival and decreases the risk of metastasis occurrence (2). In recurrent ER positive breast cancer, approximately 50% of patients treated with tamoxifen manifests intrinsic drug resistance, while the other half experiences acquired resistance during therapy (3, 4). Many studies have described mechanisms of tamoxifen resistance in breast cancer patients, such as the upregulation of ER transcriptional co-activators (5) or the expression of ER isoforms (6), the activation of several tyrosine kinase pathways such as PI3K/MAPK (7), or the dysregulation of tamoxifen metabolizing enzymes (8). Acquired mutations in the ligand-binding domain of ESR1 protein (e.g. p.Tyr537Ser) during endocrine treatment have also been associated with constitutive agonistic activity of the receptor and the unresponsiveness to anti hormonal therapies (9–11). Furthermore, gene expression analyses have been performed to derive biomarkers predictive of tamoxifen therapy outcome in both the adjuvant and the recurrent settings (12, 13). So far, none of these markers have found clinical application due to non-optimal study design, lack of extensive sample validation, or difficulty in developing assays into an accurate and standardized format (14, 15).

Proteomics-based technologies have shown to enable expansion of the depth of biomarker investigation (16), adding new layers of information to the clinical and biological profiling of diseases (17–19). Advancements in liquid chromatography and mass spectrometry (MS) instruments now enable almost full coverage of the protein-coding genome, and quantitation of even slight changes in protein expression (20, 21). Furthermore, targeted MS techniques provide accurate and absolute quantitation of target analytes, making them suitable for both biomarker verification and clinical diagnostics (22, 23). In our laboratory, we have developed a tissue proteomics biomarker discovery pipeline combining laser capture microdissection with high resolution MS analysis and label free quantitation for the analysis of breast carcinomas (24, 25), allowing us to not only establish a robust platform for biomarker discovery but also dissect the underlying biological mechanisms in epithelial tumors. Through this pipeline, we previously analyzed a cohort of snap frozen ER positive primary tumors and developed a 4-protein signature – comprising PDCD4, CNG, OCIAD1, and G3BP2 – predicting poor outcome to tamoxifen treatment with 86.7% sensitivity and 41.5% specificity

(26). However, the 4 proteins included in the classifier did not display extreme differential levels between tamoxifen outcome groups, suggesting that other factors may be involved in tamoxifen treatment outcome. In order to address this, we combined our initial MS cohorts and assessed which proteins manifested the highest change in expression between patient groups. The top candidates were selected based on significance after statistical evaluation of their abundance levels between patient groups, and their association to TTP was assessed in MS analyzed samples. Furthermore, IHC analysis of an independent cohort of tumors incorporated in a TMA provided further marker verification. In order to assess which pathways the top molecules were involved in, correlation and pathway analyses were performed. A schematic representation of the study's workflow is reported in Figure 7.1.

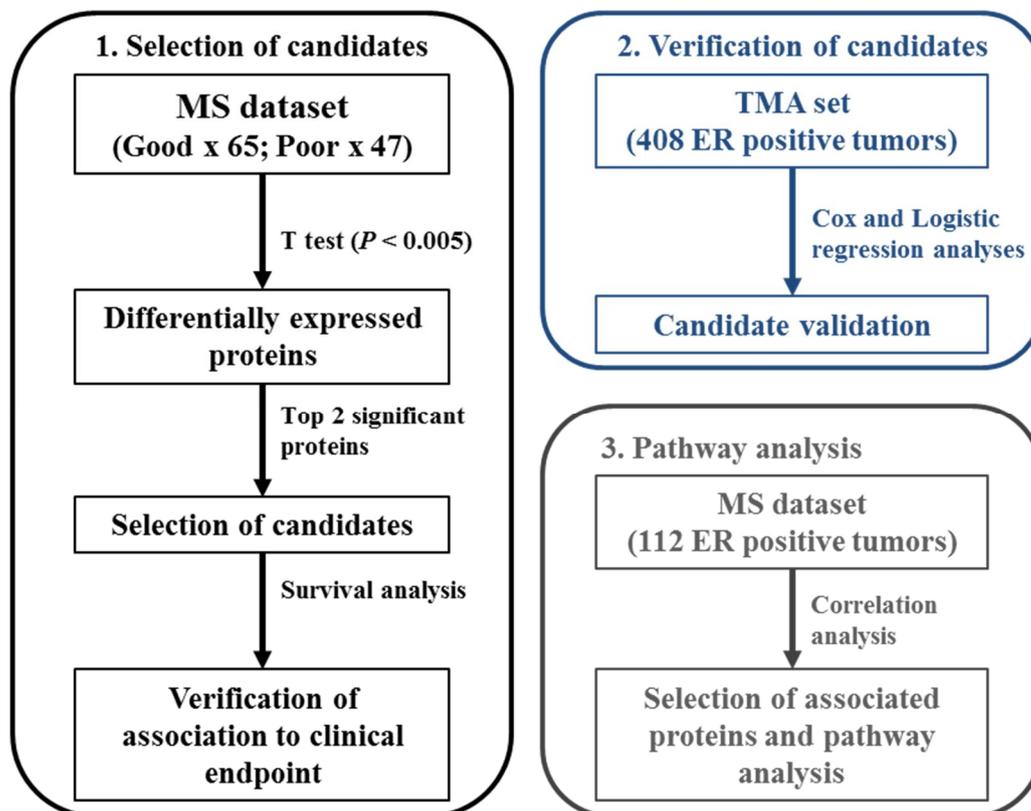


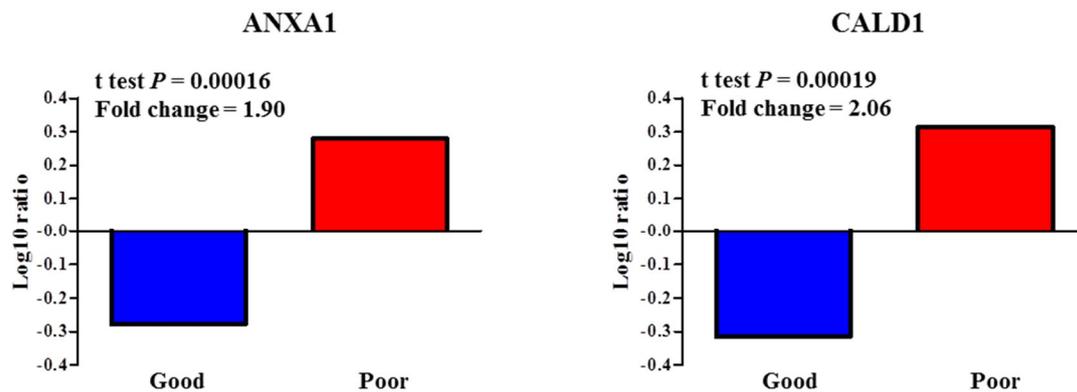
Figure 7.1. Schematic representation of analysis workflow.

## Results

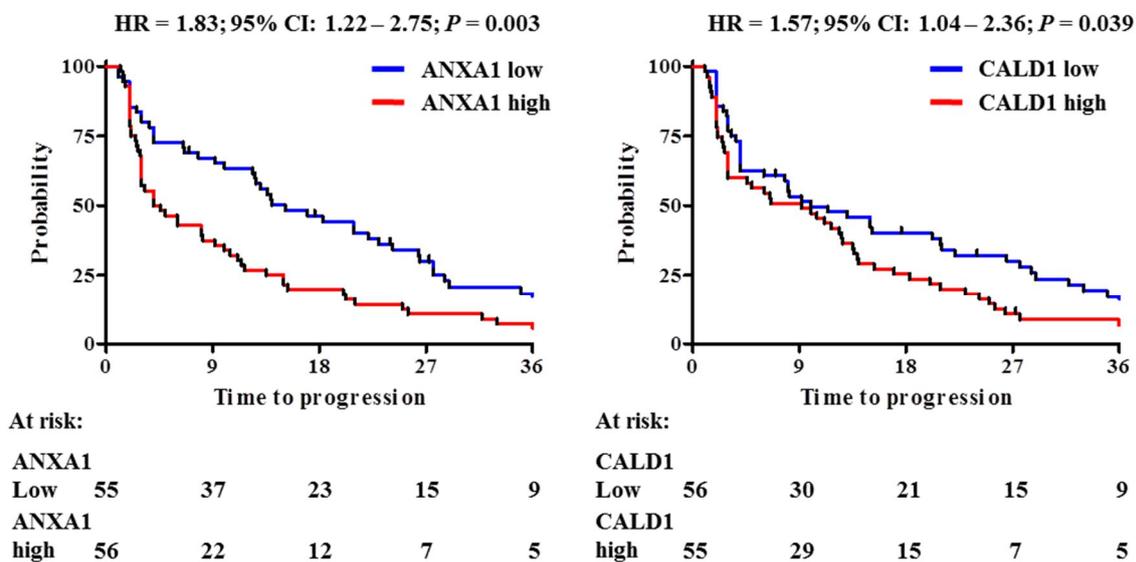
### *Analysis of MS sets*

In order to assess which proteins showed the larger change in expression between good and poor outcome tamoxifen treatment groups, statistical analysis (i.e. t test) was performed on all 1,960 quantified proteins (Table S1) in our 112 patient MS cohort (Clinical and histopathological characteristics are reported in Table S2). Based on significance levels, ANXA1 (t test  $P < 0.001$ ; fold change = 1.90) and CALD1 (t test  $P < 0.001$ ; fold change = 2.06) proteins were selected as top candidates (Figure 7.2A and Figure S1). In addition to this, survival analysis was performed on ANXA1 and CALD1-stratified patients, with TTP as endpoint. A significant difference was observed between patients who displayed high levels (based on median expression) of ANXA1 (HR = 1.83; 95% CI: 1.22 – 2.75;  $P = 0.003$ ) and CALD1 (HR = 1.57; 95% CI: 1.04 – 2.36;  $P = 0.039$ ) proteins (Figure 7.2B). These data show that ANXA1 and CALD1 are positively associated to faster disease progression after tamoxifen treatment in ER positive recurrent breast cancer.

A



B



**Figure 7.2. ANXA1 and CALD1 expression levels and survival analyses in MS cohorts.**

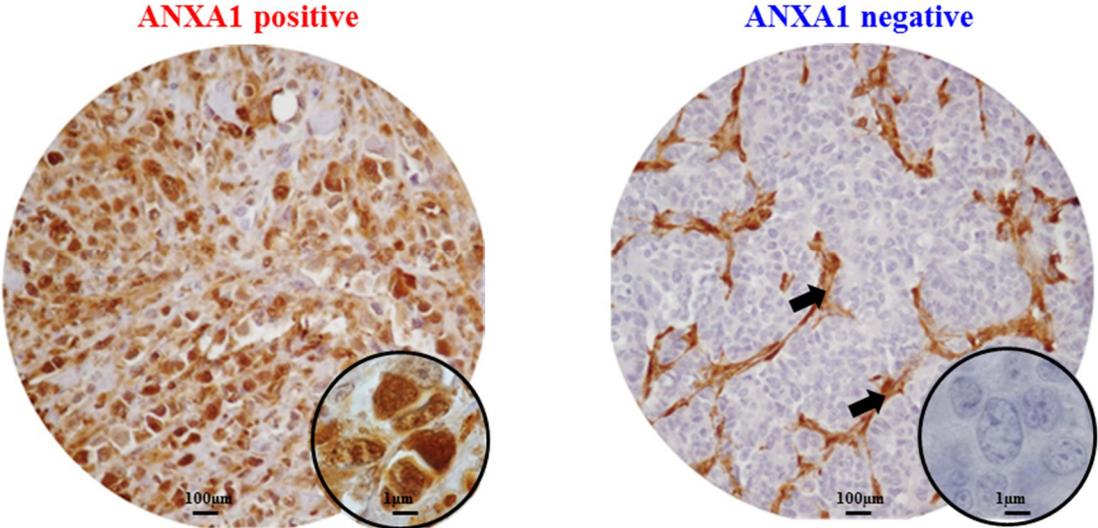
Measurement of ANXA1 and CALD1 protein levels based on previously derived proteomic data. Panel A displays Log ratio bar charts show that both ANXA1 (t test  $P = 0.00016$ ; Fold ratio = 1.90; left) and CALD1 (t test  $P = 0.00019$ ; Fold ratio = 2.06; right) were highly differentially expressed in the poor outcome group. Stratification of patients according to median protein level showed that a significant difference was observed between ANXA1 (left) and CALD1 (right) protein levels (Panel B).

### *Validation of ANXA1 and CALD1 as independent markers using immunohistochemistry*

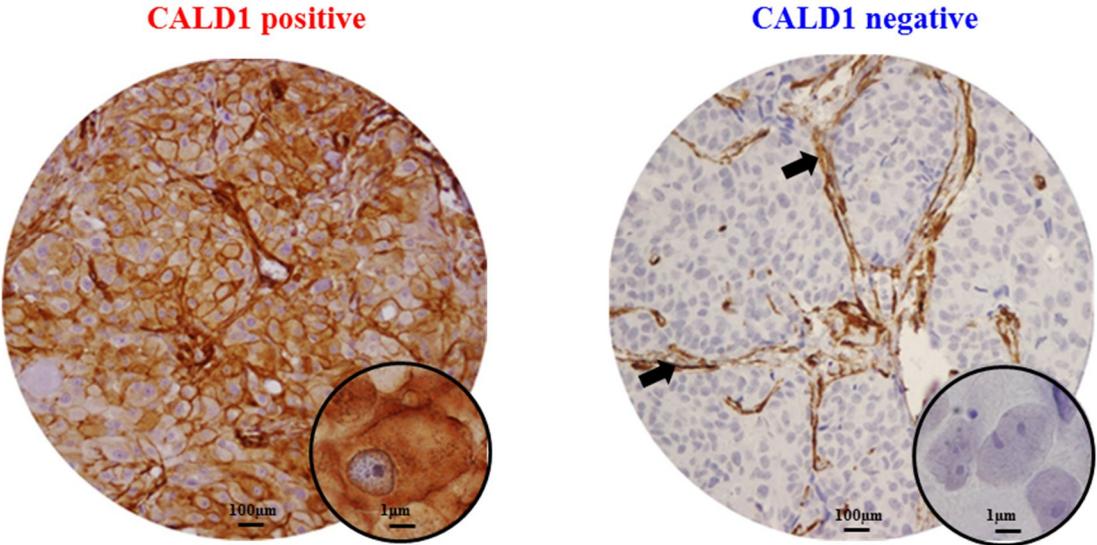
In order to further confirm our findings, ANXA1 and CALD1 protein levels were assessed through IHC staining on an independent cohort of 408 FFPE tumor tissues derived from patients that received tamoxifen as first line therapy for recurrent breast cancer (Table S3). After filtering for missing values (Figure S2), a total of 20 out of 235 tumor tissues displayed

ANXA1 positivity (Figure 7.3A). ANXA1 presence was significantly associated with shorter TTP in both univariate (HR = 2.99; 95% CI: 2.14 - 4.16;  $P < 0.001$ ) and multivariate (HR = 1.82; 95% CI: 1.12 - 3.00;  $P = 0.016$ ) Cox regression analyses (Table 7.1; Figure 7.4A).

A



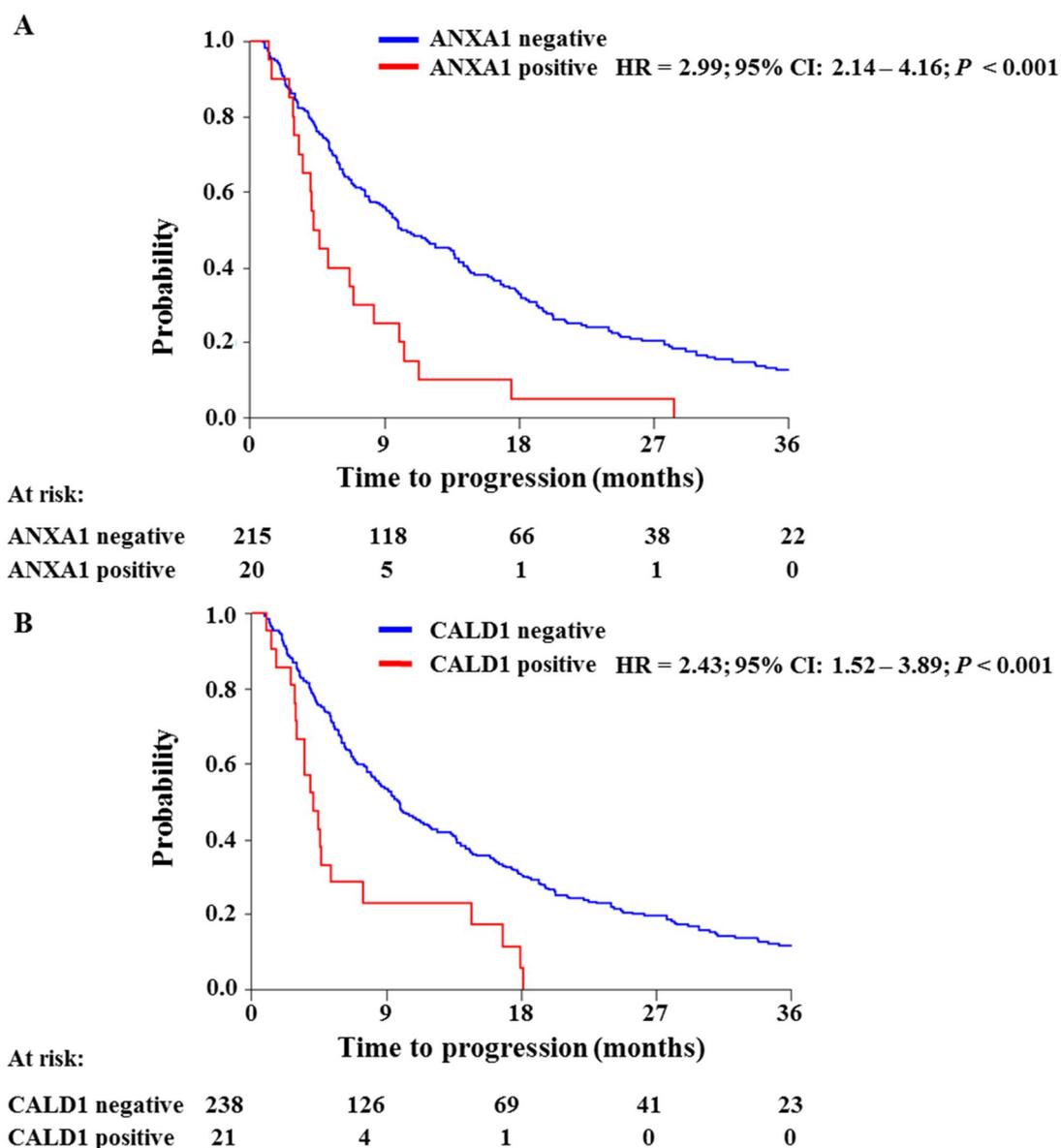
B



**Figure 7.3. Immunohistochemical stainings of ANXA1 and CALD1 proteins.**

Breast carcinomas included in the TMA displayed either ANXA1 positivity or negativity (A). Strong ANXA1 staining was found ubiquitously in stromal cells (black arrows) and was not taken into account in the survival analysis. CALD1 IHC staining was found at the membrane and cytoplasm of both carcinoma and stromal cells, but the latter was not taken into account for survival analyses (B).

CALD1 positivity (Figure 7.3B) was observed in 21 out of 259 patients and Cox regression analysis showed a significant positive correlation with shorter TTP both in univariate (HR = 2.43; 95% CI: 1.52 - 3.89;  $P < 0.001$ ) and multivariate (HR = 2.29; 95% CI: 1.40 - 3.75;  $P = 0.001$ ) analyses (Table 7.2; Figure 7.4B).



**Figure 7.4. Survival analyses of ANXA1 and CALD1 association to TTP.**

ANXA1 and CALD1 levels (i.e. negative/positive) were assessed by IHC and analyzed by Cox regression analysis and Log-rank test. Both ANXA1 (A) and CALD1 (B) levels showed significant association with short TTP in ER positive breast tumors.

In addition to this, we analyzed whether the associations of ANXA1 and CALD1 with TTP were independent of each other in the subset of 235 tumor tissues for which both measurements were available. Multivariate Cox regression analysis showed that both ANXA1 (HR = 1.90; 95% CI: 1.16 - 3.10;  $P = 0.010$ ) and CALD1 (HR = 2.29; 95% CI: 1.40 - 3.74;  $P = 0.001$ ) were independently correlated to TTP (Table S4).

**Table 7.1. Univariate and multivariate Cox regression analysis for the association of ANXA1 staining with TTP.**

	n of patients	Univariate			Multivariate		
		HR	95% CI	P	HR	95% CI	P
<b>ANXA1</b>							
Negative	272	1.00			1.00		
Positive	45	2.99	2.14 - 4.16	< 0.001	1.82	1.12 - 3.00	0.016
<b>Age</b>							
≤55 years	125	1.00			1.00		
>55 years	192	0.59	0.47 - 0.75	< 0.001	0.56	0.42 - 0.75	< 0.001
<b>Disease-free survival</b>							
≤12 months	67	1.00			1.00		
>12 months	250	0.69	0.52 - 0.91	0.008	0.72	0.51 - 1.01	0.057
<b>Dominant site of relapse</b>							
Loco-regional	43	1.00					
Bone	113	1.20	0.83 - 1.74	0.235			
Visceral	74	1.27	0.85 - 1.89	0.238			
Bone and other	87	1.25	0.85 - 1.84	0.258			
<b>PgR*</b>							
Negative	111	1.00			1.00		
Positive	204	0.51	0.40 - 0.64	< 0.001	0.71	0.52 - 0.97	0.034
<b>Her2 status*</b>							
Negative	201	1.00					
Positive	114	1.18	0.93 - 1.50	0.170			
<b>Tumor differentiation*</b>							
Good	46	1.00			1.00		
Moderate	150	1.45	1.01 - 2.06	0.042	1.15	0.79 - 1.67	0.456
Poor	118	1.95	1.35 - 2.82	< 0.001	1.24	0.82 - 1.89	0.311

\*Missing data not reported.

Tumor differentiation was assessed through Bloom-Richardson scoring system.

Acronym: PgR: progesterone receptor.

**Table 7.2. Univariate and multivariate Cox regression analysis for the association of CALD1 staining with TTP.**

	n of patients	Univariate			Multivariate		
		HR	95% CI	P	HR	95% CI	P
<b>CALD1</b>							
Negative	238	1.00			1.00		
Positive	21	2.43	1.52 - 3.89	< 0.001	2.29	1.40 - 3.75	0.001
<b>Age*</b>							
≤55 years	109	1.00			1.00		
>55 years	150	0.64	0.50 - 0.83	0.001	0.55	0.41 - 0.73	< 0.001
<b>Disease-free survival</b>							
≤12 months	45	1.00			1.00		
>12 months	214	0.72	0.52 - 1.00	0.052	0.76	0.52 - 1.11	0.158
<b>Dominant site of relapse</b>							
Loco-regional	29	1.00					
Bone	103	1.36	0.88 - 2.12	0.166			
Visceral	58	1.34	0.83 - 2.16	0.229			
Bone and other	69	1.47	0.92 - 2.34	0.105			
<b>PgR</b>							
Negative	65	1.00			1.00		
Positive	194	0.73	0.54 - 0.97	0.031	0.70	0.50 - 0.96	0.029
<b>Her2 status**</b>							
Negative	158	1.00					
Positive	79	1.18	0.91 - 1.54	0.221			
<b>Tumor differentiation**</b>							
Good	37	1.00					
Moderate	186	1.21	0.82 - 1.77	0.333			
Poor	35	1.5	1.00 - 2.26	0.051			

\*Age was assessed at start of tamoxifen therapy.

\*\*Missing data not reported.

Tumor differentiation was assessed through Bloom-Richardson scoring system.

Acronym: PgR: progesterone receptor

Furthermore, in order to evaluate whether ANXA1 and CALD1 could be used in combination we merged IHC stainings into four categories: positive/positive, positive/negative, negative/positive, and negative/negative. ANXA1 and CALD1 staining only were observed in 17 and 16 tumor tissues, respectively, while only 4 tumors showed co-expression of the two markers. For all three positive categories (i.e. ANXA1 positive/CALD1 negative; ANXA1

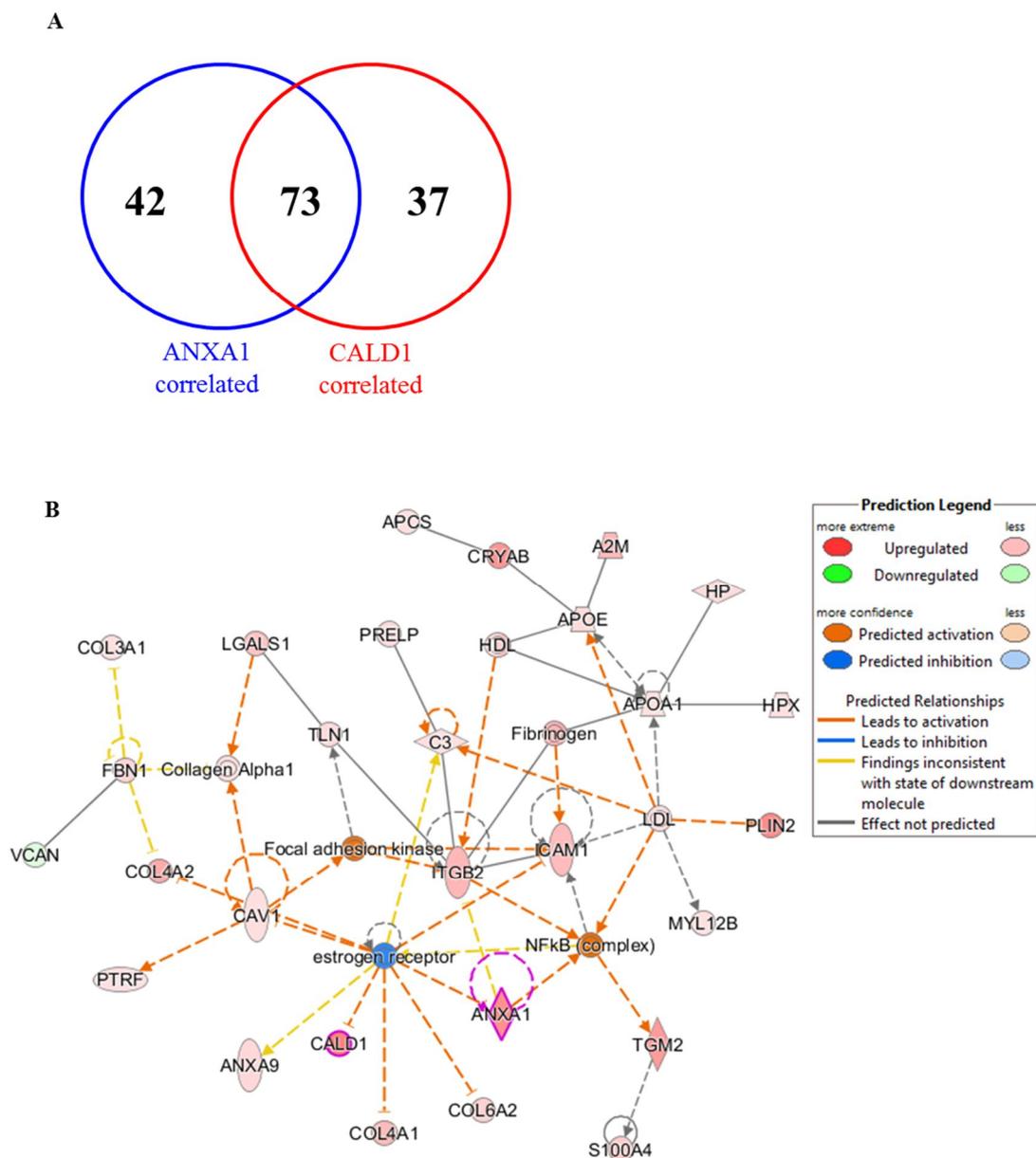
negative/CALD1 positive; ANXA1 positive/CALD1 positive) an association with shorter TTP was found (Table S5, Figure S3). However, due to the low amount of tumors comprised in each category, further verification is needed to assess whether ANXA1 and CALD1 can be used in combination to effectively identify fast progressing breast carcinomas.

Logistic regression analysis showed a significant association of ANXA1 staining to clinical benefit to tamoxifen therapy in univariate analysis (Odds ratio [OR] = 0.22; 95% CI: 0.11 to 0.45;  $P < 0.001$ ) and a borderline association after correction for traditional predictive factors (OR = 0.38; 95% CI: 0.15 to 1.01;  $P = 0.052$ ; Table S6). The association of ANXA1 with objective response was found significant only in the univariate analysis (OR = 0.20; 95% CI: 0.46 to 0.84;  $P = 0.028$ ; Table S7). A significant association was found between CALD1 staining and no clinical benefit both in univariate (OR = 0.21; 95% CI: 0.08 to 0.57;  $P = 0.002$ ) and multivariate (OR = 0.21; 95% CI: 0.08 to 0.56;  $P = 0.002$ ) logistic regression analysis (Table S8). No association was found between CALD1 staining and objective response (Table S9). Due to the fact that only CALD1 showed significant association with the type of response, while ANXA1 displayed only borderline significance, combination of ANXA1 and CALD1 stainings was not performed using tumor response as the endpoint of tamoxifen therapy. Overall, these data suggest a significant relationship between ANXA1 and CALD1 positivity and early tumor progression after tamoxifen treatment for recurrent ER positive breast cancer.

#### *Pathway analysis of ANXA1 and CALD1*

As ANXA1 and CALD1 showed to be highly expressed in the poor outcome patient group, Pearson correlation analysis of abundance levels of all proteins in the MS dataset to ANXA1 and CALD1 expression was performed. After selection of only highly correlated proteins, correlation analysis yielded a total of 115 (e.g. SERPINH1) and 110 (e.g. Vinculin) proteins correlated to ANXA1 and CALD1, respectively. Of these, 73 (i.e. more than 60% in both lists) correlated with both ANXA1 and CALD1 levels (Table S10 and Table S11 Figure 7.5A), suggesting that these proteins presented a shared biology. In the light of this we merged the two correlated protein lists and performed pathway analysis of ANXA1 and CALD1 and their associated proteins through IPA software. Canonical pathway analysis showed activation of acute phase response signaling (Z-score = 3.00; Fisher  $P = 9.59E-10$ ; Table S12), pointing out that molecules involved in inflammation might be correlated to

tamoxifen poor outcome. Upstream analysis showed that ER signaling was downregulated (Z-score = -2.111; Fisher's  $P = 1.39E-07$ ; Table S13), suggesting a link between ANXA1 and CALD1 related proteins in the disruption of ER signaling. Molecular interaction analysis indicated that both CALD1 and ANXA1 were comprised in a network along with proteins involved in cellular movement and immune response (Focus molecules: 28; P-score = 48; Figure 7.5B).



**Figure 7.5. Interaction pathways derived from proteins correlated to ANXA1 and CALD1.**

Proteins associated with both CALD1 and ANXA1 were combined into one list (A) and submitted to IPA. Molecular network analysis showed that both ANXA1 and CALD1 were involved in downregulation of ER and

activation of the NFκB pathway (B). Expressed molecules in the pathway were: A2M (alpha-2-macroglobulin), ANXA1 (Annexin-A1), ANXA9 (Annexin-A9), APCS (Amyloid P component), APOA1 (apolipoprotein-A1), C3 (Complement C3), CALD1 (Caldesmon), CAV1 (Caveolin-1), COL3A1 (collagen type III alpha 1), COL4A1 (collagen type IV alpha 1), COL4A2 (collagen type IV alpha 2), COL6A2 (collagen type VI alpha 2), CRYAB (Crystallin alpha B), FBN1 (Fibrillin-1), HP (Haptoglobin), HPX (Hemopexin), ICAM1 (Intercellular adhesion molecule 1), ITGB2 (Integrin beta 2), LGALS1 (Lectin galactoside-binding soluble 1), MYL12B (Myosin light chain 12B), PLIN2 (Perilipin-2), PRELP (Proline/arginine-rich end Leucine-rich repeat protein), PTFR (Polymerase I and transcript release factor), S100A4 (Calcium binding protein S100A4), TGM2 (Transglutaminase 2), TLN1 (Talin-1), and VCAN (Versican).

This network comprised proteins belonging to the extracellular matrix (e.g. APOE; COL4A2) and proteins involved in cell movement (e.g. MYL12B) and cell adhesion (e.g. ICAM1). Molecule activity prediction based on proteins Log ratios pointed out that molecules in the network were not only involved in ER down-regulation, but were also associated with the activation of the focal adhesion kinase and NFκB pathways. Overall, these data suggest that ANXA1 and CALD1, along with their correlated proteins, are associated with down-regulation of ER signaling and the activation of inflammation response mechanisms. However, further verification in breast cancer model systems are required to confirm these hypotheses.

## **Discussion**

Resistance to tamoxifen is still a major cause of death in patients with ER positive recurrent breast cancer. The advanced state of current proteomic technologies allows profiling of biological samples for discovery of disease biomarkers (27). Indeed also in recurrent ER positive breast cancer, using our dedicated tissue proteomics workflow, we developed and validated a 4-protein signature predictive of tamoxifen treatment outcome (26). Despite the fact that our predictor is capable of discriminating patient groups displaying good and poor outcome to tamoxifen therapy, the 4 signature proteins alone are unlikely capable of addressing the full extent of resistance mechanisms. Out of a panel of 1,960 proteins, ANXA1 and CALD1 constituted the top 2 significant proteins and were shown to be overexpressed in the poor outcome group. Furthermore ANXA and CALD1 have already been described as cell migration and markers of an epithelial-to-mesenchymal (EMT) –like phenotype in ER

negative breast cancer cell line models (28-30), though their role in ER positive tumors still needs to be functionally elucidated.

In order to confirm our MS findings, we performed IHC stainings of both ANXA1 and CALD1 in an independent cohort of FFPE ER positive tumor tissues captured in a TMA to assess the clinical relevance of these markers. Our regression analyses showed that not only ANXA1 (HR = 1.82) and CALD1 (HR = 2.40) stainings were significantly associated with shorter TTP independently of traditional predictive factors, but also contributed in stratifying patient groups independently of each other. Combination of IHC stainings suggested that ANXA1 and CALD1 could be used in concert to discriminate patients who would suffer from fast disease progression after first line tamoxifen, however the relatively small size of these groups implies that further verification in a larger patient cohort is necessary. Though association of CALD1 to breast cancer therapy outcome has been poorly assessed so far, ANXA1 expression was previously associated to BRCA1/2 mutation carriers and prediction of high mortality risk in Her2+ patients (29). Our data might suggest an additional role of ANXA1 and CALD1 in disease progression after tamoxifen therapy of ER positive recurrent breast cancers.

In order to better investigate which proteins and molecular pathways were connected to ANXA1 and CALD1, we performed pathway analysis on a cohort of ER positive tumors which were previously analyzed by high resolution MS. Proteins that were found to be correlated to ANXA1 and CALD1 levels were associated not only with inflammatory response (canonical pathway analysis), but also with activation of the NF $\kappa$ B pathway, downregulation of ER signaling and focal adhesion (molecule activity prediction). Similar findings were reported before as ANXA1 has been shown to be associated with increased metastatic potential in MCF-7 breast cancer cell lines following downregulation of ER and the expression of basal markers (e.g. vinculin) (30). Moreover, studies on breast cancer cell line models have shown that constitutive activation of the NF $\kappa$ B pathway leads to downregulation of ER signaling (31), which can have a prominent role in tamoxifen resistance, if not in a generalized anti-hormonal therapy unresponsiveness. In another study elucidating the role of ANXA1 in relation to NF $\kappa$ B activation, it was shown that its interaction with the IKK complex led to the constitutive activation of this pathway, promoting metastasis and decreasing survival in an intracardiac metastatic model (32). The constitutive activation of NF $\kappa$ B signaling in cancer cells favors not only cell survival, but also the acquisition of a more malignant phenotype (33). ANXA1 was also associated with acquisition of an EMT-like

phenotype in ER negative breast cancer cell line models through activation of TGF $\beta$  signaling (34). Although pathway analysis did not show a direct activation of the TGF $\beta$  pathway, cross-talk between TGF $\beta$  and NF $\kappa$ B has already been described in various forms of cancer, in which one pathway modulates the other via activation of binding proteins (35) or of micro-RNAs (36). In contrast, it has been shown that expression of ANXA1 was related to inhibition of NF $\kappa$ B in pancreatic and colon cancer cell lines (37), while in Adriamycin-resistant bladder cancer, ANXA1 has been reported to be downregulated (38). In the light of this, ANXA1 may be associated with the constitutive activation of NF $\kappa$ B signaling in a tissue-specific manner, which would on one side lead tumor cells to become estrogen-independent and on the other side promote a more aggressive and mesenchymal-like phenotype, though further studies need to be performed in order to confirm this function. Nonetheless, a possible novel treatment option for ANXA1 positive, tamoxifen resistant tumors would consist of blocking antibodies, which have already shown efficacy in inhibiting migration and invasion rates in pancreatic carcinoma cells (39).

CALD1 is an actin-, calmodulin- and myosin-binding protein that has been described as a cell motility suppressor by stabilizing stress fibers through F-actin binding (40, 41). Despite its role in downregulating cell motility by cytoskeletal rearrangements, CALD1 has also been described as a key component in TGF $\beta$ -driven EMT via its overexpression (42). In addition to this, downregulation of ER in MCF7 cells has been linked to the upregulation of CALD1, concomitantly with the acquisition of a new phenotype that encompasses increased growth rates, loss of cell-to-cell adhesion and a redistribution of the cytoskeletal components, resulting in increased motility (30). In addition to this, CALD1 interaction with cGMP-dependent protein kinase I $\beta$  has been shown to regulate cell invasion and migration in breast cancer cell lines (28). Similar to ANXA1, also CALD1 has been shown to display opposite roles in cancer and invasion, since its expression has been associated to reduced cell invasion in colon cancer cell lines (43). In this perspective, CALD1 may be another key effector in cytoskeletal rearrangements and the acquisition of a rapid spreading tumor phenotype in breast tissue, while countering those effects in other tissue types.

As both ANXA1 and CALD1 have been reported to be expressed in basal-type breast cancers, which are characterized by downregulation of ER and its responsive genes (44), and the fact that only a minority of the ER positive tumors that were captured in the TMA displayed expression of these markers, it is possible that these proteins may enable further stratification of ER positive breast malignancies. In the perspective of tamoxifen resistance, ANXA1 and

CALD1 may be involved in the activation of the NF $\kappa$ B pathway, which would promote cell survival by blocking intrinsic (mitochondrial-mediated) and extrinsic (death receptors-mediated) apoptotic signals and render cancer cells independent of estrogens. In addition to this, the acquisition of a rapidly spreading and fast growing tumor cell phenotype would result in a faster tumor progression, and probably an estrogen-independent phenotype.

We have here shown that ANXA1 and CALD1 are associated with tamoxifen therapy clinical outcome in recurrent ER positive breast cancer. Expression of these proteins not only correlates with shorter TTP independently of traditional predictive factors, but these markers also contributed independently of one another. In other words, ANXA1 and CALD1, alone or in combination, are able to identify groups of patients that would less likely benefit from tamoxifen therapy. In addition to this, pathway analysis suggested that ANXA1 and CALD1 are likely linked to the downregulation of ER signaling and acquisition of a more malignant phenotype with EMT-like features. Blocking such pathways would probably constitute an effective additional or substitutive therapy in patients expressing such markers, however further functional studies should be performed in order to determine causal effects of these proteins in these signaling cascades.

## **Materials and Methods**

### *MS dataset*

A total of 112 ER positive tissues were analyzed by high resolution MS after LCM-based breast carcinoma cell enrichment, as previously described (26). MaxQuant (version 1.2.2.5; search engine: Andromeda) (45, 46) results of Orbitrap .RAW data deposited in ProteomeXchange via PRIDE repository (dataset identifiers: PXD000484 and PXD000485) (47) from previously described sets (26) were imported in Microsoft Excel and normalized for inter-batch effects using ComBat algorithm in R environment (48) allowing 10 minimum observations across all samples. Outcome to tamoxifen treatment was defined based on TTP, with a 6 months cutoff: patient whose tumors progressed within ( $\leq$ ) 6 months after start of therapy were defined as poor outcome, while progression after ( $>$ ) 6 months was defined as good outcome. This set comprised 67 good and 47 poor outcome patients, respectively. Protein differential expression between patient groups was tested by t test (two sided test, unequal variances assumed) in Microsoft Excel. Top candidates were selected based on t test

P and their relation to TTP was confirmed by survival analysis through survival analysis of patients stratified according to median level of protein expression.

#### *TMA dataset*

A total of 447 FFPE tissues collected from Erasmus MC and regional hospitals were incorporated in a TMA. For statistical analysis only ER positive tumors from patients who did not receive any adjuvant hormonal therapy were included. Furthermore, patients who showed no tumors after histological revision or manifested disease progression before ( $\leq$ ) 3 weeks after start of therapy were excluded. In addition, tumors that were comprised in the MS sets were also excluded. This led to the inclusion of a total of 408 ER positive tumors, of which response data were collected according to the standard International Union Against Cancer criteria (49). Eleven (2.70%) and 51 (12.50%) patients showed complete remission (CR) and partial remission (PR), respectively. Two hundred and five patients showed no change (NC) of disease, of whom 35 (8.58%) displayed NC for less ( $\leq$ ) than 6 months, while 170 (41.66%) showed NC for longer ( $>$ ) than 6 months (defined as stable disease [SD]). A total of 141 (34.56%) patients displayed progressive disease (PD). Clinical benefit was defined as CR + PR + SD patients ( $n = 232$ ; 57%), while objective response was defined as CR + PR only ( $n = 62$ ; 15%). This retrospective study used coded primary tumor tissues, in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (<http://www.federa.org/codes-conduct>). Reporting Recommendations for Tumor Marker Prognostic Studies were followed where possible (50).

#### *Tissue micro-array*

TMA was prepared using an ATA 27 (Beecher Instruments, Sun Prairie, WI, USA). 447 paraffin-embedded primary, ER positive breast cancer tissues derived from patients who received tamoxifen as first-line therapy for recurrent disease were used to prepare the TMA. Tissue cores of 0.6 mm were taken from each tissue paraffin block and transferred in triplicate into a TMA recipient block. For each tumor tissue sample, three different areas of the tumor were taken as biological replicates. TMA slides were digitalized and analyzed using Slidepath software (Leica Microsystems, Solms, Germany).

### *Immunohistochemistry*

Five  $\mu\text{m}$  sections of FFPE tissues captured on the TMA were incubated at  $60^{\circ}\text{C}$  and washed in xylene (3 x 5 min) for de-paraffination. Washings with decreasing concentrations of ethanol were used to re-hydrate tissues as follows: 100% ethanol (1 x 5 min, 2 x 2 min), 70% ethanol (1 x 2 min), 50% ethanol (1 x 2 min), distilled water (1 x 2 min). Incubation with DAKO (Agilent Technologies Inc, Glostrup, Denmark) antigen retrieval solution diluted 1:10 in MilliQ water was performed at  $95^{\circ}\text{C}$  for 40 min. Slides were then cooled down to room temperature and washed with PBS (3 x 5 min). Slides were first incubated with 0.003%  $\text{H}_2\text{O}_2$  in PBS to block endogenous peroxidase (10 min) and subsequently with blocking solution consisting of 5% BSA in PBS for 30 min. Anti-ANXA1 (Clone ID: 29/Annexin I; Transduction Laboratories, Lexington, KY, USA) and anti-CALD1 (Clone ID: TD107; Enzo Life Sciences Inc., Farmingdale, NY, USA) mouse monoclonal primary antibodies were diluted 1:2000 (ANXA1) and 1:400 (CALD1), respectively, in DAKO Antibody Diluent, and slides were incubated for 1 h at room temperature. Slides were then washed with PBS, and DAKO Envision<sup>®</sup> secondary antibody (labeled polymer HRP -Mouse, 200  $\mu\text{l}$  per slide) solution was added to each slide and incubated for 45 min at room temperature. A washing cycle with PBS was performed for 5 min and a 1:15 solution of DAB+ chromogen in DAB+ substrate buffer was added, following incubation in the dark for 10 min. Slides were then washed in tap water for 5 min, stained with hematoxylin/eosin for 1 min each and dehydrated again through sequential washings in 50%-70%-100% ethanol and xylene of 5 min each. Cover glasses were mounted with Pertex and slides were left to dry. ANXA1 and CALD1 stained tissue sections were scored only for quantity of stained carcinoma cells due to the fact that all stained tumors displayed strong staining intensity. Scoring was performed by an experienced researcher in a blind manner, and triplicate scores were assessed and validated by a second experienced researcher, who was extensively trained by a specialized breast pathologist. Fibroblasts, endothelial cells and leukocytes displayed strong ubiquitous ANXA1 and CALD1 stainings as well, while adipocytes only stained for ANXA1 (Figure S4A-B). Only staining in breast carcinoma cells were taken into account for further analyses. Slides were digitalized and analyzed using Slidepath software (Leica Microsystems, Solms, Germany).

### *Statistical analysis of IHC staining results*

Staining scores of TMA incorporated tissues were filtered for missing data, stringently excluding cores for which triplicate measurement was not available (e.g. due to loss of core during staining procedure or not enough [ $< 35$ ] carcinoma cells observed in at least one core), leading to a set of 317 tissue samples for ANXA1, a list of 259 tissues for CALD1, and a list of 235 tissues for both ANXA1 and CALD1 (Figure S2). Due to the fact that tumor tissues displayed high quantities of stained breast carcinoma cells, both ANXA1 and CALD1 were scored as either absent or present. Association of ANXA1 and CALD1 stainings with TTP or response data were tested by Cox and logistic regression, respectively. Patient age, disease free interval, dominant site of relapse, progesterone receptor (PgR) positivity, HER2 overexpression, and degree of tumor differentiation (Bloom-Richardson) were included in all regression analyses. With the exception of disease-free survival (to correct for prognosis), variables that did not display any significant association with TTP or any response criteria were excluded from the multivariate regression models. Cox regression, logistic regression, HRs, OR, and 95% CIs were calculated in Stata (v 13.1; Stata Corp, College Station, TX, USA).

### *Extraction of proteomics data and pathway analysis*

ANXA1 and CALD1 levels were correlated with the rest of the quantified dataset through Pearson correlation. The distribution of correlation coefficients was analyzed by Z-score; cutoffs were selected according to the following formula:

$$\text{Z-score cutoffs} = \text{mean} \pm 1.96 * (\text{standard deviation})$$

Z-score transformed correlation coefficients that fell below ( $< \text{mean} - 1.96 * [\text{standard deviation}]$ ) or above ( $> \text{mean} + 1.96 * [\text{standard deviation}]$ ) the cutoffs were selected as highly correlated proteins. IPA® (IPA, Qiagen, Redwood City, [www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity)) was performed selecting Uniprot identifiers and Log10 ratios as input data (no ratio cutoff was selected). Pathway analysis was run using Ingenuity Knowledge Base as reference database. Analysis was run with the following options enabled: homo sapiens was selected as species; human tissues and breast cancer cell lines were selected as protein expression sites; Mutation and Data Sources options were kept as default (i.e. All). Interaction networks were generated including endogenous chemicals and maintaining other options as default (i.e.

number of Molecules per network: 35; number of Networks per analysis: 25). Networks that showed the highest P-score (i.e.  $P\text{-score} = -\text{Log}_{10} [\text{Fisher's } P]$ ) were subjected to Molecule Activity Predictor in IPA.

## References:

1. (EBCTCG), E. B. C. T. C. G. (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 365, 1687–717
2. Schiavon, G., and Smith, I. E. (2013) Endocrine therapy for advanced/metastatic breast cancer. *Hematol. Oncol. Clin. North Am.* 27, 715–36, viii
3. Ring, A., and Dowsett, M. (2004) Mechanisms of tamoxifen resistance. *Endocr. Relat. Cancer* 11, 643–58
4. Gradishar, W. J. (2004) Tamoxifen - What next? *Oncologist* 9, 378–384
5. Zhao, W., Zhang, Q., Kang, X., Jin, S., and Lou, C. (2009) AIB1 is required for the acquisition of epithelial growth factor receptor-mediated tamoxifen resistance in breast cancer cells. *Biochem. Biophys. Res. Commun.* 380, 699–704
6. Huang, B., Warner, M., and Gustafsson, J.-Å. (2014) Estrogen receptors in breast carcinogenesis and endocrine therapy. *Mol. Cell. Endocrinol.* 418, 240–4
7. Zhang, Y., Moerkens, M., Ramaiahgari, S., de Bont, H., Price, L., Meerman, J., and van de Water, B. (2011) Elevated insulin-like growth factor 1 receptor signaling induces antiestrogen resistance through the MAPK/ERK and PI3K/Akt signaling routes. *Breast Cancer Res.* 13, R52
8. Madlensky, L., Natarajan, L., Tchu, S., Pu, M., Mortimer, J., Flatt, S. W., Nikoloff, D. M., Hillman, G., Fontecha, M. R., Lawrence, H. J., Parker, B. a, Wu, a H. B., and Pierce, J. P. (2011) Tamoxifen metabolite concentrations, CYP2D6 genotype, and breast cancer outcomes. *Clin. Pharmacol. Ther.* 89, 718–25
9. Toy, W., Shen, Y., Won, H., Green, B., Sakr, R. a, Will, M., Li, Z., Gala, K., Fanning, S., King, T. a, Hudis, C., Chen, D., Taran, T., Hortobagyi, G., Greene, G., Berger, M., Baselga, J., and Chandralapaty, S. (2013) ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat. Genet.* 45, 1439–45
10. Polyak, K. (2014) Tumor heterogeneity confounds and illuminates: a case for Darwinian tumor evolution. *Nat. Med.* 20, 344–6
11. Robinson, D. R., Wu, Y.-M., Vats, P., Su, F., Lonigro, R. J., Cao, X., Kalyana-Sundaram, S., Wang, R., Ning, Y., Hodges, L., Gursky, A., Siddiqui, J., Tomlins, S. a, Roychowdhury, S., Pienta, K. J., Kim, S. Y., Roberts, J. S., Rae, J. M., Van Poznak, C. H., Hayes, D. F., Chugh, R., Kunju, L. P., Talpaz, M., Schott, A. F., and Chinnaiyan, A. M. (2013) Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* 45, 1446–51
12. Jansen, M. P. H. M., Foekens, J. A., van Staveren, I. L., Dirkzwager-Kiel, M. M., Ritsier, K., Look, M. P., Meijer-van Gelder, M. E., Sieuwerts, A. M., Portengen, H., Dorssers, L. C. J., Klijn, J. G. M., and Berns, E. M. J. J. (2005) Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J. Clin. Oncol.* 23, 732–40

13. Jansen, M. P. H. M., Ruigrok-Ritstier, K., Dorssers, L. C. J., van Staveren, I. L., Look, M. P., Meijer-van Gelder, M. E., Sieuwerts, A. M., Helleman, J., Sleijfer, S., Klijn, J. G. M., Foekens, J. A., and Berns, E. M. J. J. (2009) Downregulation of SIAH2, an ubiquitin E3 ligase, is associated with resistance to endocrine therapy in breast cancer. *Breast Cancer Res. Treat.* 116, 263–71
14. Beelen, K., Zwart, W., and Linn, S. C. (2012) Can predictive biomarkers in breast cancer guide adjuvant endocrine therapy? *Nat. Rev. Clin. Oncol.* 9, 529–41
15. Droog, M., Beelen, K., Linn, S., and Zwart, W. (2013) Tamoxifen resistance: From bench to bedside. *Eur. J. Pharmacol.* 717, 47–57
16. Liu, N. Q., Stingl, C., Look, M. P., Smid, M., Braakman, R. B. H., De Marchi, T., Sieuwerts, A. M., Span, P. N., Sweep, F. C. G. J., Linderholm, B. K., Mangia, A., Paradiso, A., Dirix, L. Y., Van Laere, S. J., Luider, T. M., Martens, J. W. M., Foekens, J. A., and Umar, A. (2014) Comparative Proteome Analysis Revealing an 11-Protein Signature for Aggressive Triple-Negative Breast Cancer. *J. Natl. Cancer Inst.* 106, Epub 2014
17. Swami, M. (2010) Proteomics: A discovery strategy for novel cancer biomarkers. *Nat. Rev. Cancer* 10, 597–597
18. Drabovich, A. P., Martínez-Morillo, E., and Diamandis, E. P. (2014) Toward an integrated pipeline for protein biomarker development. *Biochim. Biophys. Acta*,
19. Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M. C., Zimmerman, L. J., Shaddox, K. F., Kim, S., Davies, S. R., Wang, S., Wang, P., Kinsinger, C. R., Rivers, R. C., Rodriguez, H., Townsend, R. R., Ellis, M. J. C., Carr, S. a, Tabb, D. L., Coffey, R. J., Slebos, R. J. C., and Liebler, D. C. (2014) Proteogenomic characterization of human colon and rectal cancer. *Nature* 513, 382–387
20. Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabudde, N. a, Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L. D. N., Patil, A. H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S. K., Marimuthu, A., Sathe, G. J., Chavan, S., Datta, K. K., Subbannayya, Y., Sahu, A., Yelamanchi, S. D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K. R., Syed, N., Goel, R., Khan, A. a, Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T.-C., Zhong, J., Wu, X., Shaw, P. G., Freed, D., Zahari, M. S., Mukherjee, K. K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C. J., Shankar, S. K., Satishchandra, P., Schroeder, J. T., Sirdeshmukh, R., Maitra, A., Leach, S. D., Drake, C. G., Halushka, M. K., Prasad, T. S. K., Hruban, R. H., Kerr, C. L., Bader, G. D., Iacobuzio-Donahue, C. a, Gowda, H., and Pandey, A. (2014) A draft map of the human proteome. *Nature* 509, 575–81
21. Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582–7

22. Harlan, R., and Zhang, H. (2014) Targeted proteomics: a bridge between discovery and validation. *Expert Rev. Proteomics* 11, 657–61
23. Makawita, S., and Diamandis, E. P. (2010) The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry - Based approaches: Current strategies for candidate verification. *Clin. Chem.* 56, 212–222
24. Braakman, R. B. H., Tilanus-Linthorst, M. M. A., Liu, N. Q., Stingl, C., Dekker, L. J. M., Luider, T. M., Martens, J. W. M., Foekens, J. A., and Umar, A. (2012) Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. *J. Proteomics* 75, 2844–54
25. Liu, N. Q., Braakman, R. B. H., Stingl, C., Luider, T. M., Martens, J. W. M., Foekens, J. A., and Umar, A. (2012) Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J. Mammary Gland Biol. Neoplasia* 17, 155–64
26. De Marchi, T., Liu, N. Q., Stingl, C., Timmermans, M. A., Smid, M., Look, M. P., Tjoa, M., Braakman, R. B. H., Opdam, M., Linn, S. C., Sweep, F. C. G. J., Span, P. N., Kliffen, M., Luider, T. M., Foekens, J. A., Martens, J. W. M., and Umar, A. (2016) 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Mol. Oncol.* 10, 24–39
27. Dittmar, G., and Selbach, M. (2009) Novel insights into proteomic technologies and their clinical perspective. *Genome Med.* 1, 53
28. Schwappacher, R., Rangaswami, H., Su-Yuo, J., Hassad, A., Spitler, R., and Casteel, D. E. (2013) cGMP-dependent protein kinase I $\beta$  regulates breast cancer cell migration and invasion via interaction with the actin/myosin-associated protein caldesmon. *J. Cell Sci.* 126, 1626–36
29. Sobral-Leite, M., Wesseling, J., Smit, V. T. H. B. M., Nevanlinna, H., van Miltenburg, M. H., Sanders, J., Hofland, I., Blows, F. M., Coulson, P., Patrycja, G., Schellens, J. H. M., Fagerholm, R., Heikkilä, P., Aittomäki, K., Blomqvist, C., Provenzano, E., Ali, H. R., Figueroa, J., Sherman, M., Lissowska, J., Mannermaa, A., Kataja, V., Kosma, V.-M., Hartikainen, J. M., Phillips, K.-A., Couch, F. J., Olson, J. E., Vachon, C., Visscher, D., Brenner, H., Butterbach, K., Arndt, V., Holleczeck, B., Hooning, M. J., Hollestelle, A., Martens, J. W. M., van Deurzen, C. H. M., van de Water, B., Broeks, A., Chang-Claude, J., Chenevix-Trench, G., Easton, D. F., Pharoah, P. D. P., García-Closas, M., de Graauw, M., and Schmidt, M. K. (2015) Annexin A1 expression in a pooled breast cancer series: association with tumor subtypes and prognosis. *BMC Med.* 13, 156
30. Al Saleh, S., Al Mulla, F., and Luqmani, Y. a. (2011) Estrogen receptor silencing induces epithelial to mesenchymal transition in human breast cancer cells. *PLoS One* 6, e20610
31. Chu, S., Nishi, Y., Yanase, T., Nawata, H., and Fuller, P. J. (2004) Transrepression of estrogen receptor beta signaling by nuclear factor-kappaB in ovarian granulosa cells. *Mol. Endocrinol.* 18, 1919–1928
32. Bist, P., Leow, S. C., Phua, Q. H., Shu, S., Zhuang, Q., Loh, W. T., Nguyen, T. H., Zhou, J. B., Hooi, S. C., and Lim, L. H. K. (2011) Annexin-1 interacts with NEMO and RIP1 to constitutively activate IKK complex and NF- $\kappa$ B: implication in breast cancer metastasis. *Oncogene* 30, 3174–3185

33. Karin, M. (2009) NF- B as a Critical Link Between Inflammation and Cancer. *Cold Spring Harb. Perspect. Biol.* 1, a000141–a000141
34. de Graauw, M., van Miltenburg, M. H., Schmidt, M. K., Pont, C., Lalai, R., Kartopawiro, J., Pardali, E., Le Dévédec, S. E., Smit, V. T., van der Wal, A., Van't Veer, L. J., Cleton-Jansen, A.-M., ten Dijke, P., and van de Water, B. (2010) Annexin A1 regulates TGF-beta signaling and promotes metastasis formation of basal-like breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 107, 6340–5
35. Kan, R., Shuen, W. H., Lung, H. L., Cheung, A. K. L., Dai, W., Kwong, D. L.-W., Ng, W. T., Lee, A. W. M., Yau, C. C., Ngan, R. K. C., Tung, S. Y., and Lung, M. L. (2015) NF-κB p65 Subunit Is Modulated by Latent Transforming Growth Factor-β Binding Protein 2 (LTBP2) in Nasopharyngeal Carcinoma HONE1 and HK1 Cells. *PLoS One* 10, e0127239
36. Wang, H., Pan, J., Luo, L., Ning, X., Ye, Z., Yu, Z., and Li, W. (2015) NF- κ B induces miR-148a to sustain TGF- β / Smad signaling activation in glioblastoma. 1–14
37. Zhang, Z., Huang, L., Zhao, W., and Rigas, B. (2010) Annexin 1 Induced by Anti-Inflammatory Drugs Binds to NF- B and Inhibits Its Activation: Anticancer Effects In vitro and In vivo. *Cancer Res.* 70, 2379–2388
38. Yu, S., Meng, Q., Hu, H., and Zhang, M. (2014) Correlation of ANXA1 expression with drug resistance and relapse in bladder cancer. *Int. J. Clin. Exp. Pathol.* 7, 5538–5548
39. Belvedere, R., Bizzarro, V., Popolo, A., Piazz, F. D., Vasaturo, M., Picardi, P., and Parente, L. (2014) Role of intracellular and extracellular annexin A1 in migration and invasion of human pancreatic carcinoma cells. *BMC Cancer* 14, 1–15
40. Lynch, W. P., Riseman, V. M., and Bretscher, a. (1987) Smooth muscle caldesmon is an extended flexible monomeric protein in solution that can readily undergo reversible intra- and intermolecular sulfhydryl cross-linking. A mechanism for caldesmon's F-actin bundling activity. *J. Biol. Chem.* 262, 7429–7437
41. Mayanagi, T., and Sobue, K. (2011) Diversification of caldesmon-linked actin cytoskeleton in cell motility. *Cell Adh. Migr.* 5, 150–159
42. Morita, T., Mayanagi, T., and Sobue, K. (2007) Dual roles of myocardin-related transcription factors in epithelial-mesenchymal transition via slug induction and actin remodeling. *J. Cell Biol.* 179, 1027–1042
43. Yoshio, T., Morita, T., Kimura, Y., Tsujii, M., Hayashi, N., and Sobue, K. (2007) Caldesmon suppresses cancer cell invasion by regulating podosome/invadopodium formation. *FEBS Lett.* 581, 3777–82
44. Rakha, E. a., Reis-Filho, J. S., and Ellis, I. O. (2008) Basal-like breast cancer: A critical review. *J. Clin. Oncol.* 26, 2568–2581
45. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–72

46. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–805
47. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. a, Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O’Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R., and Hermjakob, H. (2013) The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41, D1063–9
48. Johnson, W. E., Li, C., and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–27
49. Hayward, J., and Carbone, P. (1977) Assessment of response to therapy in advanced breast cancer. *Br. J. Cancer*, 1978
50. Altman, D. G., McShane, L. M., Sauerbrei, W., and Taube, S. E. (2012) Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med.* 9, e1001216

## **Chapter 8**

### **Phosphoserine aminotransferase 1 is associated to poor outcome on tamoxifen therapy in recurrent breast cancer**

Tommaso De Marchi, Mieke A. Timmermans, Anieta M. Sieuwerts, Marcel Smid, Maxime P Look, Nicolai Grebenchtchikov, Fred C. Sweep, Jan G. Smits, Viktor Magdolen, Wim Verhaegh, Carolien H. van Deurzen, John A. Foekens, Arzu Umar A, John W. Martens.

Submitted for publication

## **Abstract**

**Purpose:** Resistance to anti-hormonal therapy is a major cause of death in estrogen receptor (ER) positive breast cancer. In a previous study, we detected a significant association between phosphoserine aminotransferase 1 (*PSAT1*) hyper-methylation and mRNA levels to good and poor outcome to tamoxifen treatment in recurrent disease, respectively. Our aim was to study the association of PSAT1 protein (PSAT1) levels to outcome upon tamoxifen treatment and to obtain more insight in its role in tamoxifen resistance molecular mechanisms.

**Methods:** A cohort of 279 formalin-fixed and paraffin-embedded (FFPE) tissues of ER positive, hormonal therapy naïve primary breast carcinomas captured in a tissue microarray (TMA) was immunohistochemically (IHC) stained for PSAT1. Staining was analyzed for association with patient's time to progression (TTP) and overall response on first-line tamoxifen given for recurrent disease by Cox and logistic regression analyses, respectively. Tumor PSAT1 mRNA levels previously assessed by reverse transcriptase quantitative polymerase chain reaction (RT-qPCR; n = 161) and Affymetrix GeneChip (n = 155) analyses were also evaluated. Association of PSAT1 and its associated genes to biological pathways on tamoxifen outcome in ER positive breast cancer were assessed by global test.

**Results:** PSAT1 positivity was significantly associated with shorter TTP (HR = 1.63; 95% confidence interval [CI]: 1.02 to 2.59; P = 0.039) independently of traditional predictive parameters. PSAT1 protein (IHC) and mRNA levels (RT-qPCR) subdivided in high and low showed a significant association (Mann-Whitney  $P = 0.009$ ). In the Affymetrix dataset, PSAT1 was associated to poor outcome on tamoxifen treatment (HR = 1.68; 95% CI: 1.20 to 2.36;  $P = 0.003$ ). In addition, global test results showed that genes associated to PSAT1 expression in relation to tamoxifen outcome belonged to cytokine and JAK-STAT signaling pathways.

**Conclusions:** We hereby report that PSAT1 protein and mRNA levels measured in ER positive primary tumors are significantly associated with poor clinical outcome in breast cancer patients treated with tamoxifen for recurrent disease.

## Introduction

ER (*ESR1* gene) positive breast cancer constitutes the majority (~70%) of all breast malignancies [1]. Targeted treatment for these patients comes in the form of anti-hormonal therapies, which are used in both the adjuvant and recurrent settings [2]. In particular, the ER antagonist tamoxifen decreases recurrence rates by almost a half (i.e. recurrence rate ratio = 0.59) and breast cancer mortality by a third (death rate ratio = 0.66) when administered for 5 years in the adjuvant setting [3]. In the recurrent setting, nearly half of the patients manifests intrinsic resistance to the drug, while the other half inevitably develops acquired resistance later on [4, 5]. Several mechanisms of tamoxifen resistance have been described, such as the upregulation of ER transcriptional co-activators [6] or the expression of ER isoforms [7, 8]. Furthermore, activation of PI3K/MAPK [9] or Her2/neu [10, 11] pathways as well as the dysregulation of tamoxifen metabolizing enzymes [12], on one side provide tumor cells with alternative growth pathways, while on the other side reduce tamoxifen active metabolite (i.e. endoxifen) levels. Mutations in ER (e.g. p.Tyr537Ser) have also been associated to its constitutive activation [7, 13, 14]. In addition to this, general cancer specific factors like DNA methylation of cytosine phosphoguanine dinucleotides (CpG), which has been established as an important mechanism that regulates gene expression in breast cancer [15, 16], have shown to play an important role in many cancer processes such as targeting and silencing genes involved in tumor suppression (e.g. BRCA1) [17], epithelial-to-mesenchymal transition (EMT), and invasion [18, 19] can all contribute to overcome or bypass tamoxifen treatment induced effects. In a previous study, we have linked hyper-methylation in the promoter region of the PSAT1 gene to a favorable outcome of tamoxifen treatment in ER positive breast cancer patients; conversely high PSAT1 mRNA levels were associated to tamoxifen resistance [20]. PSAT1 is a gene that encodes an amino-transferase enzyme involved in the conversion of phospho-pyruvate, which derives from oxidation of 3-phosphoglycerate, to phosphoserine. Phosphoserine is then converted into serine by the enzyme phosphoserine-phosphatase and further converted into glycine in order to feed the nucleotide biosynthesis pathway. Though the serine biosynthetic pathway has been shown to be a critical factor in breast cancer tumorigenesis [21] and therapy resistance [20], no study has further investigated its association to clinical characteristics. In addition to this, neither further marker verification, nor an association study of PSAT1 protein levels with tamoxifen therapy outcome, have been performed. In the light of this, the aim of this study was to verify whether PSAT1 protein levels are linked to tamoxifen therapy outcome. We assessed PSAT1 protein

levels by IHC in a cohort of FFPE tissues and analyzed its association with tamoxifen therapy outcome. Furthermore, gene expression data of a cohort of ER positive breast carcinomas was used to gain insight into the role of PSAT1 in tamoxifen resistance through global test analysis.

**Materials and Methods**

A schematic representation of the analysis workflow is shown in Figure 8.1A.

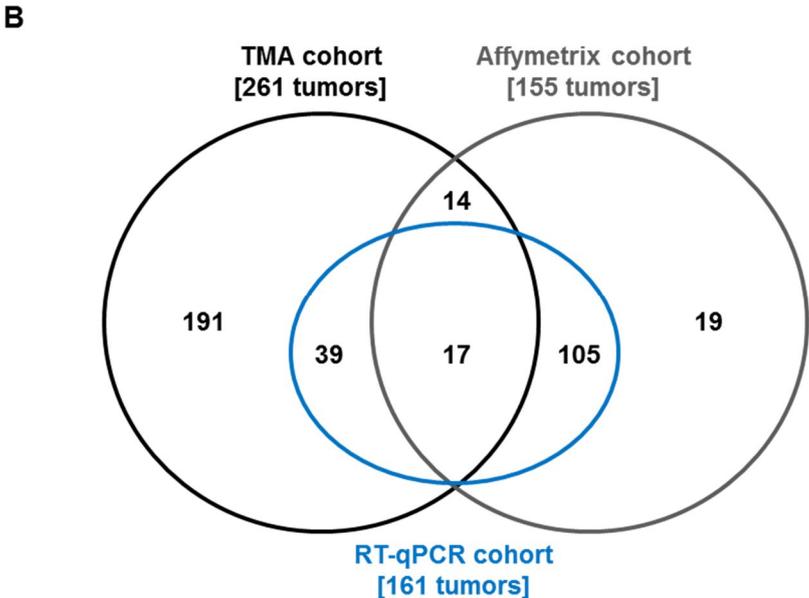
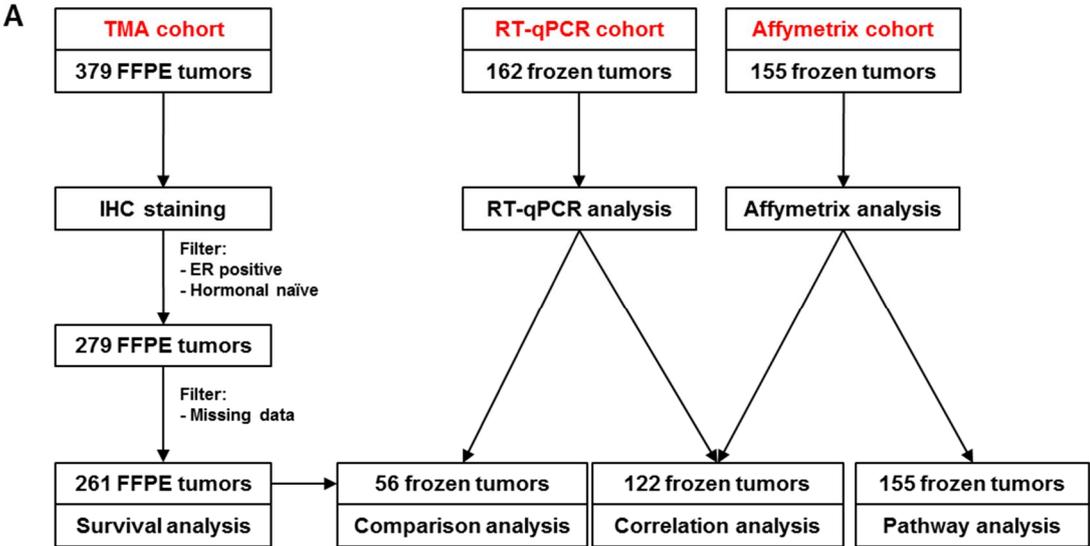


Figure 8.1. Schematic overview of experimental workflow.

Panel A: a total of 379 FFPE tissues were captured on a tissue micro-array and analyzed by IHC. After filtering for ER positivity and hormonal naïve tumors, a total of 279 samples remained. Further filtering for missing data after IHC analysis yielded a panel of 261 tumors, on which survival analysis for the association of PSAT1 protein levels to TTP was performed. Parallel to this, PSAT1 mRNA expression was measured by RT-qPCR (n = 161) and Affymetrix chip (n = 155) approaches on frozen tumor specimens. These data were used for comparison between PSAT1 mRNA and protein levels (TMA and RT-qPCR; n= 56), correlation analysis (RT-qPCR and Affymetrix; n = 122), and pathway analysis (Affymetrix only; n = 155). Panel B shows tumor sample overlap between TMA, RT-qPCR and Affymetrix sets.

Acronyms: ER: estrogen receptor; FFPE: formalin-fixed paraffin-embedded; IHC: immunohistochemistry; TMA: tissue microarray TTP: time to progression; RT-qPCR: quantitative reverse transcriptase polymerase chain reaction.

### *Patient cohorts*

The Medical Ethical Committee of the Erasmus Medical Center Rotterdam, the Netherlands, approved our study design (MEC 02.953). This study used previously described FFPE tumor tissues which were incorporated into a TMA [22]. A total of 379 breast cancer tissues derived from patients who underwent tumor resection (between 1985 and 2000), and were treated with first line tamoxifen for recurrent disease, were included in the study. A specialized breast pathologist (CHMvD) reviewed all the primary tumor tissue histologic subtypes according to the world health organization (WHO) and histologic grade was defined according to the modified Bloom-Richardson score, which takes into account tubule formation, mitotic activity and nuclear pleiomorphism [23]. Out of this set, only ER positive primary tumors from patients who did not receive any adjuvant (i.e. post-surgical resection) hormonal therapy were included for statistical analysis. This led to the inclusion of 279 tissues derived from patients with ER positive primary tumors (Table S-1), of whom response data were collected according to the standard International Union Against Cancer criteria [24]. A total of 10 (3.58%) and 42 (15.05%) patients showed complete (CR) and partial remission (PR), respectively. One hundred and fifty seven patients showed no change (NC) of disease, of whom 27 (9.68%) displayed NC for less ( $\leq$ ) than 6 months, while 130 (46.59%) showed NC for longer ( $>$ ) than 6 months (defined as stable disease [SD]). A total of 70 (25.09%) patients displayed progressive disease (PD). Clinical benefit was defined as CR + PR + SD patients (n = 182; 65.22%), while objective response was defined as CR + PR only (n = 52; 18.63%). This retrospective study used coded primary tissues, in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands

(<http://www.federa.org/codes-conduct>). Reporting Recommendations for Tumor Marker Prognostic Studies were followed [25].

For gene expression analysis, a total of 155 ER positive fresh frozen primary breast carcinomas were collected, which were derived from patients treated with tamoxifen therapy upon disease recurrence. Total RNA was analyzed by GeneChip® Human Genome U133 Plus 2.0 and Perfect Match Arrays (Affymetrix, Santa Clara, CA, USA). Previously identified [26, 27], tamoxifen therapy outcome groups in this set were defined based on TTP: patients displaying progression before ( $\leq$ ) 6 months after start of tamoxifen therapy were classified as poor outcome, while patients that showed progression of disease after ( $>$ ) 6 months were defined as good outcome. This set comprised 102 good and 53 poor outcome patients, respectively (Table S-2). For a subset of these tumors ( $n = 122$ ), and for 39 additional specimens only included in the TMA, quantitative reverse transcriptase PCR (RT-qPCR) data obtained from total RNA isolated from ER positive fresh frozen primary breast carcinomas was also available (Figure 8.1B). These RT-qPCR data were only used for correlation analyses.

### *Cell culture*

Breast cancer cell lines (CAMA-1, DU-4475, MDA-MB-175, and EVSA-T) were cultured in RPMI medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) as well as antibiotics (100  $\mu\text{g}/\text{mL}$  Penicillin, 20  $\text{ng}/\text{mL}$  Streptomycin and 80  $\mu\text{g}/\text{mL}$  Gentamycin) in a humidified atmosphere with 95% air and 5%  $\text{CO}_2$ . For protein isolation, cells were gathered from culture dishes using a scraper and suspended in a 1.5 mL tube in 1x PBS. Cells were then sonicated at 70% amplitude (Bransons Ultrasonics, Danbury, CT, USA) using a horn sonifier, and collected supernatants were transferred into a new tube and stored at  $-80^\circ\text{C}$  for further processing. For RNA isolation, cells were lysed (Qiagen lysis buffer) in the flask at 70% confluency and stored at  $-80^\circ\text{C}$  awaiting further processing.

### *RNA extraction and gene expression analysis*

To measure mRNA transcript levels, total RNA from cell cultures was isolated with the RNeasy Mini Kit (Qiagen) and from clinical tissues using RNA-Bee (Campro Scientific,

Veenendaal, The Netherlands) according to the manufacturer's instructions, as previously described [29]. For RT-qPCR, complementary DNA synthesis (RT reaction) was generated using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, Amsterdam, The Netherlands), followed by incubation at 37°C with 0.1 U/ $\mu$ L ribonuclease H (Thermo Scientific) for 30 min. Real-time PCRs were performed on cDNA generated from 10 ng total RNA and normalized using the  $dCq$  method on the average of 3 reference genes (HMBS, HPRT1 and TBP). Quantification of PSAT1 was performed using the TaqMan probe-based gene expression assay Hs00253548\_m1 specific for splice variant beta (Applied Biosystems/Life Technologies, Warrington, WA, USA) as previously described [20, 29]. For Affymetrix gene expression profiling, total RNA samples were DNase treated and cleaned with the NucleoSpin RNA II kit according the manufacturers instruction (Machery-Nagel, Dueren, Germany) and shipped to ServiceXS (Leiden, The Netherlands) for downstream processing with the 3'IVT express kit and hybridization on the Human Genome (HG) U133 Plus 2.0 array (n = 20) and HG U133 Perfect Match.

#### *Anti-PSAT1 antibody generation*

In order to select protein surface exposed epitopes for antibody production, prediction models were generated in expasy (<http://expasy.org/>) and matched with publicly available crystal structures of E. coli and B. alcalophilus PSAT1 homologues. Putative epitope regions were selected based on absence of secondary structures (i.e. excluding regions involved in  $\alpha$ -helix or  $\beta$ -strand structures) and being present in both PSAT1 splice variants [28]. A total of two peptides selected: PSAT1-A(DYKGVGISVLEMSHRSS, aa 31-47) and PSAT1-B (KLGSYTKIPDPSTWNLNP; aa 127-144). Both peptides were synthesized adding a Cys residue at the N-terminus for disulfide linkage to keyhole limpet hemocyanin (KLH). Rabbit pre-immunization sera were tested for absence of immune reaction by Western Blot analysis against full length recombinant PSAT1 (sequence including exon 8). KLH-conjugated peptides were injected into rabbits, which received boost immunizations at day 20, 30, 40, 61, 75, 90, and 104. Sera were then collected at day 120. Peptide synthesis, conjugation to KLH, and rabbit immunization steps were performed by Pineda Antibody Service (Berlin). For antibody purification, immunoaffinity columns were prepared using recombinant human full length PSAT1 (including exon 8). Recombinant PSAT1 was expressed in E. coli as a N-terminal histidine tagged protein and was purified via Ni<sup>2+</sup>-NTA (Qiagen) affinity

chromatography under denaturing and slightly reduced conditions (purity > 95%). After protein refolding, PSAT1 was recovered in soluble form in a 1x PBS and 1 mM DTT solution.

PSAT1-linked column was prepared by mixing ~2 mL of AffiGel-10<sup>®</sup> and AffiGel-15 (BioRad) in a 1:1 ratio. Sorbent mixture was then pre-treated by sequential washings: 10 mL of isopropanol, 5 mL deionized water, and 5 mL of 50 mM HEPES buffer in deionized water (pH 7.4). A total of 0.5 mg of soluble recombinant PSAT1 dissolved in 5 mL of 50 mM HEPES was mixed with the sorbent and incubated overnight at 4°C. Sorbent was then washed with 1x PBS (5 x 5 mL), followed by blocking solution incubation (100 mM glycine in 1x PBS) for 1 h at room temperature. Further washings with 1 mL of 4 M guanidinium chloride in 1x PBS (1 x 5 mL) and 40 mL of 1x PBS (1 x 5 mL) were performed. Rabbit-derived blood was centrifuged at 1711 g for 10 min and 4 mL of rabbit serum were injected onto the sorbent (three sequential times), followed by a 25 mL 1x PBS wash. Antibodies were then eluted with 4 mL of 20 mM glycine in 1x PBS (pH 2.4), and collected in tubes to which pH compensatory solution (200 µL of 1x PBS; pH 11.5) was added in order to achieve pH 7.5. Protein concentration was assessed by measuring absorbance at 260/280 nm. Eluate aliquots were then mixed with 100% v/v glycerol in a 1:1 ratio and stored at -20°C. Concentration of each antibody aliquot was 0.059 mg/mL as measured by Bradford assay.

#### *Anti-PSAT1 antibody titration*

To determine the best antibody concentration for IHC staining of breast cancer tissues, FFPE breast cancer cell lines (i.e. CAMA-1, EVSA-T, DU-4475 and MDA-MB-175) captured on TMA slides together were first stained at different anti-PSAT1 antibody dilutions (i.e. 1:10, 1:20, 1:40, 1:80, 1:160; for methods, see below under IHC staining). For each cell line, IHC stainings were compared to PSAT1 mRNA levels as measured by RT-qPCR. As the 1:40 dilution anti-PSAT1 antibody stainings of breast cancer cell lines showed comparable results with RT-qPCR data (Figure S-1), this dilution was selected as the optimal one for further ICH staining of tumor tissues.

### *Tissue micro-array*

TMA was prepared using an ATA 27 (Beecher Instruments, Sun Prairie, WI, USA). For each paraffin block, the tumor area was delineated by a specialized breast pathologist (CHMvD), after which tumor tissue cores of 0.6 mm were punched and in triplicate transferred to a recipient TMA block.

### *Immunohistochemical staining and analysis*

Sections (5 µm) of FFPE cell lines and tissues captured on the same TMA were incubated at 60°C (30 min) and washed in xylene (3 x 5 min) for de-paraffination. Decreasing concentrations of ethanol were used for tissue re-hydration: 100% ethanol (1 x 5 min, 2 x 2 min), 70% ethanol (1 x 2 min), 50% ethanol (1 x 2 min), distilled water (1 x 2 min). Incubation with DAKO (Agilent Technologies Inc., Glostrup, Denmark) antigen retrieval (pH 6.0) solution diluted 1:10 in MilliQ water was performed at 95°C for 40 min. After cooling down to room temperature, slides were subjected to sequential washes: 1x PBS (3 x 5 min), 0.003% H<sub>2</sub>O<sub>2</sub> in 1x PBS (i.e. blocking of endogenous peroxidase activity; 1 x 10 min), 1x PBS (3 x 5 min), 5% BSA in 1x PBS (i.e. blocking solution; 1 x 30 min). Anti-PSAT1 rabbit polyclonal primary antibody diluted 1:40 in Dako antibody diluent was added to slides and incubated overnight at 4°C. For the negative control, the anti-PSAT1 primary antibody was replaced by normal negative control rabbit immunoglobulin fraction (ID: X0903; Dako). Slides were then washed with 1x PBS (3 x 5 min), and DAKO Envision<sup>®</sup> labeled polymer HRP-Rabbit was added to each slide (200 µl per slide) and incubated for 45 min at room temperature. A wash step with 1x PBS (3 x 5 min) was performed followed by a development step of the HRP with 1:15 solution of DAB<sup>+</sup> chromogen in DAB<sup>+</sup> substrate buffer (100 µl per slide; 1 x 10 min in the dark). A nuclei counterstaining with Mayer's hematoxylin (Klinipath) for 30 sec was performed followed by rinsing with tap water (1 x 5 min). Dehydration, by 50% ethanol (2 x 2 min), 70% ethanol (2 x 2 min), 100% ethanol (2 x 2 min, 1 x 5 min), and xylene (2 x 2 min, 1 x 5 min). Cover glasses were mounted on slides with Pertex and were left to dry.

Slides were digitalized with the Hamamatsu Nanozoomer 2.0 HT and analyzed manually from a computer screen using Slidepath TMA database software (Leica Microsystems, Solms, Germany). PSAT1 stained breast carcinoma cells were scored by an experienced researcher in

a blind manner for both intensity (i.e. negative; weak; weak-moderate; moderate; moderate-strong; strong) and quantity (i.e. 0%; 1-5%; 6-10%; 11-25%; 26-50%; >50%) of stained carcinoma cells. Triplicate scores accompanying these images of the cores were then verified and consolidated by a second experienced researcher, which was extensively trained by a specialized breast pathologist.

### *Statistical analysis of IHC data*

PSAT1 TMA scores were stringently filtered for missing data points (i.e. lack of triplicate analysis per tissues, which derived from either lack of core or lack of enough [ $>30$ ] tumor cells in at least one core), leading to a set of 261 tissue samples. Because PSAT1 positive cases showed strong and ubiquitous staining, PSAT1 protein expression was scored as either absent (i.e. PSAT1 negative: 0% stained carcinoma cells) or present (i.e. PSAT1 positive:  $\geq 1\%$  stained carcinoma cells). PSAT1 (negative/positive) association to clinical and histopathological characteristics was assessed by Fisher's exact test (i.e. age, menopausal status, number of positive lymph nodes, disease free interval (DFI), and progesterone receptor status),  $\chi^2$  test (dominant site of relapse), and  $\chi^2$  test for trend (tumor size, tumor differentiation). Association of PSAT1 expression to TTP and response criteria was tested by Cox and logistic regression, respectively. Patient age (cutoff: 55 years), disease free interval (cutoff: 12 months), dominant site of relapse (bone, visceral, loco-regional, bone and other), progesterone receptor (PgR) positivity, HER2 overexpression, and degree of tumor differentiation (Bloom-Richardson: good, moderate, poor) were included in all regression analyses. Variables that did not display any significant association with TTP or any response criteria were excluded from the multivariate regression model (step-down analysis). Fisher exact test,  $\chi^2$  test,  $\chi^2$  test for trend, Cox regression, logistic regression, HRs, odds ratios (OR), and 95% CIs were calculated in Stata (v 13.1; Stata Corp, College Station, TX, USA).

### *Gene expression analysis*

Raw CEL files for both Affymetrix platforms were processed using fRMA with ‘robust weighted average’ (i.e. HG U133 Plus 2.0; n = 20) and ‘random effect’ (i.e. HG U133 Perfect Match; n = 135) as summarization methods. fRMAvecs from the Plus 2.0 array were used for both batches [30]. Because the resulting final set was measured in 2 separate batches, a ComBat algorithm-based normalization was applied to correct for batch effects [26]. The 155 samples gene identifiers and Affymetrix probe intensities were imported in Microsoft Excel. For probes annotated to the same genes, probes were selected based on highest variability and median level, leading to a final list of 19,042 reliable probes.

### *Statistical and pathway analyses of gene expression data*

To assess whether PSAT1 protein and mRNA levels measurements were comparable, PSAT1 Log<sub>2</sub> probe intensities and RT-qPCR dCq values were classified according to the IHC score (categories: negative vs. positive), and tested for differences by Mann-Whitney test. To assess whether *PSAT1* mRNA levels measured by Affymetrix and RT-qPCR were comparable, a Spearman correlation analyses was performed.

To assess whether pathway analysis would achieve not only biological, but also clinical significance, patients were stratified according to their median *PSAT1* expression (i.e. low: Log<sub>2</sub> intensity < median; high: Log<sub>2</sub> intensity ≥ median). Survival curves were plotted and differences in TTP were assessed by Log-rank test. Association of KEGG [31] pathways to PSAT1 and its associated genes was performed through global test [32, 33]. Significant pathways out of global test were selected based on Holm-Bonferroni corrected significance ( $P < 0.05$ ). Global test was performed in R environment (v2.11.1)

*PSAT1* mRNA expression was also associated to a 152-gene signature associated with tumor infiltrating lymphocytes (TIL) [34, 35]. Average expression of the TIL gene signature was calculated for each sample to derive a global TIL-score for every sample. *PSAT1* association to the TIL signature was evaluated by Spearman correlation. Furthermore, difference in *PSAT1* mRNA expression between high and low TIL-score (median cutoff) was evaluated by Mann-Whitney test.

Log-rank, Mann-Whitney and Spearman correlation analyses were all performed in GraphPad (v5.0).

## Results

### *Association of PSAT1 protein to clinical variables*

PSAT1 expression was observed in a small subset (N = 25, 9.6%) of ER positive breast tumors by IHC, while the majority of cores did not show any detectable, or less than 1% positively stained PSAT1 cells (N = 236, 90.7%). Examples of PSAT1 protein IHC stainings are shown in Figure 8.2A. PSAT1 was expressed both in the cytoplasm and the nucleus of carcinoma cells, with no difference between the two stainings in terms of intensity and quantity of stained cells (data not shown). Testing for association of PSAT1 protein expression with clinical and histo-pathological parameters showed that PSAT1 expression was significantly associated with poor grade tumors ( $\chi^2$  test for trend  $P < 0.001$ ) and local relapse ( $\chi^2$  test  $P = 0.009$ ; Table 8.1).

**Table 8.1. Association of PSAT1 protein expression to clinical and histo-pathological characteristics.**

	<b>Patients included in analysis</b>	<b>PSAT1 negative</b>	<b>PSAT1 positive</b>	<b><i>P</i>†</b>
<b>Total</b>	261 (100.0)	236 (100.0)	25 (100.0)	
<b>Age*</b>				
≤ 55 years	99 (37.9)	89 (37.7)	10 (40.0)	0.831
> 55 years	162 (62.1)	147 (62.3)	15 (60.0)	
<b>Menopausal status*</b>				
Premenopausal	68 (26.1)	61 (25.8)	7 (28.0)	0.813
Postmenopausal	193 (73.9)	175 (74.2)	18 (72.0)	
<b>Tumor size</b>				
T1 (≤2cm)	112 (42.9)	101 (42.8)	11 (44.0)	0.899
T2 (2-5cm) + Tx	127 (48.7)	115 (48.7)	12 (48.0)	
T3 (>5cm) + T4	22 (8.4)	20 (8.5)	2 (8.0)	
<b>Tumor Good Moderate Poor</b>				
Good	42 (16.1)	41 (17.4)	1 (4.0)	< 0.001
Moderate	143 (54.8)	136 (57.6)	7 (28.0)	
Poor	75 (28.7)	58 (24.6)	17 (68.0)	

<b>Involved lymph</b>				
<b>0</b>	87 (33.3)	78 (33.1)	9 (36.0)	0.822
<b>≥ 1</b>	166 (63.6)	151 (64.0)	15 (60.0)	
<b>Disease free interval</b>				
<b>≤ 12 months</b>	45 (17.2)	40 (16.9)	5 (20.0)	0.780
<b>&gt; 12 months</b>	216 (82.8)	196 (83.1)	20 (80.0)	
<b>Dominant site of</b>				
<b>Loco-regional</b>	28 (10.7)	22 (9.3)	6 (24.0)	0.009
<b>Bone</b>	104 (39.8)	101 (42.8)	3 (12.0)	
<b>Visceral</b>	56 (21.5)	50 (21.2)	6 (24.0)	
<b>Bone and other</b>	73 (28.0)	63 (26.7)	10 (40.0)	
<b>PgR<sup>†</sup></b>				
<b>Negative</b>	69 (26.4)	63 (26.7)	6 (24.0)	1.000
<b>Positive</b>	191 (73.2)	172 (72.9)	19 (76.0)	

\* Age and menopausal status were assessed at start of tamoxifen therapy.

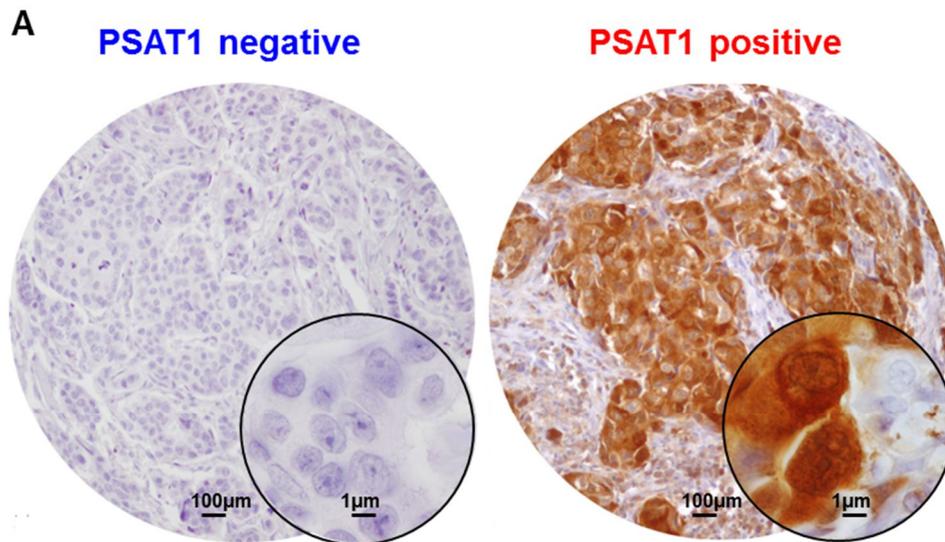
\*\* Tumor differentiation was evaluated through Scarff-Bloom-Richardson grading system

†Missing data not reported

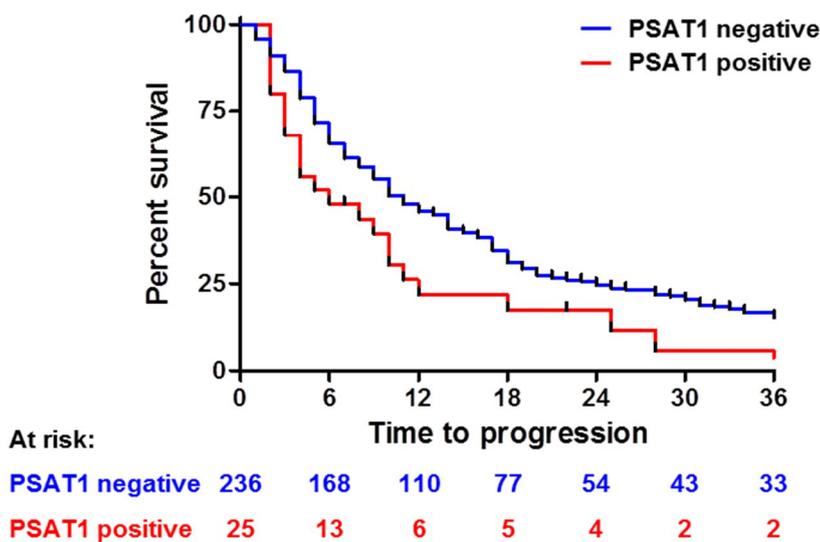
‡ PSAT1 association to clinical parameters was assessed by Fisher's exact test (age, menopausal status, number of involved lymph nodes, disease free interval and PgR),  $\chi^2$  test (dominant site of relapse), and  $\chi^2$  test for trend (tumor size, grade).

Acronyms: PgR: progesterone receptor

Next, to evaluate whether PSAT1 IHC stainings (i.e. negative vs. positive) were associated to patient outcome, Cox and logistic regression analyses were performed on the data derived from the 261 stained tissue cores. PSAT1 protein expression was significantly associated with shorter TTP both in univariate (HR = 1.77; 95% CI: 1.03 to 3.02;  $P = 0.037$ ; Figure 8.2B) and multivariate analyses independent of traditional predictive factors (i.e. DFI, dominant site of relapse and tumor differentiation; HR = 1.63; 95% CI: 1.02 to 2.59;  $P = 0.039$ ; Table 8.2).



**B** HR = 1.77; 95% CI: 1.03 to 3.02; Log-rank  $P = 0.037$



**Figure 8.2. Immunohistochemical and survival analyses of PSAT1.**

Breast carcinoma IHC stained tissues either displayed high or low PSAT1 protein levels (panel A). Kaplan-Meier analysis showed that high expression of PSAT1 protein was significantly associated to shorter TTP when compared to tumors with low PSAT1 levels (panel B).

A trend was observed in univariate logistic regression analysis for the association of PSAT1 protein levels with clinical benefit (i.e. CR +PR + SD; OR = 0.45; 95% CI: 0.19 to 1.02; univariate Logistic regression  $P = 0.057$  Table S-3), while no association with objective response (i.e. CR + PR only) was observed (OR = 0.26; 95% CI 0.15 to 1.81; univariate Logistic regression  $P = 0.304$ ; Table S-4). Overall, these data provide evidence that PSAT1

protein expression is associated with short time to progression in patients treated with tamoxifen for recurrent disease.

**Table 2. Cox regression analysis for TTP of PSAT1 stained tumors.**

	<b>n of</b>	<b>HR</b>	<b>Univariate 95% CI</b>	<b>P</b>	<b>HR</b>	<b>Multivariate 95% CI</b>	<b>P</b>
<b>PSAT1</b>							
<b>Negative</b>	236	1.0			1.0		
<b>Positive</b>	25	1.7	1.03 to	0.037	1.6	1.02 to	0.039
<b>Age*</b>							
<b>≤ 55 years</b>	99	1.0			1.0		
<b>&gt; 55 years</b>	162	0.5	0.38 to	<0.001	0.5	0.42 to	<
<b>Disease free interval</b>							
<b>≤ 12 months</b>	41	1.0					
<b>&gt; 12 months</b>	220	0.6	0.49 to	0.040	0.6	0.42 to	0.006
<b>Dominant site of</b>							
<b>Loco-regional</b>	28	1.0					
<b>Bone</b>	104	1.6	1.05 to	0.030	1.9	1.17 to	0.009
<b>Visceral</b>	56	1.5	0.92 to	0.100	1.8	1.11 to	0.017
<b>Bone and other</b>	73	1.8	1.10 to	0.019	1.9	1.17 to	0.009
<b>PgR**</b>							
<b>Negative</b>	69	1.0					
<b>Positive</b>	191	0.8	0.60 to	0.134			
<b>Her2 status**</b>							
<b>Negative</b>	210	1.0					
<b>Positive</b>	49	1.2	0.87 to	0.265			
<b>Tumor</b>							
<b>Good</b>	42	1.0			1.0		
<b>Moderate</b>	143	1.6	1.13 to	0.010	1.5	1.02 to	0.040
<b>Poor</b>	75	2.4	1.60 to	<	2.0	1.30 to	0.002

\*Age was assessed at start of tamoxifen therapy.

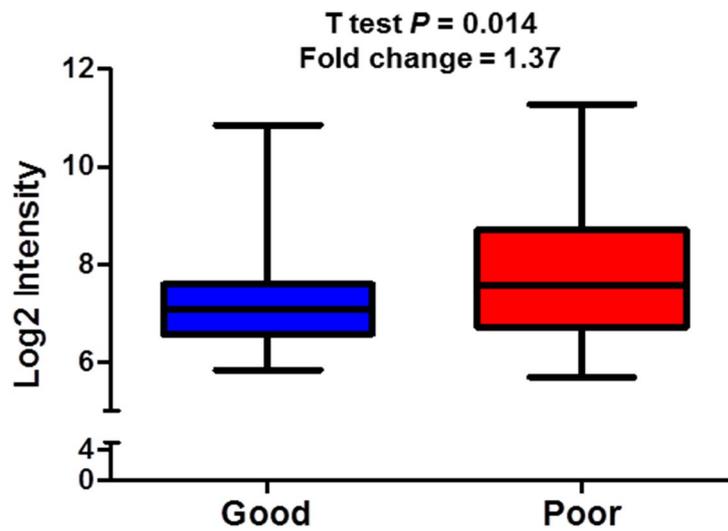
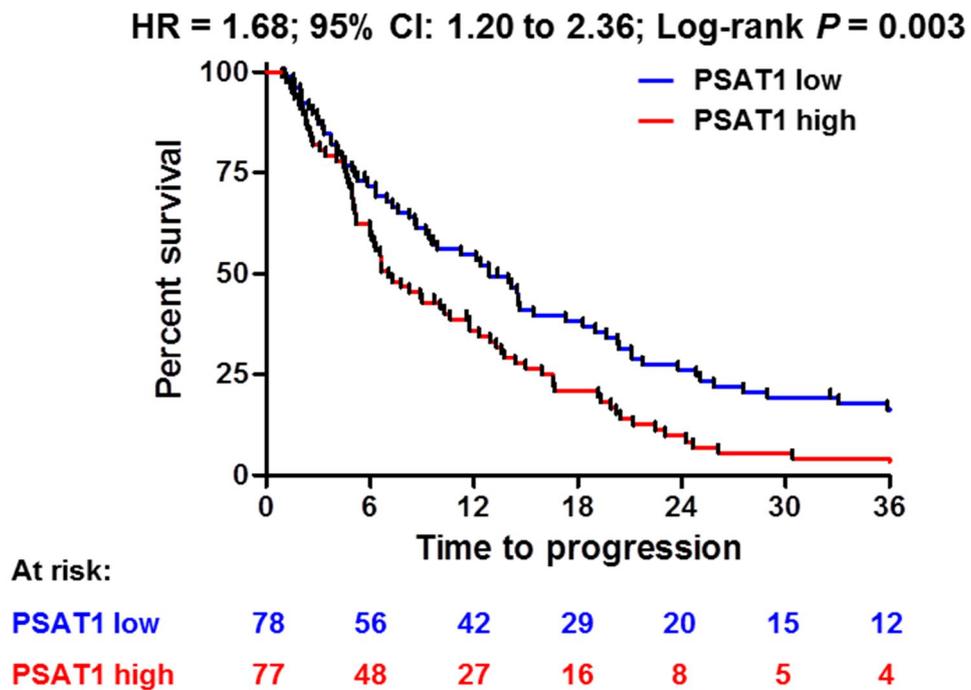
\*\*missing data not reported.

### *Comparison between PSAT1 mRNA and protein levels*

To assess whether *PSAT1* mRNA and protein expression were associated, tumors in which *PSAT1* mRNA was measured by RT-qPCR (n = 56) were stratified according to IHC results (i.e. *PSAT1* positive and *PSAT1* negative). A significant difference (Mann-Whitney  $P = 0.009$ ) was observed between *PSAT1* mRNA levels of *PSAT1* positive and negative tumors (Figure S-2A). Comparison between mRNA levels measured by Affymetrix platform and stratified according to IHC data (i.e. absent vs. present) did not show any significance (Mann-Whitney  $P = 0.133$ ), probably due to imbalances in group distributions (i.e. *PSAT1* positive = 3; *PSAT1* negative = 28; Figure S-2B) and the small overlap between the Affymetrix and TMA sets. However, *PSAT1* mRNA levels measured by RT-qPCR and Affymetrix in 122 tumors showed a significantly positive correlation (Spearman  $r = 0.74$ ;  $P < 0.001$ ; Figure S-3). These data confirm that *PSAT1* levels show a similar trend when evaluated either at the mRNA level or at the protein level.

### *Clinical significance of PSAT1 in the gene expression cohort and pathway analysis*

We used an Affymetrix GeneChip analyzed dataset (n = 155; 102 x good outcome; 53 x poor outcome) to define which genes were associated with *PSAT1*. First, we evaluated the expression of key breast cancer markers: *ESR1*, the gene encoding for ER, was found to be significantly downregulated in the poor outcome group after tamoxifen therapy (t test  $P = 0.009$ ; fold change = 0.67). Key breast cancer prognostic markers, such as *PGR* (t test  $P = 0.168$ ; fold change = 0.76) and *ERBB2* (t test  $P = 0.123$ ; fold change = 1.23), showed no change between the two patient categories. When assessing *PSAT1* levels, a significant enrichment was found in the poor outcome patient group (t test  $P = 0.014$ ; fold change = 1.37; Figure 8.3A). To confirm that *PSAT1* expression was also associated with shorter TTP in this cohort, we conservatively stratified patients at the median in *PSAT1* low and high expressing groups (cutoff: median Log<sub>2</sub> level = 7.086) and performed a survival analysis. Kaplan-Meier curves showed that patients whose tumors expressed high levels of *PSAT1* suffered from faster tumor progression when compared to the *PSAT1* low group (HR = 1.68; 95% CI: 1.20 to 2.36; Log-rank  $P = 0.003$ ; Figure 8.3B).

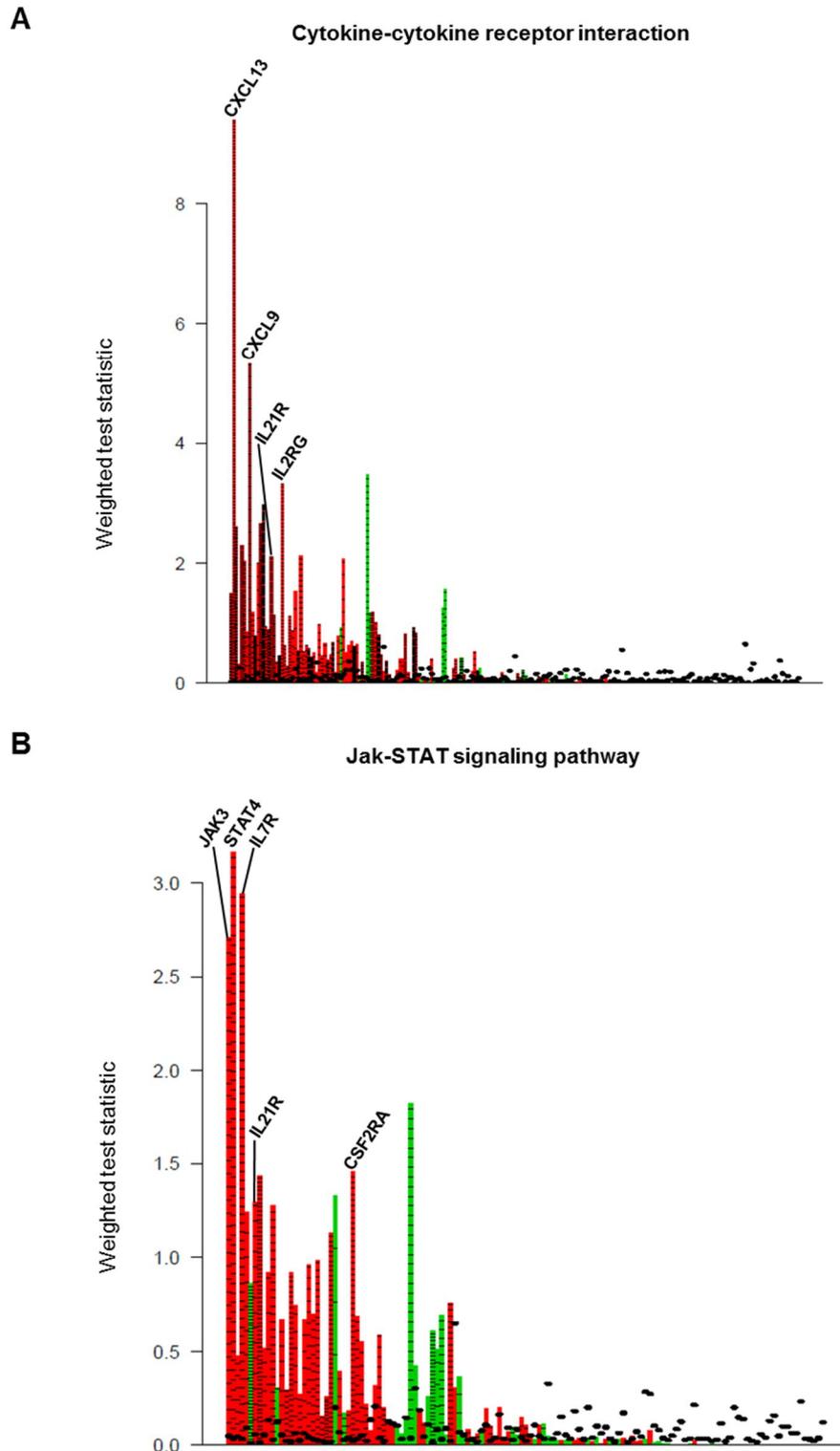
**A****B**

**Figure 8.3. Statistical and survival analyses of *PSAT1* mRNA levels by gene expression.**

*PSAT1* mRNA levels were assessed by Affymetrix GeneChip. Not only *PSAT1* showed significant enrichment in poor outcome patients (t test  $P = 0.014$ ; panel A). Kaplan Meier analysis confirmed that high expression of *PSAT1* is associated to shorter TTP (panel B).

The set was stratified based on median *PSAT1* levels (categories: *PSAT1* low, *PSAT1* high), and then analyzed by global test to investigate which (KEGG) pathways were associated with

*PSAT1* expression. A total of 8 pathways were significantly associated with *PSAT1* expression (adjusted Holm-Bonferroni  $P < 0.05$ ; Table S-5). As expected, the serine, glycine and threonine metabolism ( $P < 0.001$ ; Table S-6) was the most significantly enriched pathway which, in addition to *PSAT1*, showed enrichment of its upstream enzyme *PHGDH*. In addition to this, two additional pathways which comprised a large number of genes associated to high *PSAT1* expression were found significantly enriched: cytokine-cytokine receptor interaction ( $P = 0.001$ ; Figure 8.4A; Table S-7) and JAK-STAT signaling ( $P = 0.004$ ; Figure 8.4B; Table S-8) pathways. The cytokine-cytokine receptor interaction pathway comprised molecules involved in mesenchymal cell growth and immune cell signaling, such as *PDGFRA* and *IL4R*, respectively. On the other hand, most of the genes associated to the JAK-STAT signaling pathway comprised molecules involved in immune cell signaling, such as *JAK3* and *IL21R*. These data suggested an involvement of immune cell signaling in conjunction with serine biosynthesis.



**Figure 8.4. *PSAT1* expression associated genes in the gene expression dataset.**

The 155 tumors in the Affymetrix cohort were stratified according to *PSAT1* expression. All genes were annotated for KEGG terms and Global test was performed to assess which terms were associated to *PSAT1* expression. Panel A and B display the top 2 pathways associated to *PSAT1*: caspase cascade in apoptosis (A) and IL7 signaling (B). Each bar represents a gene in the pathway, with horizontal bars representing one standard

deviation away from the reference point. Red and green columns represent association to high and low expression of *PSAT1*, respectively

To further assess the association of *PSAT1* to the presence of immune cells, we derived the expression levels of a 152-gene signature which has been associated to tumor infiltrating lymphocytes (TILs) in breast cancer [34, 35]. We calculated the average TIL-score (see Material and Methods section) and evaluated its correlation to *PSAT1* mRNA expression in our 155 samples gene expression dataset. A significant, albeit weak, correlation between the TIL signature and *PSAT1* expression was found (Spearman  $r = 0.239$ ;  $P = 0.003$ ; Figure S-4A). Furthermore, *PSAT1* was found to be significantly enriched in the high TIL signature high expression group (median cutoff; Mann Whitney  $P = 0.009$ ; Figure S-4B).

## Discussion

Resistance to tamoxifen therapy is a major cause of death in ER positive recurrent breast cancer [36]. Several molecular mechanisms have been associated to tamoxifen resistance, but few of these have been verified on clinical specimens. In a previous study, *PSAT1* inactivation by promoter methylation, and consequently low mRNA levels, have been associated to good outcome to tamoxifen treatment [20]. We have here evaluated whether PSAT1 protein levels were associated to tamoxifen therapy outcome. In addition to this, we used gene expression data from a large cohort of ER positive breast cancer patients to further confirm *PSAT1* association to short TTP at the mRNA level, compare PSAT1 protein and mRNA levels, and performed global test analysis to potentially elucidate molecular mechanisms related to *PSAT1* expression and tamoxifen resistance.

Our IHC staining of a large cohort of paraffin-embedded breast cancers showed that only a small group of specimens displayed detectable PSAT1 protein expression, possibly constituting a subgroup within ER positive tumors. Statistical analysis of IHC stainings showed that PSAT1 expression was associated to poor tumor differentiation and loco-regional metastases. Furthermore, Cox regression analyses showed that PSAT1 expression was significantly associated to shorter TTP independently of other predictive factors (e.g. DFI). On the other side, logistic regression analyses showed that PSAT1 expression showed no association to clinical benefit or objective response, probably due to the fact that only a small portion of tumors displayed detectable PSAT1 protein expression. Small subset size could

have also influenced the comparison between PSAT1 protein and mRNA levels measured by Affymetrix GeneChip technology, leading to inconclusive results. Despite of this, comparison between PSAT1 IHC and RT-qPCR expression showed a significant association. This is probably due to the fact that, despite the small amount of samples overlapping between the TMA and the RT-qPCR sets, more accurate mRNA quantitation can be achieved through RT-qPCR compared to microarray technology.

Having established that PSAT1 protein levels are indicative of fast tumor progression tumors treated with first line tamoxifen, we sought to investigate which molecular pathways were associated to *PSATI* expression in our Affymetrix dataset. In this set, *PSATI* was found enriched in poor outcome patients and its expression was also significantly associated to shorter TTP. Global test analysis showed that, apart from the glycine, serine and threonine metabolism pathway, the JAK-STAT signaling and the cytokine-cytokine receptor interaction pathways were positively associated to *PSATI* expression. Molecules of interest in the cytokine pathway comprised *IL4R* and *IL21R*, which were enriched in poor outcome patients and positively associated to *PSATI* expression. *IL4R* has been associated to metabolic reprogramming of tumor cells in favor of an increased glucose uptake and enhancement of glutamine metabolism [37], while *IL21R* expression has been previously linked to breast cancer cell proliferation and migration in MM231 breast cancer cell lines [38]. Furthermore, *IL21* binding to its receptor promotes the activation of several signaling cascades, such as the JAK-STAT pathway, with differential effects related to tissue of expression [39, 40]. As both the JAK-STAT and cytokine pathways are directly linked to inflammation and contribute to tumor growth and spread in several cancer types [41, 42], we aimed at investigating the relation between infiltrating tumor cells and PSAT1 expression. Our analyses resulted in a significant association of *PSATI* levels to a set of genes predictive of TIL infiltration in breast cancer [34, 35], indicating a possible involvement of immune cells in *PSATI* expression and tamoxifen therapy resistance.

Glucose and glutamine are the main molecules that cells metabolize to generate ATP through glycolysis and the tricarboxylic acid cycle (TCA). During anabolic states cancer cells reprogram their metabolic pathways, from oxidative phosphorylation to glycolysis and serine production [43, 44]. Out of the latter, serine can be further converted into Glycine and subsequently in purines, while intermediate products such as  $\alpha$ -ketoglutarate can be directed towards the TCA cycle [45]. PSAT1 has a key role in serine biosynthesis, by catalyzing the oxidation of 3-phosphohydroxypyruvate into phosphoserine using glutamate. Previous studies

have shown that the serine pathway has a key role in cancer metabolism, using glycolysis-derived glucose for serine production and tumor growth [44]. PSAT1 and its related molecules (e.g. its upstream enzyme PHGDH) have been shown to be overexpressed in ER negative tumors and cell lines, conferring a metabolic-related growth advantage through production of alpha-ketoglutarate and altering the cell redox status by redirecting 3-phosphoglycerate [21, 46]. In the perspective of tamoxifen resistance, and following our global test results, the metabolic switch of tumor cells toward serine production may be an indication of a vaster tumor cell and tumor microenvironment reprogramming, involving cytokine and JAK-STAT signaling. As JAK-STAT signaling has been shown not only to induce metabolic switches in tumor cells, but also facilitates cell cycle progression by promoting the activation of cyclin dependent kinases [42, 47, 48], it is possible that PSAT1 enrichment is indicative of such upstream signaling. Still, functional studies should be performed in order to establish the hierarchical link between metabolic reprogramming, JAK-STAT signaling, and resistance to tamoxifen therapy.

We here conclude that PSAT1 protein expression is associated to rapid progression of metastatic breast cancer treated with first line tamoxifen and that it is likely associated to general tumor cell metabolic reprogramming, likely through IL-JAK-STAT signaling.

## References:

1. Siegel RL, Miller KD, Jemal A: Cancer Statistics, 2015. *CA Cancer J Clin* 2015, 65:5–29.
2. Cardoso F, Harbeck N, Fallowfield L, Kyriakides S, Senkus E: Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2012, 23 Suppl 7(Supplement 7):vii11–9.
3. (EBCTCG) EBCTCG: Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005, 365:1687–717.
4. Creighton CJ, Massarweh S, Huang S, Tsimelzon A, Hilsenbeck SG, Osborne CK, Shou J, Malorni L, Schiff R: Development of resistance to targeted therapies transforms the clinically associated molecular profile subtype of breast tumor xenografts. *Cancer Res* 2008, 68:7493–501.
5. Ring A, Dowsett M: Mechanisms of tamoxifen resistance. *Endocr Relat Cancer* 2004, 11:643–58.
6. Zhao W, Zhang Q, Kang X, Jin S, Lou C: AIB1 is required for the acquisition of epithelial growth factor receptor-mediated tamoxifen resistance in breast cancer cells. *Biochem Biophys Res Commun* 2009, 380:699–704.
7. Toy W, Shen Y, Won H, Green B, Sakr R a, Will M, Li Z, Gala K, Fanning S, King T a, Hudis C, Chen D, Taran T, Hortobagyi G, Greene G, Berger M, Baselga J, Chandarlapaty S: ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* 2013, 45:1439–45.
8. Huang B, Warner M, Gustafsson J-Å: Estrogen receptors in breast carcinogenesis and endocrine therapy. *Mol Cell Endocrinol* 2014, 418:240–4.
9. Zhang Y, Moerkens M, Ramaiahgari S, de Bont H, Price L, Meerman J, van de Water B: Elevated insulin-like growth factor 1 receptor signaling induces antiestrogen resistance through the MAPK/ERK and PI3K/Akt signaling routes. *Breast Cancer Res* 2011, 13:R52.
10. Mitri Z, Constantine T, O'Regan R: The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy. *Chemother Res Pract* 2012, 2012:743193.
11. Burstein HJ: The distinctive nature of HER2-positive breast cancers. *N Engl J Med* 2005, 353:1652–4.
12. Madlensky L, Natarajan L, Tchu S, Pu M, Mortimer J, Flatt SW, Nikoloff DM, Hillman G, Fontecha MR, Lawrence HJ, Parker B a, Wu a HB, Pierce JP: Tamoxifen metabolite concentrations, CYP2D6 genotype, and breast cancer outcomes. *Clin Pharmacol Ther* 2011, 89:718–25.
13. Polyak K: Tumor heterogeneity confounds and illuminates: a case for Darwinian tumor evolution. *Nat Med* 2014, 20:344–6.
14. Robinson DR, Wu Y-M, Vats P, Su F, Lonigro RJ, Cao X, Kalyana-Sundaram S, Wang R, Ning Y, Hodges L, Gursky A, Siddiqui J, Tomlins S a, Roychowdhury S, Pienta KJ, Kim SY, Roberts JS, Rae JM, Van Poznak

CH, Hayes DF, Chugh R, Kunju LP, Talpaz M, Schott AF, Chinnaiyan AM: Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* 2013, 45:1446–51.

15. Widschwendter M, Jones P a: DNA methylation and breast carcinogenesis. *Oncogene* 2002, 21:5462–82.

16. Rodríguez-Rodero S, Delgado-Álvarez E, Fernández AF, Fernández-Morera JL, Menéndez-Torre E, Fraga MF: Epigenetic alterations in endocrine-related cancer. *Endocr Relat Cancer* 2014, 21:R319–30.

17. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky E a, Gabrielson E, Schutte M, Baylin SB, Herman JG: Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst* 2000, 92:564–9.

18. Graff JR, Herman JG, Lapidus RG, Graft JR, Chopra H, Xu R, Jarrard DF, Isaacs WB, Pitha PM, Davidson NK, Baylin SB: E-Cadherin Expression Is Silenced by DNA Hypermethylation in Human Breast and Prostate Carcinomas *Advances in Brief E-Cadherin Expression Is Silenced by DNA Hypermethylation in Human Breast and Prostate Carcinomas*. *Cancer Res* 1995:5195–99.

19. Lombaerts M, van Wezel T, Philippo K, Dierssen JWF, Zimmerman RME, Oosting J, van Eijk R, Eilers PH, van de Water B, Cornelisse CJ, Cleton-Jansen A-M: E-cadherin transcriptional downregulation by promoter methylation but not mutation is related to epithelial-to-mesenchymal transition in breast cancer cell lines. *Br J Cancer* 2006, 94:661–71.

20. Martens JWM, Nimmrich I, Koenig T, Look MP, Harbeck N, Model F, Kluth A, Bolt-De Vries J, Sieuwerts AM, Portengen H, Meijer-Van Gelder ME, Piepenbrock C, Olek A, Höfler H, Kiechle M, Klijn JGM, Schmitt M, Maier S, Foekens JA: Association of DNA methylation of phosphoserine aminotransferase with response to endocrine therapy in patients with recurrent breast cancer. *Cancer Res* 2005, 65:4101–7.

21. Locasale JW, Grassian AR, Melman T, Lyssiotis C a, Mattaini KR, Bass AJ, Heffron G, Metallo CM, Muranen T, Sharfi H, Sasaki AT, Anastasiou D, Mullarky E, Vokes NI, Sasaki M, Beroukhim R, Stephanopoulos G, Ligon AH, Meyerson M, Richardson AL, Chin L, Wagner G, Asara JM, Brugge JS, Cantley LC, Vander Heiden MG: Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nat Genet* 2011, 43:869–74.

22. Reijm EA, Timmermans AM, Look MP, Meijer-van Gelder ME, Stobbe CK, van Deurzen CHM, Martens JWM, Sleijfer S, Foekens J a., Berns PMJJ, Jansen MPH: High protein expression of EZH2 is related to unfavorable outcome to tamoxifen in metastatic breast cancer. *Ann Oncol* 2014, 25:2185–90.

23. Elston C, Ellis I: Pathologic prognostic factors in breast cancer. I. The value of histological grades in breast cancer. Experience from a large study with long-term follow-up. *Histopathology* 1991:403–10.

24. Hayward J, Carbone P: Assessment of response to therapy in advanced breast cancer. *Br J Cancer* 1977(April):1978.

25. Altman DG, McShane LM, Sauerbrei W, Taube SE: Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med* 2012, 9:e1001216.

26. De Marchi T, Liu NQ, Stingl C, Timmermans MA, Smid M, Look MP, Tjoa M, Braakman RBH, Opdam M, Linn SC, Sweep FCGJ, Span PN, Kliffen M, Luider TM, Foekens JA, Martens JWM, Umar A: 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Mol Oncol* 2016, 10:24–39.
27. De Marchi T, Timmermans AM, Smid M, Look MP, Stingl C, Opdam M, Linn SC, Sweep FCGJ, Span PN, Kliffen M, Deurzen CHM Van, Luider TM, Foekens JA, Martens W, Umar A: Annexin-A1 and caldesmon are associated with resistance to tamoxifen in estrogen receptor positive recurrent breast cancer. *Oncotarget* 2015, 7:3098–110.
28. Baek JY, Jun DY, Taub D, Kim YH: Characterization of human phosphoserine aminotransferase involved in the phosphorylated pathway of L-serine biosynthesis. *Biochem J* 2003, 373(Pt 1):191–200.
29. Sieuwerts AM, Meijer-van Gelder ME, Timmermans M, Trapman AMAC, Garcia RR, Arnold M, Goedheer AJW, Portengen H, Klijn JGM, Foekens JA: How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. *Clin Cancer Res* 2005, 11:7311–21.
30. McCall MN, Bolstad BM, Irizarry R a.: Frozen robust multiarray analysis (fRMA). *Biostatistics* 2010, 11:242–53.
31. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M: KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2009, 38:355–60.
32. Goeman JJ, Oosting J, Cleton-Jansen A-M, Anninga JK, van Houwelingen HC: Testing association of a pathway with survival using gene expression data. *Bioinformatics* 2005, 21:1950–7.
33. Goeman JJ, Van de Geer S, De Kort F, van Houwelingen HC: A global test for groups of genes: Testing association with a clinical outcome. *Bioinformatics* 2004, 20:93–9.
34. Massink MPG, Kooi IE, van Mil SE, Jordanova ES, Ameziane N, Dorsman JC, van Beek DM, van der Voorn JP, Sie D, Ylstra B, van Deurzen CHM, Martens JW, Smid M, Sieuwerts AM, de Weerd V, Foekens JA, van den Ouweland AMW, van Dyk E, Nederlof PM, Waisfisz Q, Meijers-Heijboer H: Proper genomic profiling of (BRCA1-mutated) basal-like breast carcinomas requires prior removal of tumor infiltrating lymphocytes. *Mol Oncol* 2015, 9:1–12.
35. Massink MPG, Kooi IE, Martens JWM, Waisfisz Q, Meijers-Heijboer H: Genomic profiling of CHEK2\*1100delC-mutated breast carcinomas. *BMC Cancer* 2015, 15:877.
36. Droog M, Beelen K, Linn S, Zwart W: Tamoxifen resistance: From bench to bedside. *Eur J Pharmacol* 2013, 717:47–57.
37. Venmar KT, Kimmel DW, Cliffl DE, Fingleton B: IL4 receptor  $\alpha$  mediates enhanced glucose and glutamine metabolism to support breast cancer growth. *Biochim Biophys Acta* 2015, 1853:1219–28.

38. Wang L-N, Cui Y-X, Ruge F, Jiang WG: Interleukin 21 and Its Receptor Play a Role in Proliferation , Migration and Invasion of Breast Cancer Cells. *Cancer Genomics Proteomics* 2015, 12:211–21.
39. Wood B, Sikdar S, Choi SJ, Virk S, Alhejaily A, Baetz T, LeBrun DP: Abundant expression of interleukin-21 receptor in follicular lymphoma cells is associated with more aggressive disease. *Leuk Lymphoma* 2013, 54:1212–20.
40. Kesselring R, Jauch D, Fichtner-feigl S: Interleukin 21 impairs tumor immunosurveillance of colitis-associated colorectal cancer. *Oncoimmunology* 2012(July):537–8.
41. Yu H, Pardoll D, Jove R: STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat Rev Cancer* 2009, 9:798–809.
42. Thomas SJ, Snowden JA, Zeidler MP, Danson SJ: The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours. *Br J Cancer* 2015, 113:365–71.
43. Schulze A, Harris AL: How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature* 2012, 491:364–73.
44. Luo J: Cancer's sweet tooth for serine. *Breast Cancer Res* 2011, 13:317.
45. Amelio I, Cutruzzolá F, Antonov A, Agostini M, Melino G: Serine and glycine metabolism in cancer. *Trends Biochem Sci* 2014, 39:191–8.
46. Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo H-K, Jang HG, Jha AK, Chen WW, Barrett FG, Stransky N, Tsun Z-Y, Cowley GS, Barretina J, Kalaany NY, Hsu PP, Ottina K, Chan AM, Yuan B, Garraway L a, Root DE, Mino-Kenudson M, Brachtel EF, Driggers EM, Sabatini DM: Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* 2011, 476:346–50.
47. Demaria M, Giorgi C, Lebedzinska M, Esposito G, D'Angeli L, Bartoli A, Gough DJ, Turkson J, Levy DE, Watson CJ, Wieckowski MR, Provero P, Pinton P, Poli V: A STAT3-mediated metabolic switch is involved in tumour transformation and STAT3 addiction. *Aging (Albany NY)* 2010, 2:823–42.
48. Jäkel H, Weinl C, Hengst L: Phosphorylation of p27Kip1 by JAK2 directly links cytokine receptor signaling to cell cycle control. *Oncogene* 2011, 30:3502–12.

## **Chapter 9**

### **Discussion**

## 9.0 - Discussion

Three quarters of all breast cancer cases display estrogen receptor (ER) positivity. Endocrine therapies, such as tamoxifen, constitute primary treatment for these tumors. In the recurrent setting, nearly half of the patients that receive tamoxifen do not respond to the treatment, while the other half develops resistance over time [1,2]. Several studies elucidated many molecular mechanisms underlying tamoxifen therapy resistance in both the adjuvant and recurrent settings, such as the expression of estrogen receptor variants or mutated forms (e.g. Tyr537Ser) as well as the activation of estrogen-independent growth pathways (e.g. PI3K) [3,4]. In addition to this, several tamoxifen therapy outcome predictive gene panels have been derived from large cohort studies, however none of the markers has been introduced into a clinical diagnostic setting [5].

Through constant technological improvements, high resolution mass spectrometry (MS) has been established as a sensitive and robust platform for accurately quantifying protein abundances in biological samples, achieving coverage of almost the entirety of the human proteome [6,7]. Therefore, MS-based approaches constitute a suitable platform for biomarker discovery studies [8]. In addition, due to its compatibility with multiple enrichment methods (e.g. immunoaffinity enrichment), MS finds application in targeted quantitative analyses for both biomarker verification and functional studies (i.e. interaction proteomics) [9,10].

Still, analysis of heterogeneous biological material (e.g. breast tissue) suffers from interfering proteomes of cellular sub-populations, which can impair accurate quantification. Target cell subpopulation enrichment strategies, such as laser capture microdissection (LCM), allow bypassing this issue and assessment of molecular changes specific to (e.g.) epithelial tumor cells [11,12]. In other instances, where a previously selected analyte is measured in a complex biological matrix (e.g. tissue, plasma, serum), enrichment through antibodies efficiently minimizes interference from other compounds (e.g. co-eluting peptides) [9,13,14].

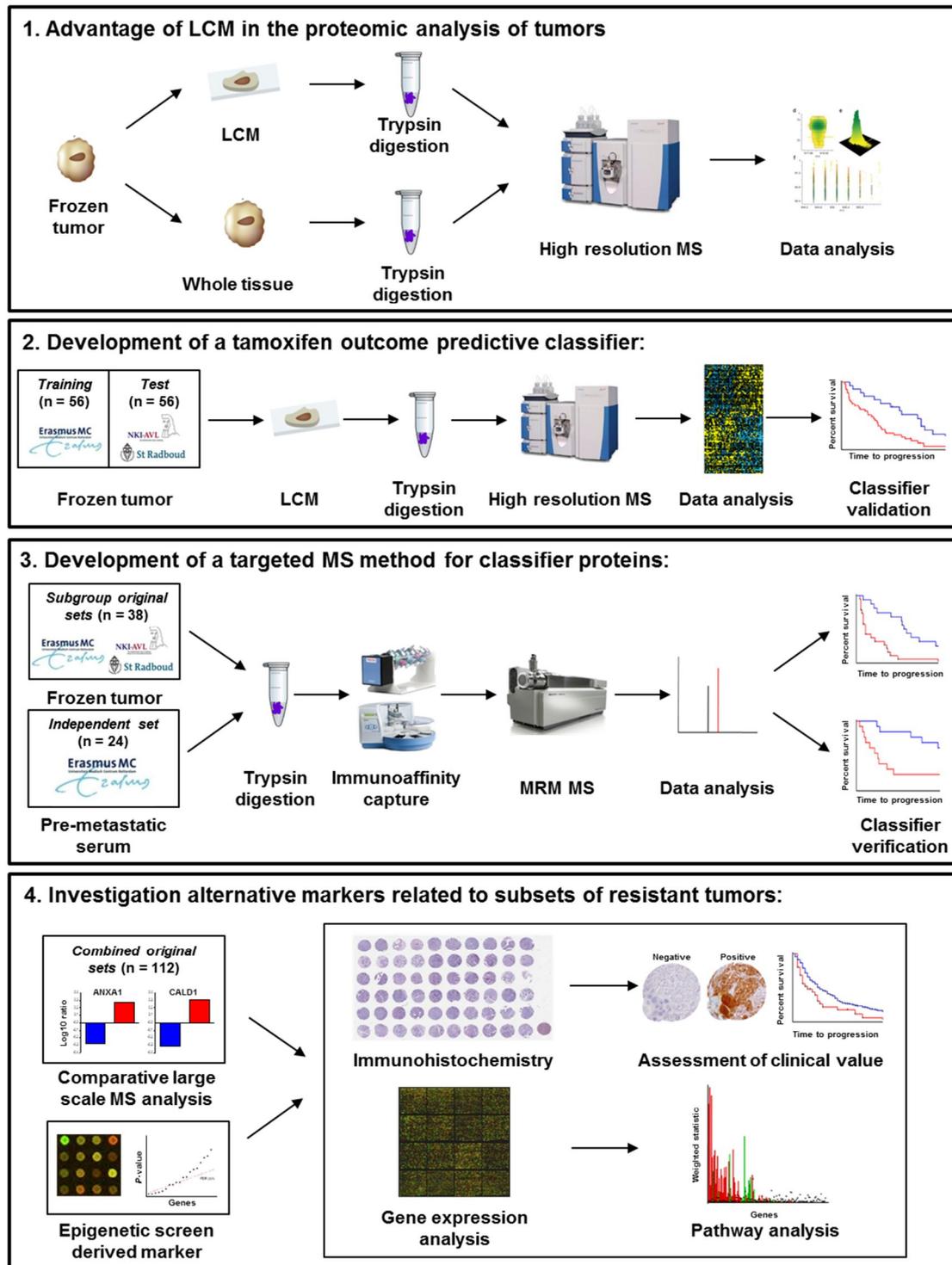
The aims addressed in this thesis are (Figure 9.1):

- Evaluation of the advantage of cell enrichment techniques in proteomics based biomarker discovery.
- Development and validation of a protein based signature for tamoxifen resistance.
- Design and development of a targeted MS based assay to quantify signature proteins.

- Analysis of pathways and potential biomarkers present in subgroups of ER positive breast cancers.

### *9.1 - High resolution mass spectrometry in the analysis of tumors: the advantage posed by enrichment strategies*

High resolution MS, also termed global proteomics, constitutes a powerful approach to assess the levels of thousands of proteins in a single biological sample (15). For that reason, such approaches constitute a robust technology in biomarker discovery studies, where the more proteins there are identified, the more the cellular pathway dynamics in a complex disease can be assessed. However, when delving into the proteome of complex specimens (e.g. tissues, blood), signals from the cell population of interest may become diluted by the so called “background proteome”, which encompasses proteins and their intensities derived from other regions of the specimen (tissue) or from entirely different cell types (e.g. leukocytes, fibroblasts or even blood vessels) (16). Therefore, while on one hand identification of low abundance proteins may be compromised due to interference of high abundant ones (e.g. albumin, hemoglobin, collagens) at the chromatography and/or MS levels, on the other hand protein quantification in a specific cell type may be impaired due to the differential expression in surrounding cell subpopulations (17). To overcome this, cell separation techniques (e.g. fluorescence assisted cell sorting [15], LCM [8]) or high abundant protein depletion (e.g. depletion columns [16]) techniques have been used to enrich for cell populations of interest followed by downstream proteomic analysis [17–19].



**Figure 9.1 Schematic representation of the experimental procedures used to address the aims of this thesis.**

Panel 1 illustrates the experimental workflow of our global proteomic comparison of microdissected tissues with their non-enriched counterparts. Panels 2 and 3 respectively display the workflow underlying the generation of a protein signature predictive of tamoxifen resistance and the development of a targeted MS method. Panel 4 recapitulates our investigation of tamoxifen resistance alternative markers that are expressed in subgroups of tumors.

Due to the fact that the use of cell enrichment techniques is labor intensive, many study designs often lack this part of the sample preparation. To date, no study examined the differences between proteomic analyses of LCM enriched and whole tissues. We evaluated whether differences exist between proteomic analyses of whole tissue lysates (WTL) compared with LCM-enriched breast carcinoma cells. We assessed the differences at the MS, peptide and protein levels between WTL and LCM material derived from 38 ER positive breast cancer tissues. Not only did the LCM set show a higher number of identified mass spectra, peptides and proteins, but also showed less redundancy at the chromatographic level. Furthermore, a significantly lower number of missing data points was observed both at the peptide and protein level. When looking at the possible contamination from proteins derived from non-tumor cells, the WTL set displayed higher abundance of blood-derived (e.g. haptoglobin) and extracellular matrix (e.g. collagen) proteins compared with LCM samples. Furthermore, the high number of missing values in the WTL set impaired not only the detection, but also the assessment of abundance differences of key breast cancer proteins, such as ER and HER2.

Overall, our findings suggested that proteomic analysis of whole tissues suffers from higher redundancy at the chromatographic and MS levels, leading to decreased protein identifications and an increase in missing data points (Table 9.1). While redundancy could be reduced by extensive sample fractionation prior to LC-MS analysis [20,21], and blood-derived proteins could be depleted [22], signal dilution derived from differential expression of proteins expressed in both the tumor and surrounding tissues cannot be avoided and, upon high resolution MS analysis, would ultimately affect accurate quantification.

Furthermore, as the tumor microenvironment plays an important role in many cancer processes such as tumor growth, spread and therapy resistance [23], future functional and biomarker studies should take such element into account. Therefore, global profiling studies aimed at defining the molecular characteristics of solid tumors and their microenvironment, especially through proteomic technologies, could employ cell enrichment techniques such as LCM to segregate different cellular types, or fluorescence activated cell sorting (FACS) to assess molecular changes between (tumor) cell subpopulations.

Taken together, the implementation of cell enrichment techniques in genomic, transcriptomic, and proteomic studies would constitute a significant asset to determine the molecular interactions between tumor cells, leukocytes and fibroblasts, define the molecular

characteristic of tumor subclones, and possibly identify predictive and prognostic biomarkers for clinical application.

**Table 9.1. Technical features of laser capture microdissected and whole tissue specimens in proteomics.**

	<b>Laser capture microdissection</b>	<b>Whole tissue specimens</b>
<b>Technical advantages</b>	specific quantitation of the tumor proteome no or little tumor cellularity restriction	can be coupled to chemical labeling high protein input less time consuming sample preparation
<b>Technical disadvantages</b>	small amount of protein derived per sample (low protein input) difficulty to use chemical labeling time consuming sample preparation	background proteome contamination only high cellularity tumors
<b>Comparative analysis</b>	high amount of identified/quantified proteins low identification redundancy low percentage of missing data minimal interference from background proteome	low amount of identified/quantified proteins high identification redundancy high percentage of missing data higher interference from background proteome

### *9.2 - Protein signature predictive of tamoxifen resistance*

Having established that LCM-based enrichment prior to downstream proteomic analysis enables better quantification of the tumor proteome over whole tissue analysis, we used our tissue proteomic workflow to develop and validate a predictive signature of tamoxifen resistance [8,24]. From global proteomic analysis of 112 (i.e. combined training and test sets) ER positive LCM-enriched tumors we derived a 4 protein signature capable of significantly predicting tamoxifen therapy outcome in the recurrent setting.

Our predictive signature presents notable characteristics: first, we included tumors from different medical centers, which likely took into account differences in frozen tumor specimen collection and storage. Second, the inclusion of tumors derived from patients who did not receive any adjuvant endocrine treatment contributed to a relatively unbiased analysis of differences in protein abundances related to the endpoint of interest: tamoxifen outcome in the recurrent setting. Third, the development of a biomarker signature consisting of only 4 protein markers significantly reduces costs and optimization efforts in the biomarker verification phase. On a technical level, the use of LCM in the investigation of differential protein abundance enabled us to dissect the breast tumor proteome with minimal interference from surrounding tissues, such as fibroblast and immune cells, as also mentioned in paragraph 9.1.

In addition to the predictive value of the 4-protein signature, these markers may also pinpoint alternative (intrinsic and de novo) mechanisms of tamoxifen resistance. The tumor suppressor PDCD4, a protein translation inhibitor, has been reported to be a target of the oncogenic long non-coding miRNA-21 [25], which regulates gene expression via 3'-UTR binding of target mRNAs. Furthermore, PDCD4 has been recently associated to aromatase inhibitor (AI) resistance in primary ER positive breast cancer [26]. In this cohort HER2 expression was associated to PDCD4 downregulation via activation of Akt and MAPK, which would phosphorylate the PDCD4 causing its inactivation and subsequent degradation. However, while several studies have pointed out HER2 overexpression as a major mechanism of endocrine therapy resistance [27], which enables tumor growth independently of anti-estrogen therapy through cross-talk mechanisms, none of the tumors analyzed in our study displayed HER2 amplification. Still, our findings pointed out that tamoxifen resistant tumors harbored dysregulation of cell cycle related pathways in which, on top of PDCD4, also CGN plays an important part. CGN has been previously reported as an epithelial differentiation marker in cancer, its function related to cell cycle regulation via interaction with GATA4, which leads to the downregulation of RhoA [28]. On the other hand, OCIAD1 and G3BP2 function in relation to endocrine therapy resistance remains unclear. OCIAD1 has been reported to be overexpressed in ovarian carcinoma and is a marker of poor prognosis [29]. OCIAD1 has also been reported to bind and interact with STAT3, an oncogenic molecule that is involved in both inflammation and promotion of pro-oncogenic signals [30], although it remains unknown whether OCIAD1 promotes its activation or its inhibition. G3BP2, a molecule involved in stress granule formation [31], has been recently reported to be involved in matrix stiffness and the acquisition of an EMT-like phenotype following its downregulation in breast cancer cell lines [32]. In the light of this, G3BP2 functional role in breast cancer and, moreover, in endocrine therapy resistance needs to be further investigated.

Taken together, these findings suggest a role of the 4 proteins in cell cycle progression dysregulation, survival signals, and EMT-like features in tamoxifen resistant breast cancers. Still, the molecular hierarchy and interplay of these pathways need to be further elucidated through functional assays in pre-clinical models.

High resolution MS approaches are capable of measuring thousands of protein abundances with high sensitivity and accuracy. However, the extensive sample preparation time, limited

throughput and limited multiplexability hinder their introduction into clinical diagnostics. In this perspective, we employed IHC as an alternative technique, which confirmed the significant association of PDCD4 protein to tamoxifen outcome in an independent cohort of formalin-fixed and paraffin embedded tissues [33]. Despite of this, validation of the 3 remaining markers (i.e. CGN, OCIAD1 and G3BP2) was unsuccessful, likely due to the limited dynamic range measurable by the technique as well as the intrinsic lack of linearity in quantifying chromophore signals.

### *9.3 - Development of a targeted MS method for quantification of the 4-protein signature.*

Following signature development, targeted quantitative assays should be generated in order to verify the clinical utility of biomarkers. Therefore, as our tissue proteomics workflow requires too extensive sample preparation time and is not multiplexable, and due to the fact that we were only able to validate one marker through IHC, we chose a targeted and more quantitative approach.

In the perspective of a clinically applicable assay, protein biomarkers can be readily measured by immuno-based assays (e.g. IHC, ELISA). Still, such techniques often do not provide accurate quantification (IHC) or require extensive optimization (ELISA). Targeted proteomic technologies, however, such as multiple reaction monitoring (MRM), have shown a high degree of reproducibility and measurement precision as well as require less laborious optimization compared with traditional immuno-assays [34,35]. Furthermore, the ability to quantify target proteins (also harboring post-translational modifications) out of several types of samples (e.g. tissue, blood, urine) broadens the applicability of targeted MS methods in the diagnostic field. MRM MS coupled to stable isotope dilution provides a highly accurate method to measure analytes in a plethora of sample types (e.g. tissue, secretions, blood) [36]. Yet, the presence of highly abundant proteins (e.g. serum albumin) may impair quantification due to interference at the chromatography level (Figure 9.2) [37,38]. In order to minimize such interference, we combined MRM to immunoaffinity captures by anti-peptide antibodies (defined as immuno-MRM [iMRM]).

<i>Phase</i>	<i>Approach</i>	<i>Characteristics</i>
<b>Discovery</b>	<b>Laser capture microdissection + High resolution MS</b>	<ul style="list-style-type: none"> <li>- sensitive identification</li> <li>- accurate quantitation</li> <li>- very low throughput</li> <li>- limited multiplex</li> <li>- extensive sample preparation</li> <li>- extensive measurement time</li> </ul>
<b>Verification</b>	<b>Immunohistochemistry</b>	<ul style="list-style-type: none"> <li>- sensitive (depending on antibody)</li> <li>- very low throughput</li> <li>- inaccurate quantitation</li> <li>- requires moderate optimization</li> <li>- very limited multiplexing</li> </ul>
	<b>Targeted iMRM MS method</b>	<ul style="list-style-type: none"> <li>- sensitive (depending on antibody)</li> <li>- moderate throughput</li> <li>- high multiplex</li> <li>- requires moderate optimization</li> <li>- accurate quantitation</li> </ul>

**Figure 9.2. Features of proteomic and immuno-based approaches in our proteomic biomarker studies.**

Our study showed that iMRM hardly suffered from sample matrix interference compared with direct (unfractionated) MRM. Still, while on the one hand immunoaffinity capture provides high sensitivity measurement of proteotypic peptides, on the other hand it necessitates relatively high amounts of input material compared to high resolution MS, rendering LCM unfeasible for such an approach.

Upon measurement of the 4 proteins by iMRM in breast cancer tissues, we were able to establish their differential abundance and, when compared to our high resolution MS data, similar sensitivity and specificity in patient stratification was achieved. This is probably due to the fact that the 4-protein markers are mainly expressed in breast epithelial cells, as observed previously in our IHC data, which limited the interference from any background proteome.

Despite of this, analyzing frozen breast cancer primary tumor tissues can be difficult due to the fact that they are not collected on a routine basis in the clinic and the possible biases introduced by different methods of sample storage. Therefore, analysis of blood-derived samples, such as plasma or serum, which can be more easily retrieved due to the non-invasive nature of its collection methodology, which renders large number of samples to be available for diagnostic purposes. Upon testing our iMRM assay on a pilot cohort of sera collected prior to start of first line tamoxifen treatment, not only were we able to detect the 4 signature proteins, but even more importantly significant patient stratification was also achieved.

In conclusion, our iMRM assay constitutes a reliable alternative method for measuring the 4 signature proteins in both patient-derived tumor tissues and sera. However, to further introduce such findings into the clinical setting and on top of clinical verification of our biomarkers in a large cohort of patients, technical issues need to be addressed, such as sensitivity and accuracy. First, antibody specificity and analyte recovery (linked to both specificity and avidity of each antibody) can be improved by selecting monoclonal antibodies for target immunocapture. Second, improvement in the resolution of MS instruments may offer a solution for the identification and quantification of low abundance species. Third, reference guidelines for peptide standard selection, protein digestion and MS measurement standardization should be followed to ensure inter-laboratory reproducibility and to ultimately translate MRM assays into the clinical laboratory [39].

#### *9.4 - Several mechanisms underlie endocrine therapy resistance.*

As mentioned in the previous paragraphs, we have shown that our 4-protein signature significantly predicts tamoxifen therapy outcome in recurrent breast cancer patients and we have developed a targeted assay to measure our signature proteins in both tumor tissues and pre-metastatic sera. Despite of this, it is unlikely that the 4 proteins can fully explain the biology behind tamoxifen resistance [41,42]. In this perspective, we assessed whether additional markers could better define the biological pathways related to endocrine resistance in our tissue samples, which led to the identification of ANXA1 and CALD1 as potential predictive markers. Upon evaluating the clinical relevance of both proteins by IHC, we discovered that only a minority of tumors expressed ANXA1 and CALD1, possibly identifying a specific tumor subtype that relates to fast tumor progression. Furthermore, most of the proteins that were associated to ANXA1 expression were also associated to CALD1, suggesting a common role for these two proteins in tamoxifen resistance. These proteins were mainly associated with downregulation of ER, cell adhesion and cell movement which confirmed previous findings of CALD1 and ANXA1 upregulation in breast cancer cell lines acquiring an EMT-like phenotype [43,44], which resembled features of basal tumors. As ANXA1 and CALD1 have been linked to regulation of basal ER negative cancer cell invasion [45–47], we can speculate whether these markers may identify ER positive tumors with basal characteristics. As basal tumors are generally characterized by poor prognosis and insensitivity to endocrine therapies due to the fact that they do not need estrogen signaling in

order to grow compared with their luminal counterparts, it is conceivable that features similar to basal tumors in ER positive cancers may render tamoxifen therapy inefficacious. Furthermore, as tumor heterogeneity and evolution studies have pointed out (e.g. [49–51]), breast tumors acquire several features (e.g. EMT characteristics, mutations) both at the primary and metastatic level. Such characteristics confer the tumor, or some of its sub-clones, the ability to escape the organ of origin and to withstand anti-cancer therapies. Therefore, it is also conceivable that ANXA1 and CALD1 may pinpoint not general tumor characteristics, but sub-clonal ones. However, in order to confirm this, studies assessing protein expression in different tumor clones need to be performed.

Taken together, our findings suggest that ANXA1 and CALD1 are not only indicative of faster tumor progression but, given the fact that tumor heterogeneity plays an important role in both the biology and the clinical course of breast cancer [42], they also possibly identify specific tumor subgroups. However, further clinical and functional validation of these findings needs to be carried out in future studies.

#### *9.5 - From epigenetic screens to gene expression - metabolically dysregulated tumors.*

Differential abundance of genes and proteins is the reflection of the fine upstream regulation of transcription factors and histone proteins, which in turn is mediated by several types of modifications (e.g. methylation, acetylation) [52,53]. These mechanisms have been shown to modulate the expression of several genes involved in a plethora of cancer processes, such as tumorigenesis and invasion [54–56]. Involvement of epigenetic mechanisms in endocrine therapy resistance has also been shown by recent studies, such as the differential chromatin-binding events of histone proteins and transcription factors (e.g. ER, PgR), that enhanced or suppressed expression of estrogen responsive genes [57–60]. Modified histone proteins and ER differential binding events to chromatin not only elucidated that epigenetic regulation is underlying endocrine therapy resistance, but also represent clinical outcome predictive markers [60]. A previous study from our laboratory on differential DNA methylation showed that phosphoserine aminotransferase 1 (PSAT1) hypo-methylation, and conversely high mRNA expression, was associated to poor patient outcome in recurrent breast cancer patients who received tamoxifen as first-line therapy [61]. We sought to investigate whether these findings would translate at the protein level and which genes were associated to high expression of PSAT1. Overall, our analysis showed that not only PSAT1 protein expression

was associated with poor outcome to tamoxifen, but that it was expressed only in a small subset of ER positive tumors, as previously shown for ANXA1 and CALD1. Furthermore, high PSAT1 mRNA expressing tumors displayed upregulation of the cytokine signaling and Jak/STAT pathways, which suggested that serine biosynthesis activation was concomitant with cascade signaling dysregulation. Correlation analysis of PSAT1 mRNA levels with a tumor infiltrating lymphocytes (TIL) signature [62] showed that tumors expressing high levels of PSAT1 also showed high expression of TIL signature related genes. This observation pinpoints the fact that metabolic dysregulation in tumor cells may derive from extracellular signals (e.g. cytokines) of immune cell origin. Several studies pointed out how metabolic reprogramming of tumor cells promoted by microenvironmental stimuli can affect the clinical course as well as the biology of the disease [30,63,64], promoting therapy resistance, tumor growth and spread. Still, it is unknown whether these stimuli originate from cancer cell-induced inflammation at the tumor site or are part of a wider signaling network. Therefore, and in order to determine the hierarchy of such signaling pathways during the clinical course of the disease, functional studies in preclinical models (e.g. co-culture experiments with activated lymphocytes) need to be performed. Such studies would not only clarify the dynamic signaling evolution of cancer and its interplay with its microenvironment, but may also define alternative therapeutic targets.

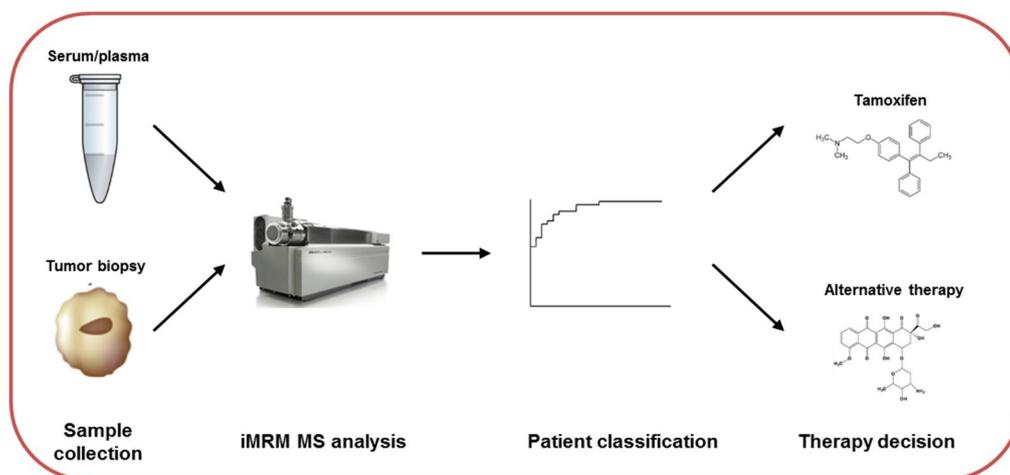
### *9.7 - Overall conclusions*

We have here described how proteomic technologies constitute a reliable platform for both biomarker discovery and verification studies in endocrine therapy treated breast cancers. In this thesis, we have first shown how LCM-based cell enrichment constitutes a better approach over whole tissue specimens to investigate the tumor proteome by high resolution MS (**Chapter 3**). In fact, a higher amount of proteins were identified/quantified in LCM samples, with minimized interference from the background proteome and highly abundant (e.g. plasma) proteins not belonging to tumor carcinoma cells. In this perspective, our findings point out that tissue enrichment strategies should be employed to specifically analyze specific cell types and subpopulations in biomarker discovery studies, though such technologies could be used as well in functional studies.

We next employed our tissue proteomic approach to develop and validate a 4-protein signature predictive of tamoxifen resistance in ER positive breast cancers (**Chapter 4**). This

protein signature predicts patients that would not benefit from tamoxifen treatment, and an alternative therapy (e.g. chemotherapy) should preferably be administered (Figure 9.3). Although tamoxifen does not constitute standard first line treatment for ER positive recurrent breast cancer, since it has been gradually replaced by AIs, our protein signature may still be predictive of endocrine treatment resistance in general. As a matter of fact, estrogen signaling modulating drugs ultimately share a common denominator, as it has been shown for previously reported epigenetic signatures [60]. In this perspective, it would be worthwhile to assess also predictive value of the 4-protein signature in a cohort of patients treated with AIs.

Following up our signature development, we aimed at validating our findings using a more clinically feasible technique. While IHC analysis verified the significant association of only one of our signature proteins (i.e. PDCD4), our targeted iMRM MS approach was able to significantly predict patient outcomes in both tissues and sera (**Chapter 5** and **Chapter 6**) by measuring all of the signature markers. This establishes our 4-protein signature as a potential diagnostic tool that would help clinicians direct the pharmacological treatment of recurrent ER positive recurrent breast cancer patients (Figure 9.3).

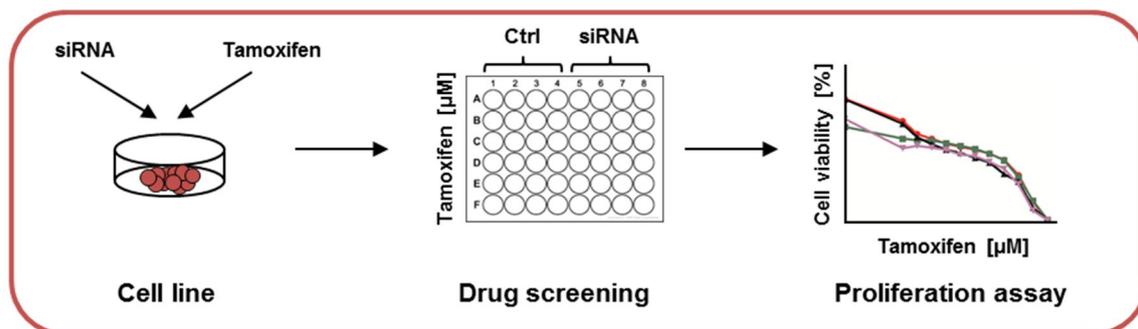


**Figure 9.3. Potential impact of the 4-protein predictive signature in a clinical setting.**

While our signature might be useful for the general identification of tamoxifen resistant tumors, other markers may be indicative of particular subgroups in which specific molecular pathways are enriched, such as the NFκB and Jak/STAT cascades (**Chapter 7** and **Chapter 8**, respectively). Furthermore, not only may these markers define subgroups within ER positive

tumors, they could also be the starting point for the development of alternative therapeutic approaches. Therefore, blocking antibodies against ANXA1 [65] or CALD1 may constitute a complementary or alternative therapeutic strategy in ER positive tumors with basal characteristics, while molecules targeting tumor-specific metabolic pathways (e.g. serine and glutamine biosynthesis) or the tumor microenvironment (e.g. TILs) may constitute an alternative therapeutic strategy in PSAT1 enriched tumors.

However, complete functional elucidation of these markers and their associated genes/proteins needs to be performed to assess the biological role of these markers in breast cancer therapy resistance. Targeted loss-of-function screens through siRNA transfection or CRISPRi [66,67] coupled to tamoxifen treatment in cell lines would be the first step to assess the biological effect of our biomarkers (Figure 9.4). In addition to this, gain of function assays through stable transfection of target genes (i.e. over-expression) would then be needed to confirm knock-down screen findings as a complementary experiment. Proliferation assays would subsequently be used to measure the effect of gene loss or gain in the presence of tamoxifen and to determine whether response to the drug is significantly impacted.



**Figure 9.4. Possible workflow for functional studies of tamoxifen resistance-associated markers.**

As the tumor microenvironment is known to play an important role in several cancer processes, such as tumorigenesis, invasion and metastasis, and drug resistance [68,69], these functional assays should however subsequently be translated into preclinical models in order to evaluate the effect of tumor surrounding tissues in tamoxifen resistance.

In conclusion, we have here shown that proteomic technologies coupled to cell enrichment techniques are capable of accurately dissecting the tumor proteome, and that this workflow was successfully used to develop and validate a tamoxifen resistance predictive protein signature. Furthermore, we have here shown that a targeted MS approach coupled to

immunoaffinity enrichment is able to accurately assess tumor protein concentrations to predict patient outcome to therapy in both primary tumor tissues and patient derived sera (at start of tamoxifen treatment for recurrent disease). Lastly, alternative markers have shown to identify tamoxifen resistant tumor subgroups, for which alternative targeted drugs could be developed.

## References:

- [1] Creighton CJ, Massarweh S, Huang S, Tsimelzon A, Hilsenbeck SG, Osborne CK, et al. Development of resistance to targeted therapies transforms the clinically associated molecular profile subtype of breast tumor xenografts. *Cancer Res* 2008;68:7493–501. doi:10.1158/0008-5472.CAN-08-1404.
- [2] Ring A, Dowsett M. Mechanisms of tamoxifen resistance. *Endocr Relat Cancer* 2004;11:643–58. doi:10.1677/erc.1.00776.
- [3] Huang B, Warner M, Gustafsson J-Å. Estrogen receptors in breast carcinogenesis and endocrine therapy. *Mol Cell Endocrinol* 2014;418:240–4. doi:10.1016/j.mce.2014.11.015.
- [4] Zhang Y, Moerkens M, Ramaiahgari S, de Bont H, Price L, Meerman J, et al. Elevated insulin-like growth factor 1 receptor signaling induces antiestrogen resistance through the MAPK/ERK and PI3K/Akt signaling routes. *Breast Cancer Res* 2011;13:R52. doi:10.1186/bcr2883.
- [5] Droog M, Beelen K, Linn S, Zwart W. Tamoxifen resistance: From bench to bedside. *Eur J Pharmacol* 2013;717:47–57. doi:10.1016/j.ejphar.2012.11.071.
- [6] Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, et al. Mass-spectrometry-based draft of the human proteome. *Nature* 2014;509:582–7. doi:10.1038/nature13319.
- [7] Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. *Nature* 2014;509:575–81. doi:10.1038/nature13302.
- [8] Liu NQ, Braakman RBH, Stingl C, Luider TM, Martens JWM, Foekens JA, et al. Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J Mammary Gland Biol Neoplasia* 2012;17:155–64. doi:10.1007/s10911-012-9252-6.
- [9] Anderson N, Anderson N. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J Proteome Res* 2004;3:235–44.
- [10] Schiess R, Wollscheid B, Aebersold R. Targeted proteomic strategy for clinical biomarker discovery. *Mol Oncol* 2009;3:33–44. doi:10.1016/j.molonc.2008.12.001.
- [11] Jr. DJJ, Rodriguez-Canales J, Mukherjee S, Prieto DA, C.Hanson J, Emmert-Buck M, et al. Approaching solid tumor heterogeneity on a cellular basis by tissue proteomics using laser capture microdissection and biological mass spectrometry. *J Proteome Res* 2010;8:2310–8. doi:10.1021/pr8009403.Approaching.
- [12] Yang F, Foekens JA, Yu J, Sieuwerts a M, Timmermans M, Klijn JGM, et al. Laser microdissection and microarray analysis of breast tumors reveal ER-alpha related genes and pathways. *Oncogene* 2006;25:1413–9. doi:10.1038/sj.onc.1209165.
- [13] Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* 2006;5:573–88. doi:10.1074/mcp.M500331-MCP200.

- [14] Whiteaker JR, Zhao L, Lin C, Yan P, Wang P, Paulovich a. G. Sequential Multiplexed Analyte Quantification Using Peptide Immunoaffinity Enrichment Coupled to Mass Spectrometry. *Mol Cell Proteomics* 2012;11:M111.015347–M111.015347. doi:10.1074/mcp.M111.015347.
- [15] Qian Y, Wei C, Lee FEH, Campbell J, Halliley J, Lee J a., et al. Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data. *Cytom Part B - Clin Cytom* 2010;78:69–82. doi:10.1002/cyto.b.20554.
- [16] Mrozinski P, Zolotarjova N, Chen H. Application Note: Human Serum and Plasma Protein Depletion—Novel High-Capacity Affinity Column for the Removal of the “Top 14” Abundant Proteins. *Agil Technol* 2008:1–6.
- [17] Neubauer H, Clare SE, Kurek R, Fehm T, Wallwiener D, Sotlar K, et al. Breast cancer proteomics by laser capture microdissection, sample pooling, 54-cm IPG IEF, and differential iodine radioisotope detection. *Electrophoresis* 2006;27:1840–52. doi:10.1002/elps.200500739.
- [18] Xu BJ. Combining laser capture microdissection and proteomics: methodologies and clinical applications. *Proteomics Clin Appl* 2010;4:116–23. doi:10.1002/prca.200900138.
- [19] Emmert-buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser Capture Microdissection. *Science (80- )* 1996;274:8–11.
- [20] Puangpila C, Mayadunne E, Rassi Z El. Liquid-phase-based separation systems for depletion, Liquid phase based separation systems for depletion, prefractionation, and enrichment of proteins in biological fluids and matrices for in-depth proteomics analysis—An update covering the period 2011–201. *Electrophoresis* 2015;36:238–52. doi:10.1002/elps.201100431.
- [21] Camerini S, Mauri P. The role of protein and peptide separation before mass spectrometry analysis in clinical proteomics. *J Chromatogr A* 2015;1381:1–12. doi:10.1016/j.chroma.2014.12.035.
- [22] Mahboob S, Mohamedali A, Ahn SB, Schulz-Knappe P, Nice E, Baker MS. Is isolation of comprehensive human plasma peptidomes an achievable quest? *J Proteomics* 2015;127:300–9. doi:10.1016/j.jprot.2015.05.010.
- [23] Gandellini P, Andriani F, Merlino G, D’Aiuto F, Roz L, Callari M. Complexity in the tumour microenvironment: cancer associated fibroblast gene expression patterns identify both common and unique features of tumour-stroma crosstalk across cancer types. *Semin Cancer Biol* 2015. doi:10.1016/j.semcancer.2015.08.008.
- [24] Braakman RBH, Tilanus-Linthorst MMA, Liu NQ, Stingl C, Dekker LJM, Luiders TM, et al. Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. *J Proteomics* 2012;75:2844–54. doi:10.1016/j.jprot.2012.01.022.

- [25] Afonja O, Juste D, Das S, Matsuhashi S, Samuels HH, Asangani I a, et al. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* 2014;9:1–9. doi:10.1128/MCB.24.9.3894.
- [26] Chen Z, Yuan YC, Wang Y, Liu Z, Chan HJ, Chen S. Down-regulation of programmed cell death 4 (PDCD4) is associated with aromatase inhibitor resistance and a poor prognosis in estrogen receptor-positive breast cancer. *Breast Cancer Res Treat* 2015;152:29–39. doi:10.1007/s10549-015-3446-8.
- [27] Hynes NE, MacDonald G. ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol* 2009;21:177–84. doi:10.1016/j.ceb.2008.12.010.
- [28] Guillemot L, Spadaro D, Citi S. The junctional proteins cingulin and paracingulin modulate the expression of tight junction protein genes through GATA-4. *PLoS One* 2013;8:e55873. doi:10.1371/journal.pone.0055873.
- [29] Sengupta S, Michener CM, Escobar P, Belinson J, Ganapathi R. Ovarian cancer immuno-reactive antigen domain containing 1 (OCIAD1), a key player in ovarian cancer cell adhesion. *Gynecol Oncol* 2008;109:226–33. doi:10.1016/j.ygyno.2007.12.024.
- [30] Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat Rev Cancer* 2009;9:798–809. doi:10.1038/nrc2734.
- [31] Matsuki H, Takahashi M, Higuchi M, Makokha GN, Oie M, Fujii M. Both G3BP1 and G3BP2 contribute to stress granule formation. *Genes Cells* 2013;18:135–46. doi:10.1111/gtc.12023.
- [32] Wei SC, Fattet L, Tsai JH, Guo Y, Pai HP, Majeski HE, et al. Matrix stiffness drives Epithelial-Mesenchymal Transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway. *Nat Cell Biol* 2015;17:678–88. doi:10.1016/B978-0-12-420118-7.00008-1.Dopamine.
- [33] De Marchi T, Liu NQ, Stingl C, Timmermans MA, Smid M, Look MP, et al. 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Mol Oncol* 2016;10:24–39. doi:10.1016/j.molonc.2015.07.004.
- [34] Addona T a, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* 2009;27:633–41. doi:10.1038/nbt0909-864b.
- [35] Kuhn E, Whiteaker JR, Mani DR, Jackson AM, Zhao L, Pope ME, et al. Interlaboratory evaluation of automated, multiplexed peptide immunoaffinity enrichment coupled to multiple reaction monitoring mass spectrometry for quantifying proteins in plasma. *Mol Cell Proteomics* 2012;11:M111.013854. doi:10.1074/mcp.M111.013854.
- [36] Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 2006;24:971–83. doi:10.1038/nbt1235.

- [37] Van Eeckhaut A, Lanckmans K, Sarre S, Smolders I, Michotte Y. Validation of bioanalytical LC-MS/MS assays: evaluation of matrix effects. *J Chromatogr B Analyt Technol Biomed Life Sci* 2009;877:2198–207. doi:10.1016/j.jchromb.2009.01.003.
- [38] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845–67. doi:10.1074/mcp.R200007-MCP200.
- [39] Hoofnagle AN, Whiteaker JR, Carr SA, Kuhn E, Liu T, Massoni SA, et al. Recommendations for the Generation, Quantification, Storage, and Handling of Peptides Used for Mass Spectrometry-Based Assays. *Clin Chem* 2016;62:48–69. doi:10.1373/clinchem.2015.250563.
- [40] Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, et al. Targeted Peptide Measurements in Biology and Medicine: Best Practices for Mass Spectrometry-based Assay Development Using a Fit-for-Purpose Approach. *Mol Cell Proteomics* 2014;13:907–17. doi:10.1074/mcp.M113.036095.
- [41] Musgrove E a, Sutherland RL. Biological determinants of endocrine resistance in breast cancer. *Nat Rev Cancer* 2009;9:631–43. doi:10.1038/nrc2713.
- [42] Polyak K. Tumor heterogeneity confounds and illuminates: a case for Darwinian tumor evolution. *Nat Med* 2014;20:344–6. doi:10.1038/nm.3518.
- [43] de Graauw M, van Miltenburg MH, Schmidt MK, Pont C, Lalai R, Kartopawiro J, et al. Annexin A1 regulates TGF-beta signaling and promotes metastasis formation of basal-like breast cancer cells. *Proc Natl Acad Sci U S A* 2010;107:6340–5. doi:10.1073/pnas.0913360107.
- [44] Al Saleh S, Al Mulla F, Luqmani Y a. Estrogen receptor silencing induces epithelial to mesenchymal transition in human breast cancer cells. *PLoS One* 2011;6:e20610. doi:10.1371/journal.pone.0020610.
- [45] Alldridge LC, Bryant CE. Annexin 1 regulates cell proliferation by disruption of cell morphology and inhibition of cyclin D1 expression through sustained activation of the ERK1/2 MAPK signal. *Exp Cell Res* 2003;290:93–107. doi:10.1016/S0014-4827(03)00310-0.
- [46] Bist P, Phua QH, Shu S, Yi Y, Anbalagan D, Lee LH, et al. Annexin-A1 controls an ERK-RhoA-NF- $\kappa$ B activation loop in breast cancer cells. *Biochem Biophys Res Commun* 2015;461:47–53. doi:10.1016/j.bbrc.2015.03.166.
- [47] Schwappacher R, Rangaswami H, Su-Yuo J, Hassad A, Spitler R, Casteel DE. cGMP-dependent protein kinase I $\beta$  regulates breast cancer cell migration and invasion via interaction with the actin/myosin-associated protein caldesmon. *J Cell Sci* 2013;126:1626–36. doi:10.1242/jcs.118190.
- [48] Pal SK, Childs BH, Pegram M. Triple negative breast cancer: unmet medical needs. *Breast Cancer Res Treat* 2011;125:627–36. doi:10.1016/j.micinf.2011.07.011.Innate.
- [49] Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, et al. Toward understanding and exploiting tumor heterogeneity. *Nat Med* 2015;21:846–53. doi:10.1038/nm.3915.

- [50] Hiley CT, Swanton C. Pruning Cancer's Evolutionary Tree with Lesion-Directed Therapy. *Cancer Discov* 2016;6:122–4. doi:10.1158/2159-8290.CD-15-1493.
- [51] Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* 2015. doi:10.1038/nm.3886.
- [52] Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell* 2012;150:12–27. doi:10.1016/j.cell.2012.06.013.
- [53] Rivera CM, Ren B. Mapping human epigenomes. *Cell* 2013;155:39–55. doi:10.1016/j.cell.2013.09.011.
- [54] Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* 2000;16:168–74. doi:10.1016/S0168-9525(99)01971-X.
- [55] Barrow TM, Michels KB. Epigenetic epidemiology of cancer. *Biochem Biophys Res Commun* 2014;455:70–83. doi:10.1016/j.bbrc.2014.08.002.
- [56] Rodríguez-Rodero S, Delgado-Álvarez E, Fernández AF, Fernández-Morera JL, Menéndez-Torre E, Fraga MF. Epigenetic alterations in endocrine-related cancer. *Endocr Relat Cancer* 2014;21:R319–30. doi:10.1530/ERC-13-0070.
- [57] Ross-Innes CS, Stark R, Teschendorff AE, Holmes K a., Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 2012;481:389–93. doi:10.1038/nature10730.
- [58] Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli G a., et al. CORRIGENDUM: Progesterone receptor modulates ER $\alpha$  action in breast cancer. *Nature* 2015;523:313–7. doi:10.1038/nature14583.
- [59] Feng Q, Zhang Z, Shea MJ, Creighton CJ, Coarfa C, Hilsenbeck SG, et al. An epigenomic approach to therapy for tamoxifen-resistant breast cancer. *Cell Res* 2014;2471:1–11. doi:10.1038/cr.2014.71.
- [60] Jansen MPH, Knijnenburg T, Reijm EA, Simon I, Kerkhoven R, Droog M, et al. Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. *Cancer Res* 2013;73:6632–41. doi:10.1158/0008-5472.CAN-13-0704.
- [61] Martens JWM, Nimmrich I, Koenig T, Look MP, Harbeck N, Model F, et al. Association of DNA methylation of phosphoserine aminotransferase with response to endocrine therapy in patients with recurrent breast cancer. *Cancer Res* 2005;65:4101–7. doi:10.1158/0008-5472.CAN-05-0064.
- [62] Massink MPG, Kooi IE, van Mil SE, Jordanova ES, Ameziane N, Dorsman JC, et al. Proper genomic profiling of (BRCA1-mutated) basal-like breast carcinomas requires prior removal of tumor infiltrating lymphocytes. *Mol Oncol* 2015;9:1–12. doi:10.1016/j.molonc.2014.12.012.

- [63] Thomas SJ, Snowden JA, Zeidler MP, Danson SJ. The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours. *Br J Cancer* 2015;113:365–71. doi:10.1038/bjc.2015.233.
- [64] Schulze A, Harris AL. How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature* 2012;491:364–73. doi:10.1038/nature11706.
- [65] Belvedere R, Bizzarro V, Popolo A, Piaz FD, Vasaturo M, Picardi P, et al. Role of intracellular and extracellular annexin A1 in migration and invasion of human pancreatic carcinoma cells. *BMC Cancer* 2014;14:1–15. doi:10.1186/1471-2407-14-961.
- [66] Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013;154:442–51. doi:10.1016/j.cell.2013.06.044.
- [67] Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* 2013;8:2180–96. doi:10.1038/nature13314.A.
- [68] Mao Y, Keller ET, Garfield DH, Shen K, Wang J. Stromal cells in tumor microenvironment and breast cancer. *Cancer Metastasis Rev* 2012;30:3–15. doi:10.1007/s10555-012-9415-3.
- [69] Qian B-Z, Pollard JW. Macrophage diversity enhances tumor progression and metastasis. *Cell* 2010;141:39–51. doi:10.1016/j.cell.2010.03.014.



## **Chapter 10**

**Samenvatting**

**Summary**

**Sommario**

## Samenvatting

Een belangrijke doodsoorzaak bij terugkerende oestrogenreceptor-positieve (ER, ESR1) borstkanker is de resistentie tegen een behandeling met tamoxifen. Hoewel een aantal studies hebben aangetoond dat er biomarkers zijn voor het voorspellen van de resultaten van een behandeling met tamoxifen, zoals mutaties of isovormen van ESR1, worden deze tot dusverre niet in de patiëntenzorg toegepast. Dankzij nieuwe technologische innovaties in de massaspectrometrie (MS) is het mogelijk om bijna het gehele menselijke proteoom nauwkeurig te meten. Hierdoor is ze geschikt voor het ontdekken van nieuwe biomarkers.

Recent hebben we in ons laboratorium een proteoom-pipeline voor weefsel ontwikkeld door laser capture microdissectie (LCM) te combineren met hoge resolutie MS. Hierdoor konden we van meer dan 4.000 samples van verrijkte epitheelcellen afkomstig van borstcarcinoom de eiwitten nauwkeurig meten. Toch is het voordeel van het analyseren van LCM materiaal boven lysaten van hele weefsel bij vergelijkende proteoomanalyses nooit in z'n geheel uitgediept. In **hoofdstuk 3** bepalen we de verschillen voor MS, op peptide- en eiwitniveau, tussen de gecombineerde LCM en specimens van heel weefsel. Bij gewone proteoomanalyse is te zien dat er een grotere hoeveelheid peptide (+47%) en eiwitten (+20%) in de LCM-specimens geïdentificeerd kon worden vergeleken met de lysaat-samples afkomstig van heel weefsel (WTL). De laatste hadden weer last van een hoge dichtheid van plasma afgeleide eiwitten en een hoger percentage ontbrekende waarden. Bovendien werden er bij het beoordelen van de specifieke verrijking van ER-positieve borstkankermarkers tussen de patiëntengroepen significante verschillen waargenomen tussen de patiënten met een slechte en een goede uitkomst op de behandeling bij de LSM-samples (bijv. HER2), maar niet bij de WTL's. Daaruit concludeerden we dat bij de proteoomanalyse van het LCM-materiaal het tumorproteoom beter ontleed wordt dan de WTL's.

Nadat we het voordeel van de LCM-techniek hadden bepaald bij het identificeren en meten van tumor-specifieke eiwitten, gebruikten we deze aanpak om een eiwit-classifier te ontwikkelen om de resultaten met tamoxifen te voorspellen voor terugkerende borstkanker (**hoofdstuk 4**). Door onze analyse van de oefencohort (bestaande uit 56 patiënten) kwamen we tot 4 veronderstelde eiwitbiomarkers die bij de testgroep de patiënten met een slechte uitkomst significant voorspeld hadden (sensitiviteit: 86,7%; specificiteit: 41,5%; hazard ratio [HR] = 2,32; 95% betrouwbaarheidsinterval [BI]: 1,29 tot 4,17; Log rank  $P = 0,004$ ). Onze 4-eiwit-signatuur kan bepalen of een patiënt met uitgezaaide borstkanker niet zou profiteren van een behandeling met tamoxifen. Onze proteoom-pipeline voor weefsel is echter niet makkelijk inzetbaar bij de patiëntenzorg, vanwege de uitgebreide voorbereiding van de

samples, het meten en de tijd die het analyseren vraagt. Daarom richtten we ons op het ontwikkelen van een test die beter toegerust is voor de klinische setting en die onze ontdekkingen technisch bevestigt. We gingen eerst na of immunohistochemie (ICH) een alternatief kon zijn, hoewel maar één van de 4 signatuureiwitten (namelijk PDCD4) met succes gevalideerd was. Dit kwam waarschijnlijk door de semi-kwantitatieve analyse van eiwitdichtheid die kenmerkend is voor IHC. Om dit probleem aan te pakken, richtten we ons op het ontwikkelen van een kwantitatieve test.

Multiple reaction monitoring (MRM)-MS biedt een zeer nauwkeurige methode voor het meten van de eiwitdichtheid in biologische samples. Echter, zoals gezegd, kan interferentie van eiwitten die overvloedig aanwezig zijn (zoals albumine in plasma en serum) een nauwkeurige kwantificatie in de weg staan. Daarom verbonden we verrijking op basis van immunoaffiniteit door middel van antipeptide antilichamen met MRM-MS om zulke interferentie zo laag mogelijk te houden (gedefinieerd als ‘immuno-MRM’ [iMRM]; **hoofdstuk 5**). Omdat voor verrijking op basis van immunoaffiniteit een hoeveelheid eiwitten nodig is die hoger ligt dan met LCM verkregen wordt, gebruikten we hiervoor de WTL-specimens. Zoals beschreven in **hoofdstuk 6**, blijkt uit onze analyse dat de iMRM-metingen van de 4-eiwit-signatuur voorspellingen doen die vergelijkbaar zijn met die van de hoge resolutie MS op de LCM-specimens. Daarentegen geeft de voorspelling gebaseerd op de data van de hoge resolutie MS op de WTL-samples aan geen onderscheidend vermogen te hebben voor de patiëntengroepen. Hieruit blijkt dat iMRM een voorspellend vermogen heeft dat vergelijkbaar is met dat van de hoge resolutie MS, maar ook dat het effectief in staat is de interferentie van hoge eiwitdichtheden te omzeilen. Bovendien wist de 4-eiwit-signatuur de resultaten van de patiënten van een klein, onafhankelijk cohort van serumsamples te voorspellen. Alles bij elkaar genomen is deze techniek beter toepasbaar geworden, doordat onze eiwit-classifier de resultaten van de patiënt kan voorspellen op basis van zowel weefsel als serum afkomstig van de borstkankerpatiënt en doordat de iMRM-samples op allerlei verschillende manieren voorbereid kunnen worden. Hierdoor is ze geschikter voor klinische diagnose.

Hoewel het veelbelovend is dat onze 4-eiwit-classifier kan aangeven welke patiënten baat hebben bij een behandeling met tamoxifen, is het onwaarschijnlijk dat deze markers een uitgebreid overzicht geven van het mechanisme achter tamoxifenresistentie. Omdat deze tumoren zowel op morfologisch als op moleculair niveau zeer heterogeen zijn, is het daarbij goed mogelijk dat meerdere mechanismen een rol spelen bij tamoxifenresistentie. Daarom

hebben we in onze gecombineerde MS-datasets gekeken welke verrijkte eiwitten aanwezig waren bij de patiënten met een slecht klinisch beloop. Hieruit volgden dat ANXA1 en CALD1 onze beste 2 kandidaten waren (**hoofdstuk 7**). Door IHC-validatie werd bevestigd dat deze markers klinisch geassocieerd zijn met resistentie tegen de behandeling en dat ze alleen tot expressie kwamen in een kleine groep ER-positieve tumoren. Aan de andere kant wees analyse van de signaalroute erop dat expressie van ANXA1 en CALD1 geassocieerd werd met de downregulatie van ER en het krijgen van een mesenchymaal-achtig fenotype.

Vervolgens bepaalden we in **hoofdstuk 8** de klinische relevantie van een eerder geïdentificeerde epigenetische biomarker, PSAT1, een enzym dat essentieel is voor de biosynthese van serine. Net als ANXA1 en CALD1 werd PSAT1 geassocieerd met een slechte uitkomst bij de behandeling met tamoxifen en kwam het alleen in een minderheid van de weefsels tot expressie. Daarbij bleek uit de algemene testanalyse dat tumoren met PSAT1-expressie verrijking lieten zien in de genen die niet alleen bij de biosynthese van serine betrokken waren maar ook bij de Jak/STAT-signalroute. Alles bij elkaar genomen kunnen deze markers wijzen op subgroepen ER-positieve tumoren met verschillende resistentiemechanismen tegen tamoxifen. Omdat deze markers prognostische en/of predictieve waarde hebben, kunnen ze gebruikt worden voor het vaststellen van nieuwe targets voor de behandeling.

Samengevat hebben we hier aangetoond dat de technologieën van de proteomica een rendabel technologisch platform bieden voor het ontdekken en verifiëren van biomarkers voor de resistentie tegen tamoxifen. Bovendien hebben we hier laten zien dat een aantal markers niet alleen geassocieerd zijn met subgroepen van tumoren die resistent zijn tegen tamoxifen, maar dat ze ook kunnen leiden tot alternatieve targets voor de behandeling.

## Summary

Resistance to tamoxifen therapy is a major cause of death in recurrent estrogen receptor (ER, ESR1) positive breast cancer. Although several studies have reported biomarkers capable of predicting tamoxifen treatment outcome, such as *ESR1* mutations or isoforms, none have been introduced in clinical diagnostics so far. Recent improvements in mass spectrometry (MS)-related technologies have shown that proteomic approaches are nowadays capable of accurately quantifying almost the entirety of the human proteome, rendering them suitable for biomarker discovery studies.

In our laboratory, we have recently developed a tissue proteomic pipeline by coupling laser capture microdissection (LCM) to high resolution MS, enabling to accurately quantify more than 4,000 proteins in breast carcinoma epithelial cell-enriched samples. Still, the advantage of LCM over whole tissues in comparative proteomic analyses has never been fully elucidated. In **Chapter 3** we assessed differences at the MS, peptide and protein levels between paired LCM and whole tissue specimens. Global proteomic analysis showed that a higher amount of peptides (+47%) and proteins (+20%) could be identified in LCM specimens when compared to whole tissue lysate (WTL) samples, which in turn suffered from high abundance of plasma-derived proteins and a higher percentage of missing values. Furthermore, upon assessing specific enrichment of ER positive breast cancer markers between patient groups, significant differences were observed between poor and good outcome patients in LCM samples (e.g. HER2), but not in WTLs. Therefore, we concluded that proteomic analysis of LCM material better dissects the tumor proteome than WTLs.

Having established the advantage of LCM in identifying and quantifying tumor-specific proteins, we used this approach to develop a protein signature to predict tamoxifen outcome in recurrent breast cancer (**Chapter 4**). From our analysis of the training cohort (comprising 56 patients) we identified 4 putative protein biomarkers, which significantly predicted poor outcome patients in the test set (sensitivity: 86.7%; specificity: 41.5%; hazard ratio [HR] = 2.32; 95% confidence interval [CI]: 1.29 to 4.17; Log-rank  $P = 0.004$ ). Our 4-protein signature could determine whether a patient with metastatic breast cancer would fail from tamoxifen therapy, though our tissue proteomic pipeline would find little use in clinical diagnostics due to the extensive sample preparation, measurement, and data analysis time required. We therefore aimed to develop a more clinically feasible assay that could provide technical verification of our findings. We first evaluated whether immunohistochemistry (IHC) could

be used as an alternative approach, though only one out of the 4 signature proteins (i.e. PDCD4) was successfully validated. This was likely due to the semi-quantitative measure of protein abundance intrinsic to IHC. Therefore, to address this issue, we focused on the development of a more quantitative assay.

Multiple reaction monitoring (MRM) MS offers a precise and accurate method to measure protein abundance in biological samples. However, as previously mentioned, interference from highly abundant proteins (e.g. albumin in plasma and serum) can impair accurate analyte quantitation. Therefore, we coupled immunoaffinity enrichment by anti-peptide antibodies to MRM MS in order to minimize such interference (defined as immuno-MRM [iMRM]; **Chapter 5**). Because immunoaffinity enrichment requires protein amounts beyond the yield of LCM, we employed WTL specimens for these experiments. As described in **Chapter 6**, our analyses showed that iMRM measurement of the 4-protein signature achieved comparable prediction when compared to high resolution MS of LCM specimens, while the one based on high resolution MS data of WTL samples showed no discriminatory power of patient groups. This showed that iMRM not only achieves similar prediction compared with high resolution MS, but that it also effectively bypasses interference from highly abundant proteins. In addition to this, the 4-protein signature was capable of predicting patient outcome in a small independent cohort of serum samples. Taken together, the ability of our protein signature to predict patient outcome in both breast cancer patient derived tissues and sera and the fact that iMRM sample preparation is highly multiplexable expand the applicability of the technique, rendering it more feasible for clinical diagnostics.

Although our 4-protein signature shows promise by distinguishing which patients would not benefit from tamoxifen treatment, it is unlikely that these markers offer a comprehensive overview of tamoxifen resistance mechanisms. Furthermore, as tumors are characterized by a high degree of heterogeneity both at the morphological and molecular levels, it is possible that multiple mechanisms may lead to tamoxifen therapy resistance. Therefore, we assessed which proteins were enriched in poor outcome patients in our combined MS datasets, which led to the identification of ANXA1 and CALD1 as the top 2 candidates (**Chapter 7**). IHC validation confirmed the clinical association of these markers to therapy resistance and showed that they were expressed only in a small set of ER positive tumors. On the other hand, pathway analysis indicated that ANXA1 and CALD1 expression was associated with the downregulation of ER and the acquisition of a mesenchymal-like phenotype.

Subsequently, in **Chapter 8** we assessed the clinical relevance of a previously identified epigenetic biomarker, PSAT1, a key enzyme in serine biosynthesis. Similarly to ANXA1 and CALD1, PSAT1 was associated to poor outcome on tamoxifen therapy and was expressed only in a minority of tissues. In addition to this, global test analysis showed that PSAT1 expressing tumors showed enrichment in genes involved not only in serine biosynthesis, but also in the Jak/STAT pathway. Taken together, these markers may be indicative of subgroups of ER positive tumors with diverse mechanisms of resistance to tamoxifen. As these markers offer prognostic and/or predictive value, they may be used to define new therapeutic targets.

In conclusion, we have here shown that proteomic technologies offer a viable technological platform for discovery and verification of tamoxifen resistance biomarkers. Furthermore, we have here shown that several markers were not only associated to subgroups of tamoxifen resistant tumors, but could also lead to alternative therapeutic targets.

## Sommario

La resistenza al tamoxifene costituisce una delle maggiori cause di morte nei casi recidivi di tumore alla mammella positivi al recettore degli estrogeni (ER, ESR1). Nonostante molteplici studi abbiano riportato diversi biomarcatori capaci di predire l'esito della terapia a base di tamoxifene, come ad esempio la presenza di mutazioni o isoforme di ESR1, nessuno di questi è stato finora introdotto a livello diagnostico. I recenti miglioramenti nelle tecniche associate alla spettrometria di massa (MS) hanno dimostrato come gli approcci di proteomica siano oggi in grado di quantificare quasi completamente il proteoma umano, rendendoli adatti a studi atti a definire nuovi biomarcatori.

Nel nostro laboratorio abbiamo recentemente sviluppato un protocollo di preparazione di tessuti umani per la proteomica, il quale combina la micro-dissezione laser (LCM) con la MS ad alta risoluzione, ed è capace di quantificare accuratamente più di 4.000 proteine a partire dall'analisi di tessuti arricchiti di carcinoma mammario. Il vantaggio nell'analisi proteomica di tessuti arricchiti tramite LCM rispetto a tessuti interi (WTL) non è stato tuttavia completamente chiarito. Nel **Capitolo 3** abbiamo valutato le differenze al livello di spettri di massa e nel numero di peptidi e proteine identificati tra tessuti arricchiti di cellule di carcinoma mammario e la loro controparte intera. La nostra analisi proteomica globale ha indicato che si possa identificare un più alto numero di peptidi (+47%) e proteine (+20%) nei campioni arricchiti tramite LCM rispetto ai WTL, che a loro volta hanno mostrato contaminazioni da parte di proteine plasmatiche e una più alta percentuale di valori mancanti. Al momento di accertare differenze di espressione di proteine specifiche per i tumori positivi a ESR1, abbiamo osservato differenze statisticamente significative (esempio: HER2) tra gruppi di pazienti – manifestanti un buono o un cattivo esito alla terapia al tamoxifene – i cui tessuti erano stati arricchiti tramite LCM. Dall'altro lato, nessuna differenza è stata osservata dalla misurazione degli stessi tessuti non arricchiti. Alla luce di questo, abbiamo concluso che l'analisi proteomica di tessuti arricchiti tramite LCM garantisce migliore risoluzione del proteoma tumorale.

Avendo stabilito il vantaggio posto dall'arricchimento tramite LCM nella quantificazione di proteine tumore-specifiche, abbiamo utilizzato questo approccio al fine di sviluppare un classificatore di proteine capace di predire l'esito della terapia a base di tamoxifene in cancro al seno recidivo (**Capitolo 4**). Dalla nostra analisi della coorte "training" (comprendente 56 pazienti) abbiamo identificato 4 presunti biomarcatori proteici, i quali hanno predetto in

maniera significativa un esito negativo alla terapia a base di tamoxifene (sensitività: 86.7%; specificità: 41.5%; rapporto di rischio [HR] = 2.32; intervallo di confidenza 95%: da 1.29 a 4.17;  $P$  del rango logaritmico = 0.004) nella coorte “test”. Il nostro classificatore, costituito da 4 proteine, ha potuto determinare se pazienti con cancro al seno metastatico potessero beneficiare di terapia al tamoxifene, tuttavia il nostro protocollo di preparazione di tessuti umani troverebbe uno scarso uso nella diagnostica clinica a causa delle lunghe tempistiche nella preparazione dei campioni, nell’analisi di MS e nel processamento dei dati. In questa prospettiva, ci siamo focalizzati sullo sviluppo di una metodologia diagnostica più idonea alla clinica e che possa provvedere una verifica tecnica dei nostri risultati. In un primo momento, abbiamo valutato se l’immunoistochimica (IHC) potesse essere usata come un approccio alternativo, tuttavia solo uno dei quattro biomarcatori (PDCD4) è stato confermato con successo. E’ molto probabile che questo sia da attribuire alla misura semi-quantitativa intrinseca alla IHC e, per questo, abbiamo indirizzato i nostri sforzi nello sviluppo di una tecnica più quantitativa.

Il “multiple reaction monitoring” (MRM) tramite MS costituisce un metodo preciso ed accurato nella misura quantitativa delle proteine in campioni biologici. Tuttavia, come menzionato in precedenza, l’interferenza generata dalla presenza di proteine ad alta concentrazione (per esempio: albumina in campioni di plasma e siero) può alterare la corretta quantificazione di un analita. Al fine di minimizzare questa interferenza abbiamo combinato l’arricchimento tramite immuno-affinità, che si avvale di anticorpi anti-peptide, con la MS MRM (definita come immuno-MRM [iMRM]; **Capitolo 5**). A causa del fatto che l’arricchimento tramite immuno-affinità richiede quantitativi proteici superiori alla normale resa della LCM, per questi esperimenti abbiamo utilizzato WTL. Come abbiamo descritto nel **Capitolo 6**, le nostre analisi hanno mostrato che la misura delle 4 proteine del classificatore tramite MS iMRM garantisce una predizione simile a quella ottenuta dall’analisi di MS ad alta risoluzione di campioni arricchiti. Dall’altro lato, la predizione basata sulle misurazioni ad alta risoluzione di campioni non arricchiti non è risultata statisticamente significativa. Questi risultati hanno dimostrato che non solo la MS iMRM è capace di predire significativamente i gruppi di pazienti non-responsivi alla terapia al tamoxifene, ma è anche in grado di aggirare l’interferenza da proteine presenti ad alte concentrazioni. Oltre a questo, la nostra lista di quattro proteine ha dimostrato una significativa predizione anche in campioni di siero derivanti da una coorte pilota di pazienti indipendente. Alla luce di questi risultati, e a

fronte del fatto che la preparazione dei campioni per la MS iMRM e' altamente 'multiplexable', l'iMRM risulta essere una tecnologia piu' adatta all'uso clinico.

Nonostante il nostro classificatore di proteine prometta una significativa discriminazione dei pazienti non-responsivi alla terapia a base di tamoxifene, e' improbabile che questi marcatori siano capaci di delineare completamente i meccanismi molecolari di resistenza alla terapia. Per di piu', visto che i tumori sono caratterizzati da un elevato livello di eterogeneità sia a livello morfologico che molecolare, è possibile che molteplici meccanismi stiano alla base della resistenza al tamoxifene. Alla luce di questo, abbiamo analizzato le nostre due coorti di pazienti combinate l'una con l'altra, identificando ANXA1 e CALD1 come i primi due candidati biomarcatori (**Capitolo 7**). La validazione tramite IHC ha confermato l'associazione di queste proteine alla resistenza al tamoxifene e ha mostrato che queste fossero espresse solo in una minoranza dei tumori. Dall'altra parte, l'analisi dei "pathway" biologici ha indicato che ANXA1 e CALD1 sono associati alla bassa espressione di ER e all'acquisizione di un fenotipo simil-mesenchimale.

Successivamente, nel **Capitolo 8**, abbiamo valutato la rilevanza clinica di un biomarcatore epigenetico precedentemente identificato, PSAT1, un enzima chiave nella sintesi di serina. Come osservato per ANXA1 e CALD1, anche PSAT1 e' stato associato a una rapida progressione del tumor metastatico trattato con tamoxifene e ne stata riscontrata l'espressione solo in una minoranza dei tessuti. L'analisi "global test" ha rivelato che nei tumori esprimenti PSAT1 vi era un arricchimento nell'espressione di geni non solo coinvolti nella sintesi della serina, ma anche di molecole coinvolte nella cascata Jak/STAT. Nel complesso, questi marcatori possono essere indicativi di sottogruppi di tumori positivi a ER aventi diversi meccanismi di resistenza alla terapia a base di tamoxifene. Inoltre, questi marcatori non solo possiedono un valore prognostico e predittivo, ma possono probabilmente essere utilizzati per sviluppare nuove terapie mirate.

In conclusione, abbiamo qui dimostrato che le tecniche di proteomica offrono una solida piattaforma per la scoperta e la validazione di biomarcatori per la resistenza al tamoxifene. Abbiamo qui inoltre mostrato che alcune molecole non erano solo associate a sottogruppi di tumori, ma potrebbero costituire obiettivi terapeutici alternativi.

# **Appendices**

**Acknowledgements**

**List of Publication**

**PhD Portfolio**

**Curriculum Vitae**

## Acknowledgements

I started my doctoral studies with the motto “*nemo ad impossibilia tenetur*” (a Latin brocard meaning: nobody has to do the impossible) in my mind. This to keep me from challenging myself with impossible endeavors, and to try to keep my feet on the ground.

Being at the end of this period in my life I can only say one thing: I was wrong. This because I realized that scientists can make the impossible possible and most adversities can be overcome, one way or the other, through sound scientific research the burning passion and commitment a scientist gives to his work. It goes without saying that this ‘make the impossible possible’ concept could not have been realized without the people around me, both outside and inside my work at the Erasmus Medical Center (EMC).

First of all, I would like to thank Arzu, my first co-supervisor. You hired me back in 2011 with a smile on your face and a lot of charisma that made me immediately accept the position. On my side, I started as a shy and uptight – at least in my perception - PhD in your group of breast cancer proteomics. This due to the fact that I came from an Italian University background (we discussed much about how things are going on there, so I won't repeat myself in here), with no background in (breast) cancer, but a will to learn as much as possible. You taught me many things in this period, not only about breast cancer and proteomics, but also about life (like patience).

Approximately by the end of my PhD period you have decided to pursue a different career, which took a lot of time in your schedule and made you less (physically) present at the EMC. Still, you managed to grant me full support during this period, showing that you are, as always, full of energy.

For this, I thank you.

Next in line is my co-supervisor, and friend, John Martens. You are a wonderful person which has always an eye for new techniques and scientific discoveries, you always care to expand your field of knowledge beyond your main expertise (genomics) and that makes you a very good scientist. On top of this, I deeply enjoyed our ‘borrel’ (drinks) evenings together with the other people in the lab and beyond, such as your ever-charming partner in life Jenny. Together you are wonderful people to go out with, from all perspectives.

On top of the “*pik orde*” (hierarchy) is my supervisor, John Foekens. When I joined the lab I was told that you advised Arzu to hire me due to the fact that I would have been a funny guy to be around to. It is curious that you had this impression of me since - initially - I had a quite terrified impression of you. This is due to your seemingly harsh comments on my presentations, manuscripts, etc.

Also in this case, I have to admit that I was wrong. You are indeed a great person to be around to (also for a smoke) and a terrific promotor. You always look at things from many perspectives and try to decipher every facet of things, a characteristic that sometimes I recognize in myself as well. You always showed me your support and encouragement and I cannot thank you enough for this.

Next, I would like to acknowledge the people that have made a significant impact during my period at the EMC.

NingQing (colloquially ‘Fangu’) is one of my fellow PhD students that accompanied me for most of this period of my life. Although I did not have a nice initial impression of you due to your (too) high confidence, and sometimes arrogance, I came to think of you as one of my dearest friends. Although you were - and still are - obsessed by science (especially since you started your Post-doc), I realized that inside you’re a wonderful and kind person. While on one hand we unceasingly worked shoulder-to-shoulder in the lab, on the other hand we shared many good moments with our friends in Rotterdam and elsewhere. No matter where our paths in life will take us, know that I will be always inspired by you.

Rene is another wonderful person I fortunately spent my PhD with. We started our graduate studies the same day and since then you taught me a lot of things, about mass spectrometry, about bioinformatics, and about the complex mechanisms of the Dutch governmental administration. I think that without you my promotion would have come much later. Due to your always present family obligations, you could not join our ‘borrel’ evenings as often as you wanted but every time you were a most pleasant company, such as in Berlin, where we roamed aimlessly for a place to eat and actually you chose a place next to the Brandenburg gate which seemed like nothing, but actually provided amazing food and drinks. From that time I recognized you have a good eye for good restaurants.

Next, I would now like to thank all the people in the lab that I've seen pass by in the lab and that have seen me do the same later on.

Diana and Katharina - two fellow PhDs during my period at the EMC - I dedicate this paragraph to you two since you were always together, much like an unique entity. While Diana has always been the sunny and 'up-in-the-air' one, Katharina has always been the most pragmatic of the two. While I learned to be easygoing from one I learned to also keep my feet on the earth from the other.

(Mr... now Dr.) Muhammad is another person of which I have a deep esteem. You came here from a very different culture but you showed to me that your values are our values as well. With your impeccable aplomb you always shed an aura of maturity on all of us fellow PhDs in the group. So much that I began to think you were like a 'PhD grandfather'.

Also Joan, our lab manager and '*factotum*', needs a special paragraph. As in the case of NingQing, and although we sit in different buildings, you became my '*buurvrouw*'. You are a wonderful person with a joyful personality, always cheering people up and helping as best as you can for matters even beyond your duty. I believe that you are one of the most kind and caring people I ever met in my life. I will always remember you once told me that you wanted a dry white wine while '*achter de geraniums zitten*' (sitting behind the geraniums means that you are retired). When you will retire I hope you get champagne instead, you certainly deserved it.

A dedicated paragraph needs to be written for our bioinformatics and statistics team: Marcel and Maxime. Marcel has been the funny guy of the group since day 1 by always making jokes, although a more fitting definition would be an Italian one: '*avere sempre la battuta pronta*' (always be ready with a joke). You always proved insightful in bioinformatics issues and at the same time made everyone laugh with your caricatures.

On the other side stands Maxime, an expert in clinical statistics. We had so many meetings during my PhD that I can hardly enumerate them, most of the times because I could not understand your workflow from the raw Stata outputs you sent me. You were always a good colleague that helped me a lot and prompted me to deepen my knowledge of statistics, even if it was probably the hard way.

Next, I would like to acknowledge the immunohistochemistry team: Mieke, Renee and Anita. You are all nice people to talk to and are very knowledgeable of your technique. In all these years I dare say that a great part of my work would have not been performed properly without you or not performed at all.

During my PhD I supervised two students, Yagmur and Titia. You both have demonstrated to be independent and enthusiastic scientists, despite the fact that I could be harsh and demanding on you. Keep up this good and strong attitude.

Going to the genomics part of the lab, I want to acknowledge first Anieta, Maurice and Jean. You are all very good scientists and a nice people to be around to: Anieta particularly at our weekend meetings in the lab (I still remember how startled you were when I showed up in the lab without you noticing it), Maurice for our corridor scientific conversations, and Jean for passing by my office for a brief chat ‘just because’.

I wish also to thank all the people that have been or still are part of the group: Els, Jozien, Anne, Kirsten, Corinne, Vanja, Wendy and Michelle.

I would then like to spend some words for the ‘newcomers’ of the lab.

Wendy, although we started around the same period, you came stably in our group a bit later. I still remember the embarrassment I felt in Ellecom (the literal ‘*middle of nowhere*’) during the Oncology course when I finally realized we were from the same department amidst people from all over the Netherlands. Although you have chosen to pursue a career in science over the possibility to join the medical ward, I’ll still call you ‘*medic (!)*’.

Marjolein, the mitochondrial DNA girl who joined us with from the company world to pursue her PhD in a highly innovative field, need also a big thank you. You are a rare source of positivity which can be contagious sometimes, I hope you keep it that way.

I would like to also thank the rest of the group: Lisanne, Jaco, Mai, Zara, Lindsay and Nick.

Outside of our lab, my most heartfelt thanks go to the people of the laboratory of neuro-oncology and mass spectrometry.

Theo, during these years you have been proven a wonderful collaborator, always interested whenever one of us would come to you with an outline for a new project. Your comments on all our work have always been insightful and very constructive, and for this I thank you.

Christoph, I believe of the people I have met here, you're the most kind and helpful of them all. On top of this, you have a deep technical knowledge of your MS babies (in particular the Orbitraps) and your help has always been invaluable during these years. I want to thank you for all your help and wish you good luck when you'll be defending your thesis!

Lennard, the MRM guy of the group (or so I'd like to think), we had a great time in Minneapolis when our groups went together to ASMS. Nice beers, nice steaks! I leave at the time in which you take an important step in your life...to you, and Diana, I wish the best!

Coskun ('shoarma meneer') and Giovanni, I had really good times with you, both in Minneapolis and in Rotterdam, whether it was at work or outside of it. I wish you the best of luck with your new position (Giovanni) and for your PhD (Coskun)!

Martijn, the person I pestered for most of my experiments with the Qtrap. You have proven to have great patience with me during these years and clearly you are one of the most knowledgeable guys in MS in the group. I want to thank you for all of this and for helping me reach my goals, and wish you the best with yours.

Next I want to thank the remaining people in Theo's lab. Marcel, with whom we had nice and funny conversations (especially when you brought your daughter at work for 'papa-dag'). Nick, for always being making silly jokes, 'grappenmaker'! Also I would like to thank the other very nice people in this lab: Diana, Lona, Ingrid and Peter.

My deepest thanks also go to the people that I've been working with during my stay at the Broad Institute, first of all Steve. It has been a pleasure and an honor being able to work with you in Cambridge (MA). You have kept the door open for me whenever I needed to speak with you and you have always been supportive of all my efforts there.

To Erik, my American supervisor at the Broad, go my deepest respects. You have taught me a lot about mass spec, about Boston and about the East Coast way. I will treasure all of these lessons.

I would also like to thank Hasmik and Rushdi, with whom I had incredible times in American coffee shops and in San Diego, respectively. You are two terrific scientists and incredibly pleasant people to be around to.

I would like to spend a few words on some valuable people from the Josephine Nefkens Institute and outside of it.

Alex(andra), you were the first friend I made in the Netherlands when I first arrived in this country from Italy. You are a wonderful person, always caring but also possessing a hot temper (I dare you say it is not true)...I am just glad I did not experience the latter on myself, yet. Putting aside these jokes, I believe you taught me a lot during my stay here, about the Netherlands and about friendship. Although we could not see each other that often, every event with you lifted my spirits, and just for this I thank you. With my moving out of the country I hope I can eventually find a way to see you in person and enjoy your personality again, in the Netherlands or elsewhere.

Matthias, a fellow PhD and ex-flat mate, you are an exceptional scientist, one I always looked upon for advice. We spent a lot of time together, giving parties in the garden and in the park. I will always keep those beautiful memories with me. Also, I wish you the best in your career, I am sure it will be outstanding.

Rene, you are a wonderful person, not only because you are a master of R, but also because you are a good friend. You are now in Barcelona, probably enjoying good (?) food and a warm weather, but I'll always remember the time we spent on board games and BBQs.

Kenneth, one of my flat mates during my stay in Boston, I really enjoyed our conversations at dinner and our trip in Vermont with Rainer, as well as our re-union in Cologne. You showed me that even when people are coming from the opposite sides of the world they can still share the same values.

Antoine, Ya, Oskar, David, Lucie, Cristina, Dave, Mark, Elena, and Eliana, you are all wonderful people I had the pleasure to meet during my PhD. I had a very nice time with you guys and I hope our roads will meet again.

Rolf, this paragraph is for you. You have become my partner just at the end of my PhD period but this was the most crucial moment for me, cause I had to finish my work at the EMC, this thesis, and look for a new job. You have supported me in all of these hectic moments and managed to get a hold on me and bring me back to reality in those moments in which I lost my nerves. I believe that your kind and mature character is what I'll need also with my new position in Sweden, so I hope you can join me there soon.

Last, I want to thank all the important people I have left back in Italy when I started my PhD at the EMC.

Simone, we've been friends since primary school and we have both chosen to take the risky step in life to pursue an academic career, become PhD students and now, Post-docs. You are one of the people I left back in Italy that I miss the most, now more than ever since you are living overseas. Although the fact that you have a very strong personality and a self-confidence that I did not see even in one of our infamous prime ministers, I value you as my dearest friend.

Marco P, Marco DPr, Max P, Max B, Rosa, Federica and Daniele. You are truly wonderful friends that I have known since I was little and I find quite remarkable to be friends for such a long time. I know that to choose a scientific career outside of the country means that I can see you only few times a year. Even then, during these times in which we get back together it feels as I never left.

Finally, I want to spend a paragraph for my family, which has supported me throughout this important period and beyond.

My mother and my father, Giuliana and Loris, have always supported me during this period of my life in which I came in contact with a different culture than my own, away from Italian cuisine (trust me, you try to eat in Italian restaurant abroad you're gonna get poisoned) and from many dear people. I guess you had a hard time accepting to see me only a couple of times a year for few weeks tops. The two of you could not be thanked without acknowledging your wonderful partners, Roberto and Manuela.

To my sister Alice, which always loved me despite our fights, goes my deepest love and respect. You have always supported me in my decisions and, despite the fact that we do not see each other. Also, you have shown to have become an independent and self-confident woman. I am proud of having a sister like that!

Also, I would like to thanks my aunts Anna and Marina, my uncle Luigi, and my grandmothers Iva and Nada.

## **Publications:**

- **De Marchi T**, Timmermans AM, Sieuwerts AM, Smid M, Look MP, Grebenchtchikov N, Sweep FC, Smits JG, Magdolen V, Verhaegh W, van Deurzen CH, Foekens JA, Umar A, Martens JW. Phosphoserine aminotransferase 1 is associated to poor outcome on tamoxifen therapy in recurrent breast cancer. Submitted for publication.

- Liu NQ, **De Marchi T**, Timmermans AM, Trapman-Jansen A, Foekens R, Look MP, Smid M, van Deurzen CH, Span PN, Sweep FC, Benedicte-Brask J, Timmermans-Wielenga V, Foekens JA, Martens JW, Umar A. Prognostic significance of nuclear expression of UMP-CMP kinase in triple negative breast cancer patients. Submitted for publication.

- **De Marchi T**, Foekens JA, Umar A, Martens JW. Endocrine therapy resistance in estrogen receptor (ER) positive breast cancer. *Drug Discov Today*. 2016 May 24. pii: S1359-6446(16)30163-5. doi: 10.1016/j.drudis.2016.05.012. [Epub ahead of print] (this thesis)

- **De Marchi T**, Kuhn E, Dekker LJ, Stingl C, Braakman RB, Opdam M, Linn SC, Sweep FC, Span PN, Luider TM, Foekens JA, Martens JW, Carr SA, Umar A. Targeted MS Assay Predicting Tamoxifen Resistance in Estrogen-Receptor-Positive Breast Cancer Tissues and Sera. *J Proteome Res*. 2016 Apr 1;15(4):1230-42 (this thesis).

- **De Marchi T**, Liu NQ, Sting C, Smid M, Tjoa M, Braakman RB, Luider TM, Foekens JA, Martens JW, Umar A. Global proteomic characterization of microdissected estrogen receptor positive breast tumors. *Data Brief*. 2015 Oct 8;5:399-402. doi: 10.1016/j.dib.2015.09.034. eCollection 2015 Dec.

- **De Marchi T**, Timmermans AM, Smid M, Look MP, Stingl C, Opdam M, Linn SC, Sweep FC, Span PN, Kliffen M, van Deurzen CH, Luider TM, Foekens JA, Martens JW, Umar A. Annexin-A1 and caldesmon are associated with resistance to tamoxifen in estrogen receptor positive recurrent breast cancer. *Oncotarget*. 2016 Jan 19;7(3):3098-110. doi: 10.18632/oncotarget.6521 (this thesis).

- **De Marchi T**, Liu NQ, Stingl C, Timmermans MA, Smid M, Look MP, Tjoa M, Braakman RB, Opdam M, Linn SC, Sweep FC, Span PN, Kliffen M, Luider TM, Foekens JA, Martens JW, Umar A. 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Mol Oncol*. 2016 Jan;10(1):24-39. doi: 10.1016/j.molonc.2015.07.004. Epub 2015 Aug 7 (this thesis).

- **De Marchi T**, Kuhn E, Carr SA, Umar A. Antibody-based capture of target peptides in multiple reaction monitoring experiments. *Methods Mol Biol*. 2015;1293:123-35. doi: 10.1007/978-1-4939-2519-3\_7 (this thesis).

- Liu NQ, **De Marchi T**, Timmermans AM, Beekhof R, Trapman-Jansen AM, Foekens R, Look MP, van Deurzen CH, Span PN, Sweep FC, Brask JB, Timmermans-Wielenga V, Debets R, Martens JW, Foekens JA, Umar A. Ferritin heavy chain in triple negative breast cancer: a favorable prognostic marker that relates to a cluster of differentiation 8 positive

(CD8+) effector T-cell response. *Mol Cell Proteomics*. 2014 Jul;13(7):1814-27. doi: 10.1074/mcp.M113.037176. Epub 2014 Apr 17.

- Liu NQ, Stingl C, Look MP, Smid M, Braakman RB, **De Marchi T**, Sieuwerts AM, Span PN, Sweep FC, Linderholm BK, Mangia A, Paradiso A, Dirix LY, Van Laere SJ, Luiders TM, Martens JW, Foekens JA, Umar A. Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer. *J Natl Cancer Inst*. 2014 Feb;106(2):djt376. doi: 10.1093/jnci/djt376. Epub 2014 Jan 7.

- Liu NQ, Dekker LJ, Stingl C, Güzel C, **De Marchi T**, Martens JW, Foekens JA, Luiders TM, Umar A. Quantitative proteomic analysis of microdissected breast cancer tissues: comparison of label-free and SILAC-based quantification with shotgun, directed, and targeted MS approaches. *J Proteome Res*. 2013 Oct 4;12(10):4627-41. doi: 10.1021/pr4005794. Epub 2013 Sep 13.

## PhD portfolio

Name PhD student	Tommaso De Marchi
Erasmus MC Department	Medical Oncology
Research School	Molecular Medicine
Supervisor	Prof. dr. John A. Foekens
Co-supervisors	Dr. Arzu Umar Dr. ir. John W.M. Martens

## PhD training

<b>Courses</b>	<b>Year</b>	<b>ECTS</b>
Biomedical Research Techniques	2011	1.60
Introduction into Clinical and Fundamental Oncology	2011	1.40
Workshop on Photoshop and Illustrator	2011	0.30
Basic Introduction Course on SPSS	2011	0.60
Molecular Medicine Course	2012	0.70
Workshop on presenting skills for junior researchers	2012	1.00
MaxQuant Summer School	2012	3.00
Course on R	2016	1.40

<b>Seminars</b>	<b>Year</b>	<b>ECTS</b>
Monthly JNI Oncology Lectures	2011-2016	1.00
JNI Scientific Meetings	2011-2016	2.50
Annual Medical Oncology Meeting	2011-2016	2.50
Annual Molecular Medicine Day	2012-2015	2.50
Cancer Genomics Center Annual Meeting	2011-2015	1.20
Netherlands Proteomics Center Annual Meeting	2012-2015	1.80
Netherlands Proteomic Platform Annual Meeting	2013, 2014	1.00
Daniel den Hoed Day	2016	0.30

<b>Teaching activities</b>	<b>Year</b>	<b>ECTS</b>
3 months supervision of Bachelor student	2012	4.00
12-months supervision of Master student	2014-2015	16.00

<b>Presentations</b>	<b>Year</b>	<b>ECTS</b>
Journal Club (Oral)	2011-2015	2.00
Annual Molecular Medicine Day (Poster)	2012-2015	2.00
Annual Netherlands Proteomics Center Meeting (Poster)	2012-2015	2.00
JNI Scientific Meetings (Oral)	2013, 2014, 2015	3.00
KWF Tumor cell biology Annual Meeting (Oral)	2011, 2012, 2014	3.00
MaxQuant Summer School (Poster)	2012	0.50

Proteomic Forum Berlin 2013 (Poster)	2013	0.50
America Association for Mass Spectrometry Annual Meeting (Poster)	2013	0.50
American Association for Cancer Research Annual Meeting (Poster)	2014	0.50
EORTC progress meeting (Oral)	2014	1.00
Netherlands Proteomic Platform Annual Meeting (Oral)	2014	1.00
Annual Medical Oncology Meeting (Oral)	2015	1.00
Cancer Genomics Center Annual Meeting (Poster)	2015	0.50
Daniel den Hoed Day (Oral)	2016	1.00

**International Conferences**

	<b>Year</b>	<b>ECTS</b>
EORTC progress meeting	2011, 2015	1.00
Proteomic Forum Berlin 2013	2013	1.00
America Association for Mass Spectrometry Annual Meeting	2013	1.00
American Association for Cancer Research Annual Meeting	2014	1.00

**Total ECTS** **65.3**

## Curriculum Vitae

Tommaso De Marchi was born on October 17<sup>th</sup> 1986 in San Vito al Tagliamento, Italy. He obtained a Bachelor degree in Health Biotechnologies from the University of Padua in 2008, with a thesis on HIV mother-to-child transmission. In 2010, Tommaso obtained a Master Degree in Medical Biotechnologies from the same institution. During his master degree final internship, Tommaso had the chance to glimpse at the field of proteomics by using mass spectrometry to analyze the fragmentation patterns of disulfide bonds. This experience sparked his interest in the field of proteomics applied to biology and medicine and ultimately led him to the Netherlands, where he joined the laboratory of Breast Cancer Genomics and Proteomics at the Erasmus Medical Center. Here he worked on breast cancer tamoxifen resistance biomarker discovery using mass spectrometry (MS)-based approaches. During his PhD period, Tommaso developed and validated a 4-protein classifier predicting first line tamoxifen therapy outcome in recurrent estrogen receptor positive breast cancer. Subsequently, in collaboration with the group of Dr. Steven A. Carr at the Broad Institute of MIT and Harvard (Cambridge, MA, USA) – which he visited in 2013 – Tommaso developed a targeted MS method to measure the 4-protein classifier in both breast cancer patient derived tissues and sera. Further studies led Tommaso to investigate additional tamoxifen resistance markers representative of resistant tumor subgroups. His PhD training resulted in the publication of two articles in the field of oncology (De Marchi et al., *Mol Oncol*, 2016; De Marchi et al., *Oncotarget*, 2016), a method chapter (De Marchi et al., *Methods Mol Biol*, 2015), and two articles in the field of proteomics (De Marchi et al., *J Proteome Res*, 2016; De Marchi et al., *Proteomics*, 2016), 1 review article (De Marchi et al., *Drug Discov Today*, 2016), and 1 manuscript in preparation. Tommaso also contributed to other publications in both the field of oncology (Liu et al., *J Natl Can Inst*, 2014) and proteomics (Liu et al., *J Proteome Res*, 2013; Liu et al., *Mol Cell Proteomics*, 2014) for a total of 8 published articles.

