Phosphoproteomics for the Mapping of Altered Cell Signaling Networks in Breast Cancer

Olga Villamar-Cruz and Luis E. Arias-Romero

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/53109

1. Introduction

Breast cancer is the most commonly diagnosed cancer in women worldwide and consequently has been extensively investigated in terms of histopathology, immunochemistry and familial history [1]. Fortunately, technological advances have enabled characterization of the molecular subtypes of breast cancer [2, 3] and this in turn has facilitated the development of molecularly targeted therapeutics for this disease.

Profiling breast cancer with expression arrays has become common, and it has been suggested that the results from early studies will lead to understanding the molecular differences between clinical cases and allow individualization of care. Breast cancer may now be subclassified into luminal, basal, and ErbB2/HER2 subtypes with distinct differences in prognosis and response to therapy. These groups of tumors confirmed long-recognized clinical differences in phenotype, but added new knowledge regarding breast cancer biology. For example, the gene expression profiling revealed that within the estrogen receptor (ER)-positive tumors at least two subtypes, luminal A and luminal B, could be distinguished that vary markedly in gene expression and prognosis [3]. Conversely, hormone receptor–negative breast cancer comprised two distinct subtypes, the ErbB2 subtype and the basal-like subtype [3, 4]. These subtypes differ in biology and behavior, and both show a poor outcome. Importantly a very similar classification of breast cancers has now been characterized using immunohistochemistry to analyze patterns of protein expression in tumor sections and suggesting that a few protein biomarkers can be used to stratify breast cancers into different groups that can be mapped to the subtypes outlined below [5-8].

Luminal breast cancers are the most common subtype of breast cancer. The luminal subtypes make up the hormone receptor–expressing breast cancers, and have expression patterns reminiscent of the luminal epithelial component of the breast [2]. These patterns
include expression of luminal cytokeratins 8/18, ER and genes associated with ER activation such as LIV1 and CCND1 (also known as cyclin D1) [2, 9]. Fewer than 20% of luminal tumors have mutations in TP53, and these tumors are often grade I [3, 9]. Within the luminal cluster there are at least two subtypes, luminal A and luminal B. Although both are hormone receptor expressing, these two luminal subtypes have distinguishing characteristics. Luminal A has, in general, higher expression of ER-related genes and lower expression of proliferative genes than luminal B [3, 4].

The basal-like subtype of breast cancer was so named because the expression pattern of this subtype mimicked that of the basal epithelial cells of other parts of the body and normal breast myoepithelial cells [2]. These similarities include lack of expression of ER and related genes; low expression of ErbB2; strong expression of basal cytokeratins 5, 6, and 17; and expression of proliferation-related genes [2, 9]. Immunohistochemical profiling using tissue microarrays has identified that a group of tumors characterized by basal cytokeratin expression are also characterized by low expression of BRCA1 [10]. Basal-like tumors are more likely to have aggressive features such as TP53 mutations and a markedly higher likelihood of being grade III (P < 0.0001) than luminal A breast cancers (P < 0.0001) [3].

Finally, the other breast cancer subtype that has been identified is distinguished by amplification of the gene encoding the human epidermal growth factor receptor 2 (ErbB2/HER2). The human ErbB/HER receptor family comprises four tyrosine kinase receptors (HER1/ErbB1, also termed the epidermal growth factor receptor (EGFR), HER2/ErbB2, HER3/ErbB3, and HER4/ErbB4) that play important roles in the progression of various types of cancers, including breast, prostate, and colon cancer [11]. Deregulation of ErbB receptor signaling leads to enhanced cell proliferation, migration, and malignant transformation. Overexpression, amplification, or mutation of the ERBB2 gene occurs in approximately 20–30% of invasive breast cancers, and is associated with disease progression, poor prognosis, increased risk of metastases and shorter overall survival [12].

ErbB2-mediated signal transduction is believed to depend largely on heterodimerization with EGFR or ErbB3, and these heterodimers activate a signaling program that drives cell proliferation, resistance to apoptosis, loss of polarity, and increased motility and invasiveness [13, 14]. Trastuzumab is a humanized monoclonal antibody targeted against the extracellular portion of ErbB2. This is the first ErbB2-targeted agent to be approved by the United States Food and Drug Administration (FDA) for the treatment of both early stage and metastatic ErbB2-overexpressing (ErbB2 positive) breast cancers [15, 16]. Subsequently, lapatinib, an orally bioavailable small molecule dual ErbB2- and EGFR/HER1-specific tyrosine kinase inhibitor (TKI), received FDA approval in combination with capecitabine for patients with advanced ErbB2 positive breast cancer [17].

Although ErbB2-targeted therapies have had a significant impact on patient outcomes, resistance to these agents is common. In clinical trials, 74% of patients with ErbB2 positive metastatic breast cancer did not have a tumor response to first-line trastuzumab monotherapy [18] and 50% did not respond to trastuzumab in combination with chemotherapy [15]. These examples illustrate the problem that inherent (de novo) resistance...
to ErbB2-targeted agents poses for effective treatment of ErbB2 positive breast cancer. Moreover, only approximately one-quarter of patients with ErbB2 positive metastatic breast cancer who were previously treated with trastuzumab achieved a response with lapatinib plus capecitabine [17]. These limitations have led to efforts to better understand the underlying cellular networks that confer resistance to these agents in order to better select patients who are most likely to benefit from specific therapies and to develop new agents that can overcome resistance.

The goal of this review is to give a concise overview of current approaches in the field of phosphoproteomics and to show how a combination of several approaches can be used to obtain a more comprehensive understanding of a given signaling pathway. A number of proteomic approaches have been developed over the years to identify aberrantly activated kinases and their downstream substrates. Most often, phosphorylation is used as a surrogate for monitoring kinase activity in cells. In the past, kinases and their activities were generally studied on an individual basis using biochemical approaches. However, technological advances in the recent past have led to development of several high-throughput strategies to study the phosphoproteome. High-throughput technologies for monitoring phosphorylation events include array-based technologies such as peptide arrays [19-21], antibody arrays [22] and mass spectrometry [23, 24]. Quantitative phosphoproteomic profiling allows researchers to investigate aberrantly activated signaling pathways and therapeutic targets in cancers. Finally, phosphoproteomic approaches can not only assist in determining the appropriate therapeutic targets but also elucidate mechanisms such as off-target effects resulting from binding of inhibitors to unintended kinases/non-kinase proteins. Here, we will discuss some of the popular approaches to characterize the kinome and the phosphoproteome along with illustrative examples where such approaches have been employed for global analysis of breast cancer.

2. Challenges of phosphoproteomics

Phosphoproteomic analysis is plagued by the same challenges facing all proteomic experiments: complexity, dynamic range, and temporal dynamics. The true complexity of the phosphoproteome has yet to be determined, but the Phosphosite database (http://www.phosphosite.org) now lists 30 000 phosphorylation sites on 17 000 proteins, and this number is steadily increasing as each large-scale phosphorylation analysis continues to identify a large number of novel sites. With so many of the proteins in the cell being phosphorylated, the dynamic range of the phosphoproteome is similar to that of the proteome (i.e., $1 \times 10^9$), but is further increased by substoichiometric modification. In addition, the temporal dynamics of protein phosphorylation regulate the rapid activation and deactivation of cellular signaling networks, further complicating analysis of the phosphoproteome. So the challenge is not simply to identify and catalog all of the phosphorylation sites, but rather to identify the site, quantify the stoichiometry, and monitor the temporal change in phosphorylation in response to a variety of cellular perturbations. Performing this task on a large number of phosphorylation sites across a broad swath of the signaling network is especially challenging, but is required to understand the mechanisms by which protein phosphorylation controls cell biology.
3. Mass Spectrometry (MS)-based approaches

Currently, the most powerful tool to interrogate the phosphoproteome is enrichment for phosphopeptides followed by reverse-phase liquid chromatography combined with tandem mass spectrometry (LC-MS/MS). When sample preparation and instrumentation are chosen appropriately, thousands of phosphorylation sites can be identified (Figure 1). Some research groups have already taken advantage of these methodologies for identifying proteins that could be useful therapeutic targets or novel molecular markers in breast cancer specimens. Many of these analyses have focused on tyrosine phosphorylation profiles due to the fact that approximately half of the tyrosine kinase complement of the human kinome is implicated in human cancers [4], and provides important targets for cancer treatment, as well as biomarkers for patient stratification. Recently, Chen et al. adapted LC-MS/MS technology to assess the tyrosine phosphorylation profile in the MCF10AT model of breast cancer progression [25]. This study identified and validated seven proteins, termed SPAG9, CYFIP1, RPS2, TOLLIP, SLC4A7, WBP2, and NSFLC1, to be authentic tyrosine kinase substrates. In addition, SPAG9, WBP2, TOLLIP, and NSFL1C were demonstrated to be authentic tyrosine phosphorylation targets of EGFR signaling, and differential expression of TOLLIP and SLC4A7 was subsequently validated in clinical breast cancer samples. Consistent with the MCF10AT model, more than 30% of the human breast cancer samples analyzed in this study displayed reduced expression of SLC4A7 compared with normal tissues. In contrast, only 25% of the samples showed increased levels of TOLLIP when normal cells become cancerous. Moreover detection of aberrant expression of TOLLIP and SLC4A7 in pre-neoplastic lesions suggests that they represent potential biomarkers that could complement mammography and histopathology for screening and early detection of breast cancer [25].

Most recently, a number of reports have demonstrated the importance of EGFR signaling in breast cancer [26-28]. Hochgrafe et al. characterized the tyrosine kinase signaling networks associated with different breast cancer subgroups [27]. By using this approach in a panel of 15 different breast cancer cell lines, the authors identified 544 phototyrosine sites in peptide sequences derived form 295 non redundant proteins, interestingly, 31 of these are novel tyrosine phosphorylation sites. Upon unsupervised hierarchical clustering using data for all tyrosine phosphorylated proteins, the 15 cell lines were clustered into two groups previously characterized as “basal” or “luminal” by transcript profiling [29]. Increased phosphorylation of several tyrosine kinases (i.e. Met, Lyn/Hck, EphA2, EGFR, and FAK) was characteristic of basal lines. In contrast, IGF1R/INSR, ErbB2, and ACK1 exhibited increased phosphorylation in luminal breast cancer cells. For all of the differentially phosphorylated kinases, increased phosphorylation was detected on sites that positively regulate kinase activity and downstream signaling. For example, Met Y1234, Lyn Y397, and FAK Y577 are activation loop sites [30], and phosphorylation of Y588 and Y594 in the juxtamembrane region of EphA2 is required for kinase activity [31]. In the case of EGFR and ErbB2, differential phosphorylation was predominantly on sites in the COOH-terminal tail that promote activation of the Ras/Raf/MEK/ERK pathway [32, 33]. A deeper analysis of the tyrosine phosphoproteome revealed a signature that characterizes the basal phenotype, and
identified a prominent Src family kinase (SFK) signaling network in basal breast cancer cells that extends not only downstream to canonical SFK substrates regulating cell adhesion and migration but also upstream to specific RTKs such as EGFR, ErbB2 and Met among others. Subsequent functional analyses determined that SFKs transmit pro-proliferative, pro-survival and pro-mitogenic signals in these cells, and that Lyn is an important regulator of cell invasion. In addition, SFKs promoted tyrosine phosphorylation of specific RTKs in these cells, and this may attenuate cellular sensitivity to therapies directed against these receptors. Consequently, these findings provide important insights into the biology of basal breast cancers and have significant implications for the development of therapeutic strategies that target this subtype of breast cancer [27].

A very elegant study performed by Zhang et al. analyzed the EGF induced protein phosphorylation events in the Human Mammary Epithelial Cell (HMMC) 184A1 [26]. In this report, a time course phosphorylation profile of 78 tyrosine phosphorylation sites on 58 proteins was generated. For each phosphorylation site, a quantitative temporal phosphorylation profile was generated by comparing the relative ratios of peak areas for the iTRAQ marker ions in the MS/MS spectrum. Of the 58 proteins identified in this analysis, 52 have been already associated with the EGFR signaling network, whereas the other six proteins have not been previously identified in either proteomic or biochemical analyses of EGFR signaling. Contained in this group are phosphorylation sites on hypothetical protein FLJ00269, hypothetical protein FLJ21610, target of myb1-like 2 protein, and chromosome 3 open reading frame 6. In addition to the six proteins that had not been previously characterized in the EGFR signaling network, the authors also identified several novel phosphorylation sites on proteins known to be in the network. The bioinformatic analysis of the data generated by this method self-organize into clusters of phosphorylation sites that correlate with well known signaling nodes reported in the literature (i.e. the Ras/Raf/MEK/ERK and PI3K/AKT signaling pathways). In a related study, the same research group analyzed the EGF- and heregulin (HRG)-induced protein phosphorylation events that control cell migration and proliferation in the context of ErbB2 overexpression in HMMCs [34]. As a result of these analyses, 332 phosphorylated peptides from 175 proteins were identified, including 289 singly (tyrosine) phosphorylated peptides, 42 doubly phosphorylated peptides (21 tyrosine/tyrosine, 18 serine/tyrosine, and three threonine/tyrosine), and one triply phosphorylated peptide (tyrosine/tyrosine/tyrosine). A total of 20 phosphorylation sites were identified on EGFR, ErbB2, and ErbB3, including nine tyrosine and two serine sites on EGFR, eight tyrosine phosphorylation sites on ErbB2, and one tyrosine phosphorylation site on ErbB3. Of the 20 phosphorylation sites on EGFR family members, Y1114 on EGFR and Y1005 and Y1127 on ErbB2 represent novel sites that have not been previously described in the literature. To correlate signals with cell response, the authors also quantified proliferation and migration rates for these same cell states and stimulation conditions. Phenotypically, ErbB2 overexpression promoted increased cell migration, but had minimal effect on cell proliferation. More specifically, EGF stimulation of ErbB2-overexpressing cells promoted migration by the phosphorylation of proteins from multiple pathways (e.g., PI3K, MAPK, catenins, and FAK), whereas HRG stimulation of ErbB2-overexpressing cells activated only a very specific subset of proteins in the canonical
migration pathway, in particular FAK, Src, paxillin, and p130Cas. In contrast, proliferation was primarily driven by EGF stimulation, and was not affected by ErbB2 expression levels [34]. Finally, Kumar et al. significantly extend their previous analysis of ErbB2-mediated signaling and cell function by using a model that predicts ErbB2 effects on HMMCs behavior by using MS phosphotyrosine data sets [28]. The results of this research showed that ErbB2 overexpression in the presence of EGF, as discussed above, produced interesting signal network changes and increased cell migration but did not affect cell proliferation [34]. These findings both highlight previously identified elements in the ErbB2 signaling network, and suggest new pathways and targets critically implicated in ErbB2-mediated signaling and its effect on migration and proliferation.

Although MS has proven to be an extraordinary tool for protein characterization, measurement of peptide intensities alone does not immediately provide quantitative information. There are several approaches to overcome this problem. Stable isotopes are incorporated either by metabolic labeling, as in the SILAC (stable isotope labeling with amino acids in cell culture) method, or by chemical derivatization (Figure 1) [35]. SILAC relies on metabolic incorporation of an isotopically labeled amino acid. Two groups of cells are grown in culture media that are identical except in one respect: the first media contains the “light” and the other a “heavy” form of a particular amino acid (for e.g. L-leucine or deuterated L-leucine). Through the use of special cell culture medium lacking the modified amino acids, the cells are forced to use the particular labeled or unlabeled form of the amino acid previously added to the medium. In each cell doubling, the cell population replaces at least half of the original form of the amino acid, eventually incorporating 100% of a given light or heavy form of the amino acid. A variety of amino acids are suitable in SILAC, including arginine, leucine, lysine, serine, methionine and tyrosine. The different cell line conditioned media can then be combined and run together in a single MS run. The advantages of SILAC include the fact that the labeling process is highly efficient, it does not require additional purifications to remove excess labeling reagent, nor does it involve multi-step labeling protocols and the sample preparation bias introduced by the comparison of two separate preparation steps is avoided. As well, SILAC allows the experimenter to use any method of protein or peptide purification (after enzymatic digestion) without introducing error into the final quantitative analysis. In one study, SILAC was utilized to examine differential membrane expression between normal and malignant breast cancer cells [36]. Approximately 1,000 proteins were identified with more than 800 of these proteins being classified as membrane or membrane-associated. Although the majority of the proteins remained unchanged when compared with the corresponding normal cells, a number of proteins were found upregulated or down-regulated by greater than 3-fold.

A few years ago, Bose et al. described a quantitative proteomic analysis to study ErbB2 signaling by using SILAC in 3T3 cells ectopically expressing ErbB2 [37]. By using this methodology, the authors identified a panel of 198 proteins that displayed increased phosphorylation levels and a group of 81 proteins that showed decreased phosphorylation levels merely by ErbB2 overexpression. The list of proteins that showed high phosphorylation levels included several well known ErbB2 downstream effectors and
modulators of pro-survival, anti-apoptotic and proliferative pathways, such as PLCγ1, the regulatory and catalytic subunits of PI3K (p85β, p85α, and p110β), the Src family member Fyn, RasGAP, and HSP90. Importantly, several known EGFR signaling proteins, which had not been previously implicated in ErbB2 signaling, were also identified, including Stat1, Dok1, and δ-catenin. The 81 proteins that displayed decreased phosphorylation levels in 3T3-ErbB2 cells included FAK, p130-Cas/BCAR1, and caveolin 1 among others. In this study, the effect of the EGFR and ErbB2 selective tyrosine kinase inhibitor (TKI), PD168393, was also quantified, the results showed that 83 of the 198 proteins that displayed increased phosphorylation when ErbB2 was overexpressed were inhibited by 100 nM of PD168393 (>1.5-fold), and 27 proteins showed a smaller degree of inhibition (1.3- to 1.5-fold), suggesting that 110 of these 198 proteins are affected by this TKI. Under these conditions, 79 proteins were not affected by PD168393, including Fyn and three subunits of PI3K. This observation raises the question of whether different arms of the ErbB2 signaling pathway have differential inhibitor sensitivity. To validate the relevance of these proteins to ErbB2 signaling in a more realistic setting, the authors used the ErbB2 positive breast cancer cell line BT-474. As expected, PD168393 also inhibited the phosphorylation of PLCγ1 and Stat1 in BT-474 cells, supporting the idea that phosphoproteins identified by performing SILAC on 3T3-ErbB2 cells may be applicable to other ErbB2-overexpressing cell lines.

Although SILAC has proven to be a very powerful method to dissect signaling in tumor cell lines, metabolic labeling has a major limitation. Whereas proteins in cultured cells can be readily labeled, those in living organisms cannot. Approaches have been developed to metabolically label worms, flies [38] and even mice [39] and rats [40], but human tissues have to this day remained 'unlabelable'. When applying proteomics to tumor biology, it is imperative to quantify a representative number of proteins, to obtain reproducible results and to study cancer-relevant proteins of low abundance. Ishihama et al. have tried to solve this problem by adding labeled cultured cells to the tissue samples [41]. However the comparison of a single cell line with a whole tissue context has several limitations. More recently, Geiger et al. mixed labeled protein lysates from several previously established cancer-derived cell lines, which together are more representative of the full complexity of a tissue proteome than a single cell line, thereby increasing accuracy [42]. Initially, they SILAC-labeled the breast cancer cell line HCC1599 and mixed the lysate with the lysate of mammary carcinoma tissue from an individual with grade II lobular carcinoma. Although they were able to quantify 4,438 proteins at least once in triplicate analysis, the ratio distribution was broad and bimodal, containing 755 proteins with more than fourfold higher expression in the tumor compared to the cell line. Next, they selected four breast cancer cell lines differing in origin, stage, ER and ErbB2 expression; and this superset of SILAC-labeled cell lines that more accurately representing the tissue was used for further analysis. The comparison of the tumor proteome with this “super-SILAC” mix, drastically improved the quantification accuracy. The distribution was unimodal and 90% of quantified proteins were within a fourfold ratio between the tumor and the super-SILAC mix (3,837 of 4,286 quantified protein groups). Furthermore, the quantitative distribution was much narrower, with 76% of the proteins in the carcinoma and the super-SILAC mix differing by only twofold or less. Although super-SILAC has not been used to analyze the tumor...
phosphoproteome yet, the results of this research accurately quantified more than a hundred protein kinases despite their low abundance. Among them were ErbB2, EGFR, AKT, Pak1 and Pak2 and nine members of the MAPK cascade, all representing pathways central to malignancy. At first view, this new method has great potential to expand the use of accurate relative proteomic quantitation methods to study molecular aspects of tumor biology and perhaps as a tool for candidate biomarker discovery, so it is conceivable that it will likely become a valuable tool for understanding the molecular and mechanistic aspects of phosphorylation in tumor samples.

As described above, quantitative MS-based phosphoproteomics has been applied to identify oncogenic kinases which may serve as potential drug targets. To validate this hypothesis, cells are often treated with selected kinase inhibitors with the goal of altering cellular phenotype, but it is often difficult to establish whether the effect was due to on or off-target effects of the compound. In order to determine the mechanism of action, it may be necessary to quantify the specificity of the inhibitor. Two groups have pioneered the use of immobilized kinase inhibitors with broad specificity to enrich a substantial subset of protein kinases from total cell lysates followed by quantitative mass spectrometry. Daub et al. developed a kinase inhibitor pull-down technique in combination with phosphoproteomics to map and quantify more than one thousand phosphorylation sites on human protein kinases arrested in S- and M-phase of the cell cycle [43]. Researchers at Cellzome employed Kinobeads™ to enrich protein kinases and then performed competition-based assays using specific kinase inhibitor drugs such as imatinib (Gleevec), dasatinib (Sprycel) and bosutinib in BCR-Abl positive K562 cells [44]. Recently, Zhang et al. modified this approach in order to develop more potent inhibitors of the kinase AXL, which has an important role in mediating breast cancer cell motility and invasivity [45]. In this study, the authors used a chemical library of kinase inhibitors in order to identify small molecular inhibitors with selective activity on the AXL tyrosine kinase, the chemical compound NA80x1 which has previously been reported to have inhibitory activity against Src kinase [46], inhibited AXL kinase activity in a dose-dependent manner, with an IC50 of 12.67 ± 0.45 μmol/L. Then, NA80x1 and a structurally similar, but much more potent inhibitor of Src and Abl kinases termed SKI-606, were chemically modified and attached to an affinity purification resin. To identify the specific targets (and some other off-targets) of these inhibitor derivatives, SILAC labeled proteins from the breast cancer cell line Hs578T were used for in vitro association experiments with the immobilized chemical compounds. The protein eluates from the respective affinity purifications were mixed and digested, and the resulting peptide fractions were analyzed by MS. In total, 146 different proteins were identified with at least two unique peptides in the MS experiments. Among them, 43 proteins were found to specifically bind to the immobilized compounds and 32 were kinases. In addition to known targets such as Src/Abl family kinases Src, Lyn, Arg, and the RTK AXL, which was functionally characterized as a cellular target in this study, a variety of other inhibitor-interacting proteins were identified, including eight more tyrosine kinases (such as FAK and four Eph receptor kinase family members) as well as nine members from the STE group of kinases involved in mitogen-activated protein kinase (MAPK) signaling (including six MAP4K/STE20 kinase family members and two MAP2K family members). This study is a
clear example of how MS can help to identify off-targets of small molecular kinase inhibitors in order to develop more specific and potent chemicals for cancer therapies.

**Figure 1.** Mass Spectrometry based approaches. The upper panel shows the pipelines of a prototypical proteomics experiment. Proteins are extracted from a biopsy or tumor sample and digested with trypsin to obtain peptides. The resulting peptides are resolved by reverse phase liquid chromatography (LC) and subsequently, analyzed by tandem mass spectrometry (MS/MS). Finally, the matched peptides allow the identification of the proteins using databases. The lower panel shows the schematic outline of the SILAC method. Separate cultures of cells are grown in normal medium ($^{12}$C$_6$-arginine) or in medium containing arginine labeled at all six carbons with $^{13}$C ($^{13}$C$_6$-arginine). The cells in normal medium are left unstimulated whereas cells in the $^{13}$C-arginine medium are stimulated with an agent that activates signaling. The cells are harvested and equal amounts of lysate protein mixed together. In most cases, steps to enrich phosphoproteins and/or phosphopeptides after trypsin digestion are needed to detect low-abundance phosphopeptides. The peptides are resolved by LC-MS/MS and the data are used for automated database searching to identify peptides (and their corresponding protein) and to detect phosphopeptides.
4. Protein microarray approaches (non-MS)

To monitor previously identified phosphorylation sites, the combination of phosphospecific antibodies and western blotting has been the gold standard. However, until recently the limited throughput of this approach, with only one phosphorylation site investigated at a time, has driven the development of other, high-throughput approaches.

Arrays using phosphospecific antibodies to investigate phosphorylation sites have been developed [47, 48] and used to interrogate dozens of phosphorylation sites simultaneously [49]. As this technology requires antibodies with high-affinity and specificity, currently only a limited number of phosphorylation sites can be analyzed [50]. However, further development might lead to an even broader application of microarray technology for phosphoprotein studies.

Protein microarray formats can be divided into two major classes: forward phase arrays and reverse phase arrays (Figure 2) [51]. In a forward phase array, each spot contains one type of immobilized capture molecule, usually an antibody. Each array is incubated with one test sample containing the protein mixture of interest.

**Figure 2. Protein microarray platforms.** Forward phase arrays (top) immobilize a bait molecule such as an antibody designed to capture specific biotinylated proteins representing a specific treatment or condition. In this specific case, the bound analytes are detected by fluorescently labeled biotin. Reverse phase arrays immobilize the test sample analytes on the solid phase. An analyte specific labeled ligand (e.g., antibody; lower left) is applied in solution phase. Bound antibodies are detected by signal amplification (lower right).
sample such as a cellular lysate or serum sample representing a specific treatment condition, and multiple analytes from that sample are measured simultaneously. In contrast, the reverse phase array format immobilizes an individual test sample in each array spot, in a way that this array is comprised of hundreds of different patient samples or cellular lysates. In the reverse phase array format, each array is incubated with one detection protein (e.g., antibody), and a single analyte endpoint is measured and directly compared across multiple samples [47, 51-55].

5. Forward phase protein arrays

The most popular class of forward phase protein arrays in cancer research is the antibody array. A common application of antibody arrays is the identification of biomarkers or molecules that are potentially valuable for diagnosis or prognosis or as surrogate markers of drug response. The multiplex capability of antibody arrays allows the efficient screening of many marker candidates to reveal associations between proteins and disease states or experimental conditions. Multiplexed measurements also allow the evaluation of the use of multiple markers in combination. The use of combinations of proteins for disease diagnostics may produce fewer false positive and false negative results as compared with tests based on single proteins. Antibody microarrays, by increasing the number of proteins that can be conveniently measured in clinical samples, could more significantly take advantage of the benefit of using combined markers in diagnostics. Other example applications of antibody microarrays in cancer research are to evaluate the coordinated changes of members of signaling pathways or to measure changes in expression levels of a class of proteins, such as angiogenesis factors.

Only a few studies using antibody arrays for breast cancer research have been reported. One of the first studies was performed by Hudelist et al., who employed a high-throughput protein microarray system which contains 378 well characterized monoclonal antibodies printed at high density on a glass slide in duplicate in order to compare the gene expression pattern of malignant and adjacent normal breast tissue in a patient with primary breast cancer [56]. Using this technique, the authors identified a number of proteins that show increased expression levels in malignant breast tissues such as casein kinase 1ε, p53, annexin XI, CDC25C, eIF-4E and MAP kinase 7. The expression of other proteins, such as the multifunctional regulator 14-3-3ε was found to be decreased in malignant breast tissue, whereas the majority of proteins remained unchanged when compared to the corresponding non-malignant samples. Moreover, the protein expression pattern was corroborated by immunohistochemistry, in which antibodies against 8 representative proteins known to be involved in carcinogenesis were employed in paraffin-embedded normal and malignant tissue sections deriving from the same patient. In each case, the results obtained by IHC matched the data obtained by antibody microarray system. In another report [57], 224 antibodies revealed proteins that are related to doxorubicin therapy resistance in breast cancer cell lines. A decrease in the expression of MAP kinase-activated monophosphotyrosine, cyclin D2, cytokeratin 18, cyclin B1 and heterogeneous nuclear ribonucleoprotein m3-m4 was found to be associated with doxorubicin resistance. Other
recent investigations helped identify a marker involved in invasion (interleukin (IL)-8) [58]. Studying the serum proteome from metastatic breast cancer patients and healthy controls with recombinant single-chain variable fragment (scFv) microarrays [59], breast cancer was identified with a specificity and sensitivity of 85% on the basis of 129 serum analytes.

Although a number of companies have already developed phospho-antibody arrays for breast cancer research, there are only a few reports of the use of this technology in breast cancer. In 2008, Eckestein et al. [60], studied the cellular mechanisms of resistance to cisplatin using MCF-7 cells as a model system. Cisplatin-resistant MCF-7 breast cancer cells were selected by exposure to sequential cycles of cisplatin that mimic the way the drug is used in the clinic. To investigate the phosphorylation status of the EGFR receptor family, a phosphoreceptor tyrosine kinase (phospho-RTK) array was used. In this assay, monoclonal capture antibodies, specific for a variety of RTKs, were spotted in an array format, and phosphorylation of EGFR family members was subsequently detected by a pan anti-phosphotyrosine antibody conjugated to horseradish peroxidase. In nonresistant cells the EGFR was phosphorylated at a low level. In contrast, in cisplatin resistant MCF-7 cells both the EGFR and ERBB2 receptors were strongly phosphorylated. The phospho-RTK array detected very low ErbB3 and ErbB4 phosphorylation in both MCF-7 and cisplatin resistant MCF-7 cells, suggesting, that these receptor subtypes are not activated in cisplatin-resistant breast cancer cells. By using similar arrays, the authors examined the Ras/Raf/MEK/ERK, PI3K/AKT, JNK and p38 signaling pathways, which are downstream effectors of EGFR in a number of cell systems. The analysis of these pathways showed that the Ras/Raf/MEK/ERK and PI3K/AKT pathways are hyperactive in the cisplatin-resistant breast cancer cells, whereas the JNK and p38 pathways were not affected. Similarly, this study shows that cisplatin-resistant breast cancer cells have an inactivation of the p53 pathway and display high levels of BCL-2. A transcriptional profile of the cisplatin-resistant breast cancer cells also showed that these cells have an upregulation of the amphiregulin gene, the expression and secretion of this protein is also elevated and this mechanism creates an autocrine loop that confers resistance to cisplatin.

A more recent study using this technology showed that activation of the PI3K-AKT pathway in tumors is modulated by negative feedback, including mTORC1-mediated inhibition of upstream signaling [61]. The authors clearly demonstrate that AKT inhibition induces the expression and phosphorylation of multiple receptor tyrosine kinases in a panel of different breast cancer cell lines. The results of this research suggest that receptor activation of PI3K-AKT causes AKT-dependent phosphorylation of FOXO proteins, which downregulate the expression of some of the receptors that are tightly coupled to PI3K, including ErbB3, IGF1R, and IR. In addition, AKT activation leads to activation of TORC1 and S6K, which feedback inhibits IRS1 expression and other non defined regulators of receptor signaling, resulting in down modulation of the signaling pathway. Thus, AKT inhibition will result in activation of FOXO-dependent transcription of receptors and inhibition of S6K-dependent inhibition of signaling with resultant activation of multiple receptors. The downstream effects of AKT will be suppressed, but other RTK-driven signaling pathways will be activated. In contrast, TORC1 inhibition blocks S6K-dependent feedback, activates IGF and ErbB kinases, but not their expression, and, thus, activates both AKT and ERK signaling. These findings have important basic and therapeutic implications.
6. Reverse phase protein arrays

Probing multiple arrays spotted with the same lysate concomitantly with different phosphospecific antibodies provides the effect of generating a multiplex readout. The utility of reverse phase protein microarrays lies in their ability to provide a map of known cell signaling proteins. Identification of critical nodes, or interactions, within the network is a potential starting point for drug development and/or the design of individual therapy regimens [62, 63]. The array format is also amenable to extremely sensitive analyte detection with detection levels approaching attogram amounts of a given protein and variances of less than 10% [51, 64]. Detection ranges could be substantially lower in a complex mixture such as a cellular lysate; however, the sensitivity of the reverse phase arrays is such that low abundance phosphorylated isoforms can still be measured from a spotted lysate amount of less than 10 cell equivalents. This level of sensitivity combined with analytical robustness is critical if the starting input material is only a few hundred cells from a biopsy specimen. Due to all this advantages, the reverse phase protein array has demonstrated a unique ability to analyze signaling pathways using small numbers of cultured cells or cells isolated by laser capture microdissection from human tissue procured during clinical trials [47, 53, 54, 65].

In a landmark study, Boyd et al. investigated how signaling pathways are differentially activated in different breast cancer subtypes [66]. In this study, the phosphorylation status of 100 proteins was examined in a panel of 30 different breast cancer cell lines. These cell lines have previously been classified into the three major molecular subtypes using a combination of gene expression data and ErbB2 status [67]. Briefly, cell lines were assigned to luminal or basal-like classes using gene expression data, and ErbB2 amplification status was assigned by means of quantitative reverse transcription to identify cell lines with more than four copies of the 17q12-q21 locus. Then, the phosphorylated protein status from the 30 breast cancer cell lines was analyzed by reverse phase protein arrays. In order to reduce dimensionality of the data and find patterns that might be related to the differential activity of signaling pathways in particular subtypes of breast cancer, the principle component analysis (or PCA, which convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components) was used. The results of this analysis showed that the global proteomic signature determined by this method largely separates basal-like cell lines from ErbB2 amplified and luminal cell lines along the second principal component. Also, with the exception of the ErbB2-amplified line BT474, the majority of the luminal lines are separated from the ErbB2 lines. This analysis suggests that the phosphorylated protein end points in this analysis are significantly correlated because the first three principal components can account for 61% of the variance in the data and also that distinct pathways may be activated in the different subtypes. Moreover, this analysis suggests that specific pathway activation events may be present in the different molecular subtypes. In particular, basal-like lines were found to be distinct from luminal and ErbB2-amplified lines in having low levels of pPTEN and high levels of total EGFR, pPyk2 Y402, and pPKC-α S567. ErbB2-amplified cell lines were distinct from the other subtypes in having high levels of pERBB3, pFAK, and pEGFR Y1173, and luminal cell lines were distinct in having higher levels of phosphorylation of p70S6K S371.
and A-RAF S299. In addition, this analysis revealed patterns of pathway activation that are not obvious from published gene expression analyses. In particular, basal-like cell lines were found to have high levels of phosphorylation of non-receptor tyrosine kinases, such as c-Abl and Pyk2, and in addition showed generally high levels of ERK1/2 phosphorylation and high total EGFR expression. In contrast, ErbB2-amplified cell lines were found to have high levels of phosphorylation of components of the EGFR pathway (e.g., Shc, ErbB3, EGFR), as well as other receptor tyrosine kinases (e.g., c-MET). Finally, luminal cell lines that do not have apparent amplification of ErbB2 showed generally higher levels of activation of downstream signaling pathway components in the AKT/mTOR pathway (e.g., p70S6K).

A potentially important application of reverse phase protein array technology is the more personalized administration of targeted therapies based on the signaling status of a given patient's tumor. The assumption is that if a patient's tumor is addicted to the continued activation of a particular pathway for continued growth and survival [68], then phosphorylation at key nodes in that pathway may serve as hallmarks, indicating the presence of an activated pathway and the potential for therapeutic intervention with inhibitors targeting that pathway. Similarly, PI3K is a key transducer of growth factor signals from receptor tyrosine kinases, as well as a frequently mutated oncogene, suggesting that PI3K inhibitors might have beneficial effects in treating cancers driven by pathologic alterations of this pathway [69]. The results reported by Boyd et al., suggest that activation of these pathway modules occur in a subtype-specific manner and can provide the basis for therapeutic intervention. If this is true, basal tumors, which display high levels of EGFR, activated ERK1/2, and phosphorylation of Src-activated effector kinases, such as c-Abl and Pyk2 would be potential candidates for combined therapies with antibodies and/or small molecule inhibitors used in clinical trials. These findings also highlight the potential utility of reverse phase protein arrays in confirming pathway modulation upon therapeutic intervention and applications in examining pharmacodynamic biomarkers of drug response. For example, it is well documented that an inhibitor of all isoforms of the class I catalytic subunit of PI3K, GDC-0941, results in potent and selective inhibition of multiple nodes in the PI3K/AKT pathway and, thus, that reverse phase protein arrays might have utility monitoring surrogate markers of compound activity. Conversely, the results of this study also showed that a selective MEK inhibitor results in potent down-regulation of pERK1/2 and actually increases signaling through the PI3K/AKT axis. This result highlights the fact that signaling pathways are dynamically linked networks and that perturbations in one pathway may have unforeseen consequences on interacting pathways that may affect response to therapeutic agents [70].

In a more recent study, Iadevaia et al. used a reverse-phase protein array to measure the transient response of the MDA-MB-231 breast cancer cell line after stimulation by insulin-like growth factor (IGF-1) [71]. The experimental results showed that when active, IGFR propagates the signal downstream through the Ras/Raf/MEK/ERK (MAPK) and phosphoinositide-3-kinase/AKT (PI3K) signaling pathways. The signals from the MAPK and PI3K cascades are routed to the mTOR pathway through tuberous sclerosis (TSC2) inactivation. Phosphorylated mTOR activates p70S6K, which inactivates the insulin receptor substrate (IRS-1) through a negative feedback loop.
The experimental results indicate that combined inhibition of the MAPK and PI3K/AKT pathways optimally inhibited the signaling networks and decreased cell viability. In contrast, combined inhibition of the MAPK and mTOR cascades led to significant activation of p-AKT and increased cell viability. Although several other kinases and pathways may potentially regulate the viability of the MDA-MB-231 cells, the experimental results indicated that simultaneous inhibition of the MAPK and PI3K/AKT pathways was sufficient to significantly reduce cell proliferation. The procedure is currently being used to identify and validate drug combinations that can inhibit aberrant networks in a panel of human cancer cell lines. **Figure 3** summarizes some of the deregulated signaling pathways described by the use of Phosphoproteomics.

**Figure 3. Altered signaling pathways in breast cancer.** This interaction map was created in the String 9.0 program (http://string-db.org) and summarizes some of the most commonly affected signaling pathways in breast cancer. Predicted functional links, consist of different colored lines: one color for each type of evidence. In this specific case, pink lines represent experimental evidence, blue lines represent interactions already published in databases and green lines text data mining.
7. Clinical implications

Cancer is among the leading causes of death worldwide. Therefore, the design of effective strategies to successfully implement personalized cancer medicine in clinical practice needs to face substantial challenges in the future. One of the biggest challenges in cancer research is the fact there is currently an insufficient number of effective rationally targeted drugs to implement this strategy broadly, at the time of this review, at least 50 distinct selective kinase inhibitors had been developed to the level of a phase I clinical trial, some of them have already been tested in breast cancer patients and it is expected that many more will be developed as cancer phosphoproteome analysis efforts continue to identify additional potential targets (Table 1).

<table>
<thead>
<tr>
<th>Kinase</th>
<th>Alteration</th>
<th>Therapeutic Agent</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Receptor Tyrosine Kinases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>Amplification, mutations gefitinib, erlotinib</td>
<td>[72]</td>
<td></td>
</tr>
<tr>
<td>ErbB2/Her2</td>
<td>Amplification</td>
<td>lapatinib, trastuzumab</td>
<td>[73]</td>
</tr>
<tr>
<td>MET</td>
<td>Amplification</td>
<td>PF2341066, XL184, SU11274</td>
<td>[74]</td>
</tr>
<tr>
<td>FGFR2</td>
<td>Amplification, mutations PKC412, BIF1120</td>
<td>[75]</td>
<td></td>
</tr>
<tr>
<td>AXL</td>
<td>Increased activation</td>
<td>R428</td>
<td>[76]</td>
</tr>
<tr>
<td>IGF1R/INSR</td>
<td>Overexpression</td>
<td>BMS-754807</td>
<td>[77]</td>
</tr>
<tr>
<td>EphA2</td>
<td>Overexpression</td>
<td>None available</td>
<td></td>
</tr>
<tr>
<td><strong>Non Receptor Tyrosine Kinases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ack1</td>
<td>Increased activation</td>
<td>None available</td>
<td></td>
</tr>
<tr>
<td>FAK</td>
<td>Overexpression</td>
<td>None available</td>
<td></td>
</tr>
<tr>
<td>Src/Lyn/Hck</td>
<td>Overexpression</td>
<td>dasatinib, AZD05030</td>
<td>[78]</td>
</tr>
<tr>
<td><strong>Serine/Threonine Kinases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI3K</td>
<td>Mutations</td>
<td>BEZ235</td>
<td>[79]</td>
</tr>
<tr>
<td>mTOR</td>
<td>Increased activation</td>
<td>everolimus</td>
<td>[80]</td>
</tr>
<tr>
<td>PLK</td>
<td>Overexpression</td>
<td>GSK461364</td>
<td>[81]</td>
</tr>
<tr>
<td>Aurora Kinases A and B</td>
<td>Overexpression</td>
<td>MK5108</td>
<td>[82]</td>
</tr>
<tr>
<td>Raf</td>
<td>Increased activation</td>
<td>sorafenib</td>
<td>[83]</td>
</tr>
<tr>
<td>MEK</td>
<td>Increased activation</td>
<td>PD0325901</td>
<td>[84]</td>
</tr>
<tr>
<td>ERK1/2</td>
<td>Increased activation</td>
<td>None available</td>
<td></td>
</tr>
<tr>
<td>Pak1</td>
<td>Amplification, overexpression</td>
<td>None available</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Oncogenic Kinases as Therapeutic Targets in Breast Cancer.

The current phosphoproteomic goals imply the identification of phosphoproteins, mapping of phosphorylation sites, quantitation of phosphorylation under different conditions, and
the determination of the stoichiometry of the phosphorylation. In addition, knowing when a protein is phosphorylated, which kinase/s is-are involved, and how each phosphorylation fits into the signaling network, are also important challenges for researchers in order to understand the significance of different biological events. The new phosphoproteomic technologies are fundamental for cataloguing all this information, and it is heading towards the collection of accurate data on phosphopeptides on a global scale. In addition, the possible difficulties to get sufficient amount of specific phosphorylated proteins of specific low abundant protein-kinases in vivo which might limit the usability of the phosphoproteome analysis, must be pointed out. The concept of personalized cancer medicine also has significant implications for the drug development industry, which is beginning to recognize and appreciate the need to alter the current business model for drug development and clinical testing. Moreover, the clinical success of such kinase inhibitors as imatinib, erlotinib, and lapatinib has validated this strategy and has prompted a virtual explosion in the development of additional kinase inhibitors for cancer therapy. Importantly, though, with these successes has also come the realization that these agents are generally effective for a relatively small subset of treated patients, often defined by a common genomic, proteomic and/or phosphoproteomic denominator present within the tumor cells. Such findings have highlighted the potential importance of identifying defined patient subpopulations before treatment with kinase inhibitors to optimize clinical outcomes.

Finally, it is important to state that to develop clinical proteomic applications using the identified proteins and phosphoproteins, collaboration between research scientists, clinicians and diagnostic companies, and proteomic experts is essential, particularly in the early phases of the biomarker development projects. The proteomics modalities currently available have the potential to lead to the development of clinical applications, and channeling the wealth of the information produced towards concrete and specific clinical purposes is urgent.

8. Concluding remarks

Cancer has been described as both a proteomic and a genomic disease [66]. Only those genetic defects creating a survival advantage increase the tumorigenic potential and are reflected in an altered functional state [19, 67]. Thus, the current challenges of cancer treatment, e.g. why do some patients respond to cancer drugs, while others do not, can only be answered with comprehensive efforts and by integrating knowledge on genetic and chromosomal aberrations, clinical data, IHC, and quantitative protein profiling.

Phosphoproteomics has played a significant role in our ability to understand molecular mechanisms that govern human cancers. Various technological platforms are now available for phosphoproteomic studies enabling us to address different aspects of tumor biology governed by phosphorylation-mediated signaling pathways. These studies have clearly taken us beyond looking at mutations or other genetic variations commonly observed in cancers and are providing us insights into functional consequences of these changes in
conferring survival advantages to cancer cells. Such studies are already being used as the basis for determining therapeutic options. With an ever increasing list of kinase inhibitors being developed by pharmaceutical companies, such strategies have become vital not only to determine the targets of these inhibitors but also to study their off-target effects. We foresee phosphoproteomics emerging as a vital technique in clinical research to assist in diagnosis, prognosis and treatment of cancers. The major challenge ahead is to develop this technology further to make it amenable for use in the clinic with as few sample processing steps as possible.

There are several issues, however, that must be carefully and promptly addressed if we are going to fulfill the dream of bringing individualized cancer care closer to reality. First of all, we must acknowledge the value of long-term research and provide the appropriate legal and ethical framework to encourage the collaboration among all the stakeholders in the cancer ordeal. Bridging the gap between basic and clinical research, facilitating the engagement of the industry, creating new infrastructures and bio banks, as well as the creation of innovative clinical trials are among the items that require urgent action. The aim of cancer research is to improve the life expectancy and quality of life of patients and we must make every effort to coordinate current activities in order to achieve this goal.

Author details

Olga Villamar-Cruz and Luis E. Arias-Romero*

Cancer Biology Program, Fox Chase Cancer Center, Philadelphia, PA, USA

Acknowledgement

We gratefully acknowledge the helpful comments from E. Arechaga-Ocampo, C. Perez-Plasencia and our anonymous reviewers.

9. References


* Corresponding Author


[53] Zha, H., et al., Similarities of prosurvival signals in Bcl-2-positive and Bcl-2-negative follicular lymphomas identified by reverse phase protein microarray. Lab Invest, 2004;84(2) 235-244.


