Philippine Journal of Biochemistry and Molecular Biology
2020; 1(1): 12-28
http:// http://www.psbmb.org/pjbmb.html
doi: 10.5555/pjbmb.ph.2020.01.01.12

PJBMB

Review Article

Mass Spectrometry and Proteomics as Emerging Technologies for Breast Cancer

Maritess D. Cation,[1,2,3] and Maria Cristina Ramos[1,4,5]

[1]The Graduate School, University of Santo Tomas, Manila 1015 Philippines;
[2]Institute of Chemistry, Academia Sinica, Taipei 11529 Taiwan;
[3]Faculty of Pharmacy, University of Santo Tomas, Manila 1015 Philippines;
[4]Research Center for the Natural Sciences, University of Santo Tomas, Manila 1015 Philippines;
[5]College of Science, University of Santo Tomas, Manila 1015 Philippines

Email address:
mdcation@ust.edu.ph

To cite:
Cation, MD; Ramos, MC.2020. Mass Spectrometry and Proteomics as Emerging Technologies for Breast Cancer. PJBMB. Vol. I, No. 1, 2020, pp. 12-28. doi: 10.5555/pjbmb.ph.2020.01.01.12

Received: 07 16, 2020; **Accepted**: 10 27, 2020; **Published**: 12 02, 2020

Abstract: Breast cancer among women has shown a steady increase in incidence and mortality rates in the Philippines, and around the globe. To date, there are a few and limited biomarkers approved for diagnosis and target for therapy. Some tumor tissues do not express any valid biomarkers in clinical tests, and patients from this group are unlikely to respond well to hormone therapy. Here, we presented a comprehensive literature resources citing potential biomarkers found from omics-based assays. More importantly, we also presented a rich list of significantly expressed novel protein biomarkers found through mass spectrometry and proteomic analysis. By applying mass spectrometry technology, we can achieve deep and large proteomic profiles from cells and tissues. The latest developments in mass spectrometry and its application will bring a big impact in breast cancer research and drug discovery as we find novel proteins and its association to various pathways linked to the hallmarks of breast cancer.

Keywords: Proteomics, mass spectrometry, breast cancer, biomarker, tumor marker, proteogenomics, omics

INTRODUCTION

The World Health Organization during the World Cancer Day 2020 recently warns the public of probable spike in breast cancer estimated to affect more than 15 million women globally by 2030 [1]. In the Philippines and most developing countries in Asia, a slow but steady increase in the rate of incidence, mortality and breast cancer recurrence is manifested. Health experts around the world have successfully implemented different strategies to mitigate and control breast cancer. They encourage individuals to conduct self-examination of the breast, and to undergo an annual breast cancer screening test to assess risk, and screen for breast cancer before its onset [2]. Breast cancer today is treated according to the presentation of specific breast cancer biomarkers in each patient. Breast cancer is detected with very few validated biomarkers - estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2). However, there are a few patients who do not show evidence of any biomarkers with these in clinical laboratory tests, and therefore do not have good prognosis. Another drawback affecting accuracy in current clinical laboratory diagnosis in cancer is the availability of limited tissue biopsy for examination. The tissue obtained depending on its size may or may not be sufficient for the numerous pathological tests to perform. The biopsy tissue samples would naturally have a heterogeneous form which may lead to poor conclusive information on the properties of the cancer as a whole. These challenges are now being addressed with modern technology and sophisticated instrumentation such as the mass spectrometry paired with software for computational tools. It is regarded as fast and emerging technology useful in discovering new and potential protein biomarkers for breast cancer.

A. Standard Breast Cancer Screening Methods

The female breast is made of milk-producing glandular tissues, lobules, ducts, and fatty tissues [3]. A thorough self-examination can observe changes on the skin, around the breast area, and lymph nodes near the armpit. Visualizing the lump can be done by breast mammography using low-dose X-rays to identify irregularities in the breast density and mass formation in the breast. Breast ultrasound is a better alternative to mammography since the later although noninvasive is painful and is not well tolerated by other patients. Magnetic resonance imaging (**MRI**) is another sophisticated but not too commonly used pictorial imaging of the breast. MRI provides a more accurate breast cancer staging results than the first two methods [4]. The images derived from these imaging methods are inspected by technicians and medical doctors to describe the nature of the size of the lump, malignancy of the disease, or make recommendation for additional laboratory examinations. The way this examination is done even with experienced personnel may pose inaccurate diagnosis, carry the risk of false-positive results, and discrepancies due to the limitations of the imaging method itself, and the subjective judgment of the observer. Nevertheless, these imaging technologies combined with regular self-examination have contributed to reduce mortality among breast cancer patients.

If the preliminary breast cancer screening results show suspicious lumps, other test may be recommended and a biopsy will be done [5]. Biopsy is an invasive procedure which collects breast tissue samples by open surgery or by needle biopsy collection. Immunohistochemistry (**IHC**) is the standard assay platform for the pathology of tissue biopsy and surgical resection specimen. In breast cancer, IHC for ER, PR, and HER2 are performed using laboratory produced antibodies. IHC may arrive at varying results due to the heterogeneity of the breast tissue, choice of antibody, age of fixative and reagents, different manufacturers, and the kind of immunostaining methods applied. In situ hybridization (ISH) or fluorescence in situ hybridization (**FISH**) test is done to complement IHC test to measure HER2 more accurately when IHC results are negative or at the borderline in terms of HER2 scoring.

B. Gene and Protein Biomarkers Associated to Breast and Cancer Development

The mammary gland undergoes three distinct stages in which hormones and growth factors play critical role in the complex biological process called mammogenesis. At the instance of embryo development, the skin's ectoderm will form mammary placodes until puberty [6]. The first stage forms an immature mammary gland in a process called "ductal morphogenesis" which is regulated by ER and insulin-like growth factor 1(**IGF**-1) [7], [8]. Upon reaching puberty, the growth factor-like amphiregulin (**AREG**) becomes more abundant for proper development of the mammary epithelial cells. Ovaries secrete estrogen hormone to promote fat cells to proliferate and deposit around the breast connective tissues thereby enlarging the size of the breast [9]. As a person matures physically, progesterone, estrogen and other growth hormones are involved in the development of the breast, and production of milk during pregnancy. As a female adult becomes pregnant, prolactin and progesterone promote milk production. The alveolar at this time switches to secrete more milk during the lactation period. As the weaning period ceases to stop, milk production and supply begins to decrease with reduced demand. At this point involution takes place to return the mammary glands back to its normal pre-pregnancy condition [10].

As the breast ages, there is a decrease in the ducts embedded in strands of collagen, depletion of mammary stem cells, and replacement of interlobular connective tissues and glandular epithelium tissues with fat. Early onset of lobular involution in premenopausal women have significantly decreased the risk of developing breast cancer compared to those with delayed involution [11]. These stages in the normal mammary gland development, functioning and aging all require numerous chains of biochemical events involved in several signaling pathways. When any of these biomolecules undergo uncontrolled changes or mutation it becomes contributory to the risk of developing breast cancer.

Breast cancer are diagnosed in clinics for detection and analysis of the status of abundance of hormone receptors. Breast cancer tumors found to have an over expression or high abundance of these protein biomarkers will be called **ER+, PR+,** or **HER2+.** If none of these three receptors are found in the breast cancer tumor then such disease is considered as triple negative breast cancer (**TNBC**). With such presentation or absence of these receptor, breast cancer is subtype D accordingly in clinics. Table 1 summarizes the intrinsic breast cancer subtypes into 4 major groups [12].

.

Table 1. Intrinsic breast cancer Subtypes of breast cancer according to receptor and gene detection

Group	Breast Cancer Subtype	Biomarker/s overexpressed
1	Luminal A	ER+, PR+/PR-
2	Luminal B	ER+, PR+/PR-, and HER2+
3	HER2 positive, nonluminal	HER2+
4	Triple-negative **or** basal-like	No known biomarker

Estrogen has long been implicated to support growth of breast cancer because it binds directly to ERs found on the cell membrane. Estrogen receptors are transcription factor proteins encoded by the estrogen receptor 1 (**ESR1**) gene. It is composed of domains which support hormone binding, and activation of transcription for the

synthesis of proteins. The **ER+ breast cancer** manifests abundant ERs that bind with estrogen promoting deregulated breast cancer growth. Anti-estrogen drug is an anti-estrogen hormone therapy acting as estrogen antagonist to block estrogen from binding to the ERs. Progesterone is an important steroid hormone which regulates changes in the reproductive system. Most breast cancers express both ER and PR together, as a result PR is usually studied always alongside with ER. The progesterone receptor positive (PR+) kind of breast cancer would not only be PR+ but often be ER+, too. **Hormone receptor negative** or TNBC do not over express ER, PR nor HER2 receptors. TNBC to date remains to have the worst prognosis and available hormone antagonistic therapies in most situations will not work with this type of cancer.

In a study by Chiang et al., they found that in ER+ invasive breast carcinomas, AREG was contributory to the increased invasiveness of breast cancer cells. Suppressing AREG expression in transformed human breast epithelial cells in nude mice showed decrease formation of tumor [13]. AREG mechanism have not been completely understood but it is hypothesized to cross the membrane to the stromal fibroblasts activating the epidermal growth factor receptor (**EGFR**), and induce expression of fibroblast growth factors (**FGFs**) leading to cellular proliferation. In adult female breast, some hyperplastic enlarged lobular unit (**HELU**) studies revealed both ER and AREG are upregulated and are suspected to induce self-propagating growth leading to tumorigenesis [14]. Such findings support that AREG plays a role in breast cancer development particularly for ER+ tumors [15], [16].

Another study identified progesterone and its relation to the tumor necrosis factor ligand superfamily, member 11 gene which encodes the receptor activator of nuclear kappa-B ligand (RANKL) and its receptor – RANK. RANK, RANKL and progesterone are all found to be essential in the milk production of the mammary gland during pre-pregnancy stages [17]. The RANK/RANKL signaling pathway is also regulated by progesterone being the main upstream regulator of the RANK/RANKL pair. The RANK/RANKL pair expression also were found to control stem cell expansion and generate the pro-growth response, and drives cell proliferation in progestin-dependent breast cancers [18], [19].

Risk of breast cancer was investigated from isolated extracellular matrix (**ECM**) from post-weaning mammary glands. The study revealed some mass forming ECM fragments which increases the metastasis of breast cancer in mice, and the invasiveness of breast cancer cells in culture [20]. Post-lactation involution processes have also been connected to breast cancer and tumorigenesis in few transcriptional profiling studies [21]. It is assumed that involution may lead to a tumor microenvironment that alters preneoplastic mammary cells leading to transient increase in breast cancer risk after pregnancy [22].

A few of the linked oncogenic pathways to breast cancer are the ER signaling pathway [23], HER2 signaling pathway [24], nuclear factor-kappaB (NFκB) signaling pathway [25], mitogen-activated protein kinase (MAPK) signaling pathway [26], phosphatidylinositol 3-kinase/protein kinase B/mammalian target of rapamycin (PI3/AKT/mTOR) signaling pathway [27], and notch signaling pathway [28]. When sudden changes in the body's physiology and conditions take place, these signaling pathways maybe altered, hijacked or dysregulated by activities of the cells leading to uncontrolled cell growth, cell invasiveness, and suppresses cell apoptosis [12]. These pathways are considered in most breast cancer studies and some drugs have been developed to inhibit downstream signaling to control breast cancer.

Other studies look at the genetic DNA for preventive and prophylactic approach to breast cancer treatment. Germline or somatic mutations in genes are suspected to code for growth of the cancerous cells [29]. One of the genes linked to breast cancer is the human epidermal growth factor 2 gene, also called HER-2/neu or ErbB-2 gene (**ERBB2**). The amplification and deregulation of the ERBB2 gene leads to an overexpression of HER2 in tumor cells. Burstein et al., found out that HER2 overexpression may reach as high as 50-fold higher than its normal abundance in 30% of the invasive breast cancer tumor tissues used in this study [30]. The HER-2 protein is found to activate molecular pathways supporting cellular proliferation and metastasis of breast cancer.

Some breast cancers are hereditary and it is passed from an autosomal dominant pattern from any one of the parents to offspring such as the mutations presented by the breast cancer gene 1 (**BRCA1**) and breast cancer gene 2 (**BRCA2**) [31]. Presence of these genes increases one's risk to developing breast cancer and morbidity [32]. At normal conditions, the breast cancer genes function by encoding proteins to control and repair damages in the DNA or fix other gene mutations. BRCA genes also prevent uncontrolled cellular division and growth suppressing rapid tumor formation. But when BRCA genes are mutated it leads to forming abnormal and nonfunctional versions of the corresponding BRCA1 and BRCA2 proteins.

From the abovementioned biomarkers still very few are currently being considered for breast cancer clinical diagnosis to date. These biomarkers are either genes or proteins, with known cellular regulation and binding properties controlling downstream signaling activities. In some instances, these molecules are modified by external factors leading to its distorted structure affecting its role to maintain the cell's normal homeostasis. With increasing demand for potential biomarkers for diagnosis, a high throughput and multiplex analysis method in the detection is an emerging need. Biomarker discovery by **mass spectrometry (MS)** is becoming a popular method of

choice, and as well as an effective tool in disease diagnosis. This method is capable of analyzing tissue samples that are heterogeneous in nature, and are usually available in small or limited quantities. Experienced MS users were able to achieve wider and deeper analytical capability in many different kinds of samples. MS has also delivered analysis with high sensitivity, reproducibility and repeatability, and accuracy and precision over multiplex analysis of samples. In MS, profiling of the biopsy of both tumor (**T**) and its adjacent non-tumor tissues (**NT**) can also be done simultaneously. This method can deliver better identification and differentiation measures of breast cancer-associated biomarkers in every patient.

C. Genes to protein biomarkers in breast cancer clinical diagnosis

Prognosis estimates the course and outcome of cancer such as the likelihood of recurrence, remission and survival. An individual breast cancer prognosis is done in several phases, first tumor biopsy is subjected to histomorphology to examine its type, grade and size of the cancer, and its presence in the lymph nodes. The second phase detects protein biomarkers, ER, PR, and HER2 expression status. In cancer prognosis, about 20-30% of breast cancer tumors have malignant breast cancer cells that have unusually high concentration of HER2 receptor proteins. Overexpression of HER2 found on tissues are contributory to unusually rapid cell proliferation. This is carried on by the dimerization of HER2 with other EGFRs leading to the activation of the growth factor signaling pathway driving cancer to grow [33]. Some HER2 breast cancer patients show improvement with first generation adjuvant therapy like trastuzumab, while others would require other HER2 targeted therapies such as antibodies pertuzumab and adotrastuzumab emtasine, or a kinase inhibitor like lapatinib [34].

Another protein biomarker used in breast cancer care is **Ki-67** antigen expressed by the marker of proliferation Ki67 (**MKi67**) gene. This is used as a prognostic biomarker for measuring proliferation, predicting drug response, and drug resistance [35]. Ki67 concentration in tumor positive cells increases as cells prepare to divide and form new cells. Hormone receptors together with Ki-67 are quantitatively measured using IHC based assays [36]. Results from the IHC assay have brought some concerns such as the nonlinear nature of IHC staining due to the heterogeneous nature of the tumor tissue, the antibodies itself and slide scoring applied, and the different subcellular location of the different biomarkers. Genomics have addressed these issues and led to the development of various gene-based technologies for tumor biomarker assessment. There is a free online tool called “Predict,” that may be used to provide patients and doctors varied treatments options for post-operative early invasive breast cancer patients [37]. Estimating prognosis have also been successfully done with genomics assays approved by USFDA to profile gene expression and has been used as one of the more accurate assays for diagnosis and estimating prognosis in breast cancer. In various conditions and stages of breast cancer, several multianalyte tests are used routinely in the laboratory examination of patient samples. Current gene-based assays include *Oncotype DX, urokinase plasminogen activator (uPA)-PAI-1, BBDRisk Dx, Immunohistochemistry 4 (IHC4), BreastSentry, MammaPrint, EndoPredict, Breast Cancer Index, Natera's Signatera Molecular Monitoring (MRD), Rotterdam Signature 76-Gene Panel, HERmark, NexCourse IHC4, Mammostrat, Symphony (Agendia), GeneSearch BLN, Insight TNBC type, and Prosigna (PAM50).* Each assay has its own advantages and applies only to specific subgroups of patients to predict prognosis and plan for the most suited adjunct therapy treatment.

There are also ind*ividual genes USFDA approved for prognosis* in some types of breast cancers *such as neurotrophic tropomyosin receptor kinase (**NTRK**)* genes, and phosphatidylinositol 3-kinases *catalytic subunit alpha (**PI3KCA**)* gene. NTRK genes were observed upon the fusion of proteins encoded upon by three NTRK genes in secretory breast carcinoma [38]. The PIK3CA gene commonly found mutated in ER+ breast cancer alters the activity of class IA phosphatidylinositol 3-kinase (**PI3K**). The mutation causes PI3K to downstream activate the PI3K/AKT/mTOR pathway involve in breast cancer [39]. Patients with advanced breast cancer having PI3K mutation are given some PI3K inhibitor like alpelisib along with fulvestrant [40]. These findings suggest that detecting genes is useful in determining therapies for lowering the risk of breast cancer onset and recurrence. On the other hand, incorporating genetic testing into breast cancer care requires accredited laboratories to perform this clinically validated tests, technical skills of analyst, and modern instrumentation to arrive at accurate results.

Complementing genomics is the emerging proteomics approach to diagnosis and prognosis. The study of proteomics has brought remarkable advancement in the better understanding of the disease since proteins are directly involved in all cellular processes. The carcinoembryonic antigen (CEA) family of glycoproteins is the most widely used antigen found in the tumor tissue. One of the major subfamilies of these genes is the **CEA** cellular adhesion molecule **(CEACAM)** family belonging to the immunoglobulin superfamily. The adhesion properties of these molecules with one another or with other molecules suggest that alternations in cell adhesion play an important role in cancer metastasis. CEA levels are measure to aid in breast cancer diagnosis, clinical staging, responsiveness to chemotherapy or radiotherapy treatments, and monitoring recurrence in post-operative patients [41]. The cancer antigen 15-3 (**CA 15-3**), and cancer antigen 27.29 (**CA 27.29**) are normally

expressed in healthy cells to work in controlling abnormal cellular growth. In cancer cells, elevated CA 15-3 amounts in blood are used to monitor the stage of breast cancer, and measure the effectivity of a breast cancer therapy [42]. A more sensitive biomarker than CA 15-3 for metastatic breast cancer is CA 27.29. This highly polymorphic glycoprotein is found to be less glycosylated in breast cancer tumor cells than its usual form [43]. The biomarker is expressed throughout malignant epithelial cells of the breast and measured together with other tumor markers to improve specificity in disease staging, track effectivity of therapy, and monitor breast cancer recurrence.

USFDA have approved protein tumor markers for breast cancer longitudinal study including the circulating tumor cell analysis of epithelial cell adhesion molecule (**EpCAM**), **CD45** antigen, and cytokeratins (**CK8, CK18,** and **CK19**). Circulating tumor cells (**CTCs**) though exist in minute concentrations in the blood, urine, stool or other body fluids have occupied a big role in many cancer research seeking to find prognostic and therapeutic values in metastatic breast cancer [18], [44]. EpCAM expressed by normal CTCs, is a membrane protein being considered to become a novel drug target for gene therapy. In a study done by Osta et. al, they compared the primary and metastatic breast cancer against normal breast tissue and discovered that EpCAM mRNA expression levels to be differentially overexpressed by 100 to 1000-fold in breast cancer tumor tissue [45]. Such tumor markers may also be measured periodically to monitor good response of cancer therapy by detecting a significant decrease in the concentration of circulating tumor marker. **CD45** antigen also known as protein tyrosine phosphatase, receptor type, C (**PTPRC**) is a signaling protein involved in a number of cellular processes including cellular division and growth. It is also found to regulate functions of some antigen receptor complexes and kinases necessary for antigen receptor signaling [46]. In breast cancer CD45 are important in the diagnosis and can be used to monitor changes relative to the effectivity of the applied drug for therapy. CK8, CK18, and CK19 are proteins expressed by the epithelial cells normally lining the breast tissue. Measuring changes in the levels of these protein biomarkers concentration would determine the cancerous condition of breast-associated adenocarcinomas [47], [48]. Lustberg *et al. also* reported in their findings that circulating atypical cells detected in blood have high expression of CK8, CK18, CK19, and CD45 biomarkers [49]. Researchers continue to discover possible mechanisms of these CTC biomarkers relevant in understanding tumor metastasis. The first and only approved immunotherapy currently used in breast cancer clinics is the detection of **programmed cell death-1 receptor (PD-1). PD-1** is an important signaling protein found on the surface of immune cells and some tumor cells Its role is to provide immune inhibitory signals to fight pathogens and cancer cells. By blocking PD-1 with drugs such as atezolizumab, PD-1 can help boost the body's immune system [50]. In some forms of cancer, IHC analysis is performed to detect **PD-1 ligand (PD-L1)** overexpression from patient samples. The JAK1/JAK2-STAT1/STAT2/STAT3-IRF1 pathway regulates the expression of PD-L1 upon secretion of interferon gamma (IFN-γ) by the T cells [51]. The molecule PD-1 once bound to PD-L1 delivers a negative modulatory signaling pathway to activate T cells. The activation of due to PD-1 and its ligand PD-L1 in the PD-1/PD-L1 pathway induces downregulation of T-cell activity, cell proliferation, induction of tolerance to antigens, and trick the immune cells from destroying damaged cells [52]. In metastatic TNBC, blocking PD-1 or PD-L1 with specific antibodies have shown good response and control over cancer growth [53]. IHC when compared to MS-based analysis of PDL1, IHC was found to show lower concentrations of the glycosylated protein. This result may lead clinicians to ignore PDL1 expression and may misdiagnosis patients with inappropriate immunotherapy. In this respect, MS-based absolute quantification delivers better sensitivity to detect protein expression levels even when it is in its glycosylated form [54].

Proteomics and genomics, together termed as proteogenomics, in conjunction with computational tools, newer technologies are leading the way to finding better solutions in discovering specific biomarkers to combat breast cancer. This underscores as well the importance of biomarker in the more advanced stages, recurring types and those that are rare, and have difficulty in finding appropriate and precise treatment. Collectively, these attempts in proteogenomics is leaning towards an improved quality-of-life among breast cancer patients, and an eradication of the disease in the near future.

D. Mass Spectrometry-based Proteomics Approach

Research in breast cancer has not found a biomarker which provides measure for accurate breast cancer diagnosis and typing. Protein molecules are considered best biomarkers for they have different functions and are involved in many biochemical reactions in the body including the body's fight to prevent diseases including cancer. To perform analysis of the proteins, mass spectrometry was widely adopted for proteomics research. In the past years, peptides cannot even be analyzed in MS since peptide samples cannot be ionized in its gaseous form. With the latest application of ionization technology even **peptides** and other simple molecules to as low as femtomole quantities in liquid states can possibly be analyzed and sequenced today in MS.

MS-based proteomics strategies can either be top-down, or bottom-up proteomics [55]. Proteins are first extracted from biological samples like cell lines, tissues, tumor specimens, and other derivatives. These proteins are enzymatically digested into peptides and then analyzed by liquid chromatography−tandem mass

spectrometry (**LC−MS/MS**). In **top-down proteomics** it analyzes intact proteins separated first by isoelectric focusing (IEF) according to its isoelectric pH, then by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) according to its size. The separation of proteins can be viewed by performing protein staining or by using fluorescent tags. The two-dimensional electrophoresis (2-DE) technique is low throughput and is limited by protein solubility and sensitivity of detection. In this technique a big amount of sample is needed for this assay. Its minimum detection sensitivity is down to a few numbers of proteins with molecular weights of at least 120 kDA per run. Specific proteins of interest are excised from the gel, purified, and analyzed using the mass spectrometer [56]. An improved protein profiling strategy was later developed called the "**bottom-up**" or "**shotgun**" **proteomics** [55] which refers to reconstruction of protein information from peptide sequences. In shotgun proteomics high performance liquid chromatography (**HPLC**) in tandem mass spectrometry is used for characterization of proteins from complex samples [57], [58], clinical biomarkers discovery [59], identification of post-translational modifications (PTMs) [60] and protein-protein interactions [61] to explore biological system [62].

Peptides are digested proteins extracted from samples and digested usually with trypsin, or other proteases. A final step requires peptide purification by washing away salts, detergents and inactivated enzymes to eliminate matrix effects in LC-MS leading to ion suppression or co-elution of distinct peptides. Purified peptides are analyzed with optimized workflows and LC-MS method settings to achieve efficient mass fragmentation datasets for identifying peptide sequences and proteins. The resulting analysis will be matched to protein sequences from protein sequence databases [63] and spectral libraries of fragment ions and precursor ions [64], [65]. Spectral libraries are acquired a priori from single shot and fractionated samples from tissues or cell lines from an organism. The spectral libraries are used for peptide identification and quantification for protein sequencing at 1% false discovery rate [66]. MS engineering technology have continuously achieved better results over the years. Latest chromatography columns have been designed with varying dimensions, retention characteristics, particle size, and selectivity of the column material. These enhancements make it possible to detect a broader protein range, better peak resolution and shorter gradient time. Using LC-MS technology proteome profiling acquired from reproducible analysis have presented improved sensitivity in peptide and protein identification and quantitation, specificity and discrimination between protein isoforms. Datasets derived from MS analysis of the peptides are submitted to software capable of automatically sequencing the MS/MS (also written as MS2) spectra of the fragmented ions. Peptide identification is done by comparing the mass spectrum of peptides to mass spectra predicted from public sequence databases or mass spectral libraries.

The development of quantitative methods has become the focus of MS-based proteomics research, with the aim of achieving high precision and accuracy in quantitation, as well as a high reproducibility and low number of missing values. Protein quantification is usually performed in two major approaches: the use of stable isotope labelling and label-free techniques. These methods provide opportunity for sample multiplexing, and quantitation analysis based on relative intensities of the reporter ions. Most proteomics quantification workflows involve chemical labelling techniques for relative and absolute quantitation, such as Stable Isotope Labeling by Amino Acids in Cell Culture (**SILAC**), Isobaric Tags for Relative and Absolute Quantification (**iTRAQ**), and Tandem Mass Tags (**TMT**). Labelling strategies allow multiplexing few samples at the same time, however, they are limited by the high cost of isotope labels, varying labeling efficiency and software for data analysis. Some of the techniques of labeling include protein or peptide, chimeric recombinant protein, isobaric and metabolic either in vivo or in vitro. SILAC [67] is used in vivo, while iTRAQ [68], and TMT [69] are used in vitro.

SILAC requires cells cultured with light medium using normal arginine (blue color Arg-0 isotope), and medium with heavy arginine (red color Arg-6 isotope). These isotopes of arginine are metabolically incorporated into the proteins while cells are growing under certain experimental conditions. Afterwards, cells are harvested and proteins are extracted from each set-up. Mass spectra of the corresponding peptides in both medium and light cultures are analyzed. Combination of both light and medium sample peak intensities in the mass spectrum will give the ratio of its relative protein abundance. If the protein would have a ratio of 1 this means that the abundance in both light and medium samples are the same. If the protein ratio is less than 1, then the protein abundance is greater in the medium sample than in the light sample, and vice versa if ratio is more than 1. These information from SILAC may lead to the identification of differentially expressed proteins in the sample. iTRAQ is also used for cells, tissues and other samples. It uses stable isotopes in iTRAQ reagents to be covalently bonded to the N-terminus and side chain amines of proteotypic peptides. iTRAQ enables relative quantification of very complex tissue mixtures thus it is been used in MS analysis in various applications such as comparison between normal and adjacent tissues, drug treated samples of cancer patient samples, biological replicates. TMT is currently the latest add-on in the label-methods for MS/MS analysis and have great advantage for its ability for multiplex analysis surpassing that of what iTRAQ may offer. TMT can perform analysis with several samples within one experiment producing MS data sets that defines the proteome of each sample. By TMT method analysis can have a reduced overall analytical time and eliminates

variations from one sample run to another. For the protein quantitation and identification, the isobaric set of six mass tags with five isotopic substitutions called tandem mass tags six-plex (6-plex-TMT) were the first to be used as chemical labels for cell lines or patient samples. There are different TMT products now made available capable of conducting as high as 16-multiplex (16-plex Pro-TMT) analyses. TMT method was also compared to label-free DIA method in a study performed in the lab of Muntel [69]. They found that MS label-free approach yields high protein quantification and identification with less than 2% missing values, while TMT approach have achieved higher quantitation accuracy.

These labelling approach to proteomics improve sensitivity and the analysis when compared to other approach display fewer missing fragment ion intensities in the analysis. Table 2, highlights the characteristics of the three labeling techniques.

Table 2. Characteristics of SILAC, iTRAQ, and TMT

	SILAC	iTRAQ	TMT
Labelling Method	Stable isotope - labeled lysine and arginine	Isobaric reagents	Stable isotope labels
Sample	Living cells	Peptides	Proteins and peptides
Advantages	Less sample required Can be used in live cell lines	High throughput (4-/8-plex) High sensitivity Good reproducibility	High throughput (up to 10-plex) High sensitivity Efficient separation ability Good reproducibility and repeatability
Disadvantages	Expensive reagents Limited to cell line samples	Expensive reagents	Expensive reagents Low scan speed requirement decreases sample throughput
Reference	[67]	[70]	[71]

MS-based proteomics strategies make it also possible to perform measurement of proteins that have undergone some form of PTMs including ubiquitylation, phosphorylation, and acetylation to name a few among hundred types [72]. PTMs enable energy-efficient protein function optimization and diversifying its functionality at the cellular level to regulate cellular processes and pathway signals for growth, proliferation, and apoptosis. In a pool of peptides, PTMs are mapped by shotgun sequencing, or are fished out with a variety of affinity-enrichment or by covalent capture techniques, or its combination. In doing so, specificity is improved and co-eluting distinct peptides prevented. An approach to improve shotgun sequencing approach in LC/MS is by performing a prefractionation step. This will improve the dynamic range and expands the LC/MS-MS capacity to detect beyond the four to five orders of magnitudes [73]. By doing so, an increase in peptide coverage and identification of low abundant proteins (LAPs) to the high abundant proteins (HAPs) [74] happen. Protein abundance in some samples like in a single cell protein, can vary from 50 to 1,000,000 relative quantities. When the HAPs and LAPs are together, proteomic analysis becomes difficult since the bigger molecules can dominate in the sample matrix and hide the appearance of LAPs. There is also a big consideration in research of LAPs for some studies suggesting that these are found to leak into the blood in very small quantities. Circulating tumor cells are known to contain LAPs which may yield potential breast cancer biomarkers.

PTM-specific enrichment approach improves sensitivity and specificity of MS analysis [75]. In this manner, phosphoproteins are isolated from proteomic mixtures either by anti-phosphoamino acid antibodies or PTM-specific affinity to the phosphate groups. After targeted enrichment, these phosphoproteins are digested to yield peptides which is later enriched by antibodies, titanium dioxide (TiO_2), or by immobilized metal affinity chromatography (IMAC). At this point peptide complexity and heterogeneity is addressed by separating it into components in ion exchange chromatography as final step prior to LC-tandem MS analysis. The result of this leading to the detection of thousands of phosphorylated sites [76].
Samples may also be processed without using isotopic labels because of some practical reasons and cost-consideration. In label-free MS, peptides are directly analyzed, and proteins are quantified on the basis of precursor ion signal intensity or spectral counting [80].Recently, label-free quantitative proteins strategy has become a stand-alone method or in combination with enrichment or other labelling methods. In label-free approach, no chemical labels or tags are added to peptides before it is submitted to MS analysis. Fragmentation and mass analysis of these peptides are obtained either by data dependent (DDA) or data independent acquisition (DIA) [77]. **Data dependent acquisition** (**DDA**) is a highly selective method commonly applied to many MS instruments including triple quadrupole, Orbitrap, and tandem Q-ToF or ToF/ToF. Large-scale precursor ions enter the quadrupoles. The quadrupole emits simultaneously high and low energies to facilitate collision-induced dissociation of precursor ions. Upon fragmentation, accurate mass measurement is detected in the instrument. In DDA, the peptide precursors are scanned and in each scan about 9-10 precursors having the highest intensities at MS1 are selected for further fragmentation for the sequential MS/MS. However, in DDA method only those with high intensity precursors are selected to enter the mass analyzer. The low intensities fragments are ejected out in a trajectory and will never reach the detector for measurement. DDA may eliminate some important peptides at this stage. It may also miss some precursors to be identified after MS/MS analysis creating some difficulties in quantitation.

The **data independent acquisition (DIA)** method is an answer to the earlier difficulties faced by using DDA. In

DIA, the entire mass spectrum produced from MS1 ion intensities is fragmented and undergoes MS/MS analysis. Precursors are taken part by part selected over a very narrow m/z range. This is known as isolation windows, and the acquisition method termed as sequential windowed acquisition of all theoretical

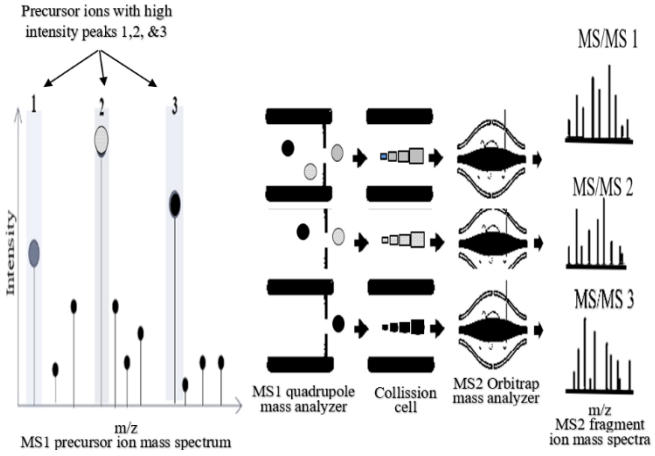

Figure 1. Data Dependent Acquisition. In the figure the high intensity precursors 1,2,3 in MS1 full scan will undergo sequential isolation and MS/MS fragmentation to produce a mass ion spectrum. (Abbreviation: HCD,High-enery collision dissociation)

fragmented ion mass spectrometry (**SWATH MS**) [77]. It is a high throughput label-free proteotyping technique applied in the mass analysis of every precursor ions found within narrow isolation windows. The precursor ions are fragmented over pre-selected smaller ranges of precursor m/z ratio range. All precursor ions enter sequentially the collision cells before reaching the final mass analyzer.

Due to the highly complex and convoluted DIA MS/MS spectra, it is difficult to interpret for peptide identification by conventional database searching tools without pre-processing. Instead the data analysis mainly relies on a prior knowledge information derived from fragment ion spectra of the targeted peptides. Commonly, there are two primary ways to interpret MS DIA data: (1) peptide-centric analysis (library query targeted analysis methods) and (2) spectrum-centric analysis (library-free approaches) [78]. Peptide-centric analysis needs a prior-knowledge information “peptide query parameters (PQPs)” stored in a spectral library. The library contains information of each targeted peptide consists of precursor m/z, peptide sequence of fragment ions, highly confident fragmented ions, and their relative intensity, and the normalized retention time. A high-quality and comprehensive **spectral library** is required for MS label-free identification and quantification. The

Due to the highly complex and convoluted DIA MS/MS spectra, it is difficult to interpret for peptide identification by conventional database searching tools without pre-

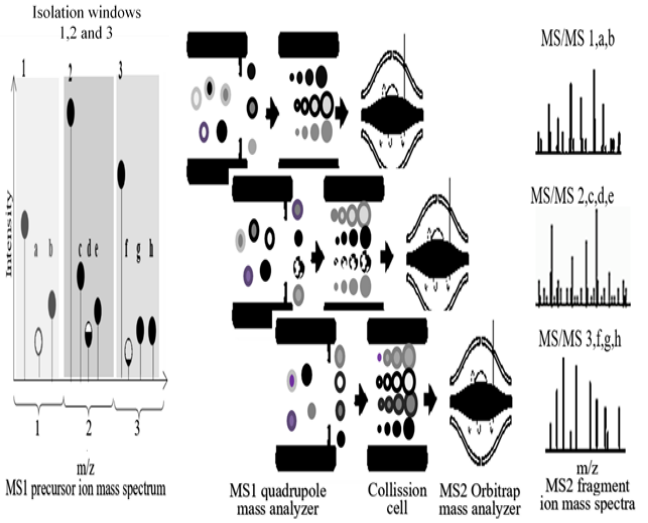

Figure 2. **Data Independent Acquisition**. In the figure all precursors produced in MS1 full scan will undergo MS/MS fragmentation in a sequential narrow m/z isolation windows 1,2 and 3.

processing. Instead the data analysis mainly relies on a prior knowledge information derived from fragment ion spectra of the targeted peptides. Commonly, there are two primary ways to interpret MS DIA data: (1) peptide-centric analysis (library query targeted analysis methods) and (2) spectrum-centric analysis (library-free approaches) [78]. Peptide-centric analysis needs a prior-knowledge information “peptide query parameters (PQPs)” stored in a spectral library. The library contains information of each targeted peptide consists of precursor m/z, peptide sequence of fragment ions, highly confident fragmented ions, and their relative intensity, and the normalized retention time. A high-quality and comprehensive **spectral library** is required for MS label-free identification and quantification. The spectral library is built from actual MS experiments in DDA mode acquisition and processed in a software like Maxquant [79] and Spectronaut [80] to generate a database of mass spectra. To achieve a normalized retention time, a set of shared endogenous or synthetic spike-in peptides such as the indexed retention time (**iRT**) are spiked in sample for accurate prediction of peptide retention time. This creates MS datasets with consistent, comprehensive and validated digital map of the entire proteome. False discovery rates are set to control the propagation of errors in MS analysis [81]. Some organism-scale spectral libraries or are now publicly available include *Homo sapiens*, *Methylobacterium extorquens (strain PA1), Drosophila melanogaster, Solanum lycopersicum* and *Streptococcus pyogenes* to mention a few [82]. To date, spectral libraries are continuously being expanded to include other organisms and specific human tissues to achieve deeper proteomic coverage.

As DIA is widely being used at present in MS data acquisition simultaneous development of software tools aided by artificial intelligence (AI) are also in progress. Common software like OpenSWATH [83], Skyline [78],

and MSPLIT-DIA [84] have enabled processing and simplified the analysis of highly convoluted SWATH-MS data. On the other hand, PQPs information can also be computationally predicted in silico spectral libraries by using deep neural networks to predict MS/MS spectrum, retention time, and fragment ion intensity. These in silico spectral libraries are used to interpret DIA data in deep learning methods such as DeepMass [85], Prosit [86], and DeepDIA [87]. These peptide-centric analysis approaches are sensitive and have more comprehensive DIA data than what is normally found in sample-specific generated spectral libraries. However, they limit peptide identification only to analytes found present in silico libraries. Thus, tools designed to detect peptides from DIA data without libraries will soon be introduced.

In spectrum-centric analysis, DIA spectra are most commonly interpreted using classical database search strategies [88]–[90]. This approach detects precursor ion chromatographic features and deconvolves the DIA fragments into pseudo-MS/MS spectra, which can then be directly searched with the traditional sequence database. Current spectrum-centric analysis DIA data software tools include Pulsar in Spectronaut [80], Group-DIA [91], PECAN [92], PEAKS [87], and DIA-Umpire [93]. However, these library-free based tools are not as sensitive as the library-based approaches [94], although these strategies have its advantages in identifying new peptide variants in DIA data sets. Further, spectrum-centric library-free search has demonstrated its potential in identifying novel peptides. To date, most published studies are still using the spectral library-based targeted extraction approach [94].

E. Proteomics for breast cancer molecular subtyping

Biomarkers have tremendously increased over the past decades, due to the advancement of high-throughput platforms for investigating the molecular characteristics of tumor tissue biopsies. With powerful bioinformatics tools integrated to aid in breast cancer research, it enabled us to further dissect the tumor to its molecular level and expand the spectrum of breast cancer subtypes. Microarray-based gene expression profiling has helped in determining breast cancer from its histopathologic type to its molecular subtype [95]. Today, ER+ and ER- breast cancer subtypes are considered as different diseases and are treated depending on the presentation of significant biomarkers [96]. Due to these development in genomics research on breast cancer biomarkers, The Cancer Genome Atlas (TCGA) Network has refined subtypes of breast cancer based from the extensive profiling of protein expression levels, microRNAs, and DNA gene mutations [97]. The molecular subtypes luminal A, luminal B, HER2-enriched, and basal-like breast cancers (Table 1) are undergoing paradigm changes at the molecular level to help in the modernization of breast cancer treatment [97].
Early proteomic studies of clinical breast tumor samples got low proteome coverage yield from low cohort sizes. Rezaul et. al, have identified about 1000 proteins from each of the 6 patient samples from which they found more than 200 differentially expressed proteins between the ***ER+*** and ***ER-*** breast cancer tissues. Their study discover the potential biomarkers ***Fascin, death-associated protein 5, Iprin-α1, and β-arrestin 1*** specifically found expressed only in the ***ER-negative subgroup*** [98]. Cha et al., studied 18 breast cancer samples of different subtypes and detected 298 significantly changing proteins that are associated to the transitional changes from the normal epithelial tissue conditions to a highly invasive malignant tumor. In their work, they found several proteins which are involved in alterations of downstream transcription factors (**TFs**) affecting regulatory pathways in breast cancer which can be used distinguishes ***malignant tumor from matched normal tissues*** [99].

With ongoing improvement in MS technologies, sample preparation, and bioinformatics tools have extended MS capacity to analyse larger cohort sizes. These developments led to improve both quality and quantity of proteomic data. In the lab of Liu et al., they made use of laser capture microdissection–nanoscale LC–MS/MS approach to study fresh frozen paraffin embedded breast tissues from 126 ***TNBC*** breast cancer samples. They were able to quantify about 3500 proteins, and identified 10 upregulated- potential protein biomarkers for novel therapies of distant metastatic TNBC. The study found that these proteins are involved in cell metabolism, cell death, immune response, transport of macromolecules, and biological processes linked to cancer progression. These proteins are apoptosis-inducing factor 1, mitochondrial (**AIFM1**), AP-1 complex subunit gamma-1 (**AP1G1**), AP-1 complex subunit mu-1 (**AP1M1**), F-actin-capping protein subunit beta (**CAPZB**), UMP-CMP kinase (**CMPK1**), catenin alpha-1 (**CTNNA1**), echinoderm microtubule-associated protein-like 4 (**EML4**), ferritin heavy chain 1 (**FTH1**), **GANAB**, , and syntaxin-12 (**STX12**) [100].

De Marchi et al., profiled ***ER+*** breast cancer patients, by comparing primary tumors with and without lymph node involvement to their matched normal noncancerous tissues. Their study contributed to the understanding of important functional insights of proteins expressed among the ER+ tamoxifen resistant population. The analysis has obtained more than 9000 proteins, where they were able to identify four important proteins namely, cingulin (**CGN**), Ras GTPase-activating protein-binding protein 2 (**G3BP2**), the programmed cell death protein 4 (**PDCD4**), and ovarian carcinoma immunoreactive antigen domain-containing protein 1 (**OCIAD1**). They suggested these to be potential biomarkers in predicting prognosis for the tamoxifen-susceptibility in recurrent breast cancer [101].

Tyanova et al., used SILAC MS labelling approach in the analysis of luminal ***ER+, HER2+,*** and ***TNBC*** tumor subtypes. Their team used support vector machine (SVM)-based classification for the 10,000 protein groups they identified from 40 tumor tissue samples. The functional proteomic profiles from theses samples were able to distinguish one subtype from another such as proteins related to cell growth, cell–cell communication energy metabolism, and mRNA translation. From this analysis, they identified ***19 specific proteins*** to differentiate between the breast cancer subtypes [102].

Some studies have found integration of genomics and proteins effective in breast cancer subtyping. The Clinical Proteomic Tumor Analysis Consortium (CPTAC) breast cancer study with 77 breast cancer samples from the TCGA cohort [103] accurately quantified a total depth of about 11,000 proteins from combined method of iTRAQ labelling approach and fractionation of samples. Comparison between the ***intrinsic breast cancer subtypes*** with the mRNA intrinsic subtypes found relatively similar abundances in the hormone receptors ER, PR, and HER2, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform (**PK3CA**), GATA-binding 3 (**GATA3**), and cellular tumor antigen p53 (**p53**). They were also able to show an unsupervised clustering of samples based on the proteomic level and identified these into three main clusters – luminal enriched, basal enriched, and stromal enriched. Their findings revealed that very similar subtype-defining features can be observed in both RNA-seq and labeled MS protein analysis although different tissue sections of the same tumors were used [103].

Johansson et al., by applying nanoLC-MS/MS method have accurately identified about 14,000 proteins, and 13,000 genes [104]. From these data 9995 of these genes identified across all 45 breast cancer samples were used for unsupervised clustering and arrived at 6 distinct proteome-based consensus core tumor clusters (**CoTC**). These clusters were closely associated to previously determined subtype in 50 transcripts (PAM50). Using the PAM50 subtyping, the clusters identified are **CoTC1** and **CoTC2** (basal-like tumors), **CoTC3** (luminal A), **CoTC4** and **CoTC6** (dominated by luminal and HER2+ tumors), and **CoTC5** (normal-like breast cancer). On the other hand, the depth and quality of proteome profiling of these clusters led to identify novel immunohistochemical biomarker candidates can help achieve better patient stratification approach to treatment. Moreover, their discovery also identified the link between tumor extracellular matrix composition to immune cell infiltration to prognosis, established a proteome-based framework for assessing prognosis, and discovered neoantigens to improve breast cancer therapies [104].

Bouchal et al. used SWATH-MS to analyze 96 breast cancer tissues and cell lines. . In their method, a spectral library was generated from breast cancer tissues and cells lines as a prerequisite to their breast cancer sample analysis [105]. In their findings they were able to identify about 2,800 protein groups, observed differences over the intrinsic breast cancer tumor subtypes, and discovered a greater depth of proteome variability between different breast cancer subtypes. Three differentially expressed proteins are considered to contribute strongly to improve the present breast cancer classification including **CDK1**, inositol polyphosphate 4-phosphatase type II (**INPP4B**), and **ERBB2**. Among the proteins found in the datasets, these proteins have shown the strongest correlation at the protein and transcript levels. This study was able to generate the first and only breast cancer spectral library, along with other proteomic datasets, have provided abundant sources of information for potential biomarkers research in breast cancer [105].

Rare breast cancer subtypes like mucinous carcinoma, cribform carcinoma and tubular carcinoma respond well to few endocrine therapies. Most of the endocrine responsive subtypes usually belong to luminal A subtypes and achieve good prognosis even without neoadjuvant chemotherapy [106]. However, the following rare subtypes namely metaplastic, apocrine, adenoid cystic and medullary do not have good prognosis to endocrine therapies [107]. Although there is a small proportion of patients under these categories the chance to find specific biomarkers for these rare breast cancer subtypes is possible in the near future. Table 3, presents these current findings in summary.

F. Recent breast cancer biomarkers discovery by different MS based technologies

Mass spectrometry applied in breast cancer research has led significant findings of proteins. Blood serum of breast cancer patients was analyzed in ***tandem mass spectrometry*** and the study revealed protein disulfide isomerase family A, member 3 (**PDIA3**) as minimally invasive protein markers for breast cancer. PDIA3 participates in the formation of the major histocompatibility complex (MHC) class I peptide loading complex which is involved in the formation of antigens. PDIA3 was also found in MCF-7 breast cancer cell line and patients with metastatic breast cancer [108], [109] having high expression. The elevated concentration of PDIA3 is primarily due to the cell's response to stress at breast cancer state. In other findings, they discovered that such high expression of PDIA3 is correlated to TP53 gene mutation and high expression of Ki-67 antigen, which are all known to enhance proliferation of cancer cells and growth [109]–[111]. Furthermore, individuals that have TP53 germline mutation are found to be at higher risk for early-onset breast cancer [112]. A study among Malaysian women was able to use the same MS approach and found novel proteins that are differently expressed in either stage 2 or stage 3 breast cancer cohorts only [113]. The stage

Table 3. Potential protein biomarkers for breast cancer molecular subtyping

Proteins quantified	No. of sample	Proteins differentially expressed	Significant findings	Reference
1000	6	200	Potential biomarkers for ER- BC - fascin, death-associated protein 5, Iprin-α1, and β-arrestin 1	[98]
18	18	298	Proteins involved in alterations of downstream transcription factors (TFs) in malignant tumor tissues	[99]
3500	123	10	(**AIFM1**), (**AP1G1**), (**AP1M1**), (**CAPZB**), (**CMPK1**), (**CTNNA1**), (**EML4**), (**FTH1**), **GANAB**, (**STX12**) were found on distant metastatic TNBC	[100]
9000	38	4	(CGN), (G3BP2), (PDCD4), and (OCIAD1). are potential biomarkers for the tamoxifen-susceptibility in recurrent BC	[101]
10000	40	19	Developed SVM based BC subtyping	[102]
11000	77		Proteomic analysis was comparable to mRNA BC subtyping	[103]
14000	45	9995	Identified CoTC for BC subtyping, comparably similar to PAM50 gene subtyping	[104]
2800	96	3	**CDK1**, (**INPP4B**), and **ERBB2** are strongly correlated at the protein and transcript levels in BC	[105]

2 breast cancer tumor when compared to its normal adjacent tissue shows to have high expressions of the **prolyl 3-hydroxylase 1 (P3H1), transmembrane emp24 domain-containing protein 10 (TMED10), peptidyl-prolyl cis-trans isomerase FKBP10 (FKB10), peptidyl-prolyl cis-trans isomerase FKBP9 (FKBP9), immunoglobulin superfamily containing leucine-rich repeat protein (ISLR), MOB kinase activator 1A (MOB1A), protein enabled homolog (ENAH), collagen alpha-1(V) chain (COL5A1), CAP-Gly domain-containing linker protein 1 (CLIP1), protein canopy homolog 4 (COL4A1), perilipin-4 (PLIN4), and zinc finger CCCH domain-containing protein 18 (C3H18).** These proteins in the stage 2 breast cancer tumor tissues are involved in many signaling pathways linked to invasion, proliferation, and migration. Protein found significantly expressed in the stage 3 breast cancer tumor includes **TAR DNA-binding protein 43 (TADBP), V-type proton ATPase subunit E 1 (VATE1), eukaryotic peptide chain release factor subunit 1 (ETF1), nucleoside diphosphate kinase 3 (NME3), deoxynucleoside triphosphate triphosphohydrolase SAMHD1 (SAMH1), protein SEC13 homolog (SEC13), protein enabled homolog (ENAH), DNA-dependent protein kinase catalytic subunit (PRKDC), golgi resident protein (GCP60), transmembrane glycoprotein NMB (GPNMB), rho GTPase-activating protein 1 (RHG01), LEM domain-containing protein 2 (LEMD2), prefoldin subunit 1, coiled-coil domain-containing protein 58 (CCCD58), inhibitor of nuclear factor kappa-B kinase-interacting protein (IKIP), MOB kinase activator 1A (MO1A), and MOB kinase activator 1B (MOB1B)**. Several among these proteins in the stage 3 breast cancer tumor tissues are found to function in supporting metastasis. Going for smaller quantities, a nanoLC-MS/MS technology was used in another study and revealed that EGF-like repeat and discoidin I-like domain-containing protein 3 (**EDIL3**) was elevated in the circulating extracellular vesicles [114]. The protein was observed to have a critical role in the integrin-FAK signaling cascade. When EDIL3 is inactivated in the MDA-MB-231 breast cancer cell lines integrin-FAK signaling pathway was also suppressed. The inactivity of EDIL3 resulted in controlling intracellular signal transduction which minimizes potential cellular invasion.

Phosphorylation of proteins is involved in the regulatory process in cells for proper functioning to occur. Phosphoproteins from biofluids are considered in many related studies as potential biomarkers for breast cancer. However, it is very challenging to find them in their native forms since they are usually denatured due to the presence of phosphatases. This is also the reason why perhaps most phosphoproteins are never detected that easy [115]. But in a recent study of Chen, et.al., they have successfully presented in their study a strategy to discover these phosphoproteins in human plasma by isolating them first from extracellular vesicles (EV) instead [115]. **SILAC** method was also employed in another phosphoproteomic study of kinase suppressor of ras-1 (**KSR1**) regulated phosphoprotein isolated from MCF-7 cell lines [116]. KSR1 phosphoprotein belongs to the RAF family of pseudokinases which is considered to play prominent roles in breast cancer growth. In the laboratory, KSR1 was observed to activate **B-Raf (BRAF) proto-oncogene** catalytic activity upon the binding of the formed protein complex between KSR1 and **mitogen-activated protein kinase (MEK)**. BRAF upon activation is directly involved in many cellular regulatory processes including cell proliferation and regulation of transcription. A mutant BRAF gene also results to a malfunctioning protein **B-raf** and has been implicated in several cancer studies including breast cancer [117].

In a study done by Jiang et al., MDA-MB-231 metastatic breast cancer cell lines were labeled by **iTRAQ** [118] to study the properties of Ras-related protein Rab-1B (**RAB1B**). Mechanistically, it was found that a down-regulation of RAB1B in cell lines activated TβR signaling pathway causing the expression of TGF-β receptor 1 (**TGFβR1**) protein levels to be elevated. Though little is

known about **TGFβR1**, one study supported it role as a tumor suppressor in the early stages of breast cancer, yet ironically it is also the same oncoprotein that may promote growth of tumor among invasive breast cancer patients [119]. In another study that also used iTRAQ was able to profile metastatic breast cancer tissues and revealed that decorin (**DCN**) and the heat shock protein 90 beta family member 1 (**HSP90B1**) as potential breast cancer biomarkers. DCN and HSP90B1 were both associated in biological pathways related to tumorigenesis promoting cell proliferation, migration, and transcription. DCN's oncogenic role was found related to tumor microenvironment showing significant interactions with EGFR and MAPK. HSP90B1 overexpression was best illustrated in a study of Huang, et al employing proteogenomics integration. They suggested that genomics alone with proteomics diagnosed by MS is not enough to show some of the phosphoprotein events that are involved in the regulatory events in the cell. Both DCN and HSP90B1 proteins are continuously being investigated at the gene and protein levels to study their roles in promoting tumor invasion and metastasis especially among TNBC patients [118]. At present, these oncoproteins are also considered indicators for poor breast cancer prognosis. The study of Lawrence et al., combined proteomics with genomic aberrations affecting protein expression [124]. They successfully discovered potential biomarkers for drug sensitivity screening based on combining the strengths of these two methods [124].

TMT can also be effective with enrichment protocols for deeper phosphoproteome analysis. Using **6-plex TMT**, Chen et al., incorporated MS approach in their drug discovery study to investigate the expression inhibition effect on the C-X-C chemokine receptor type 4 (**CXCR4**) by the **ginsenoside Rg3**, a compound extracted from Panax ginseng [120]. MDA-MB-231 cell line treated with ginsenoside Rg3 peptides were labeled with 6-plex TMT. The mixture was divided into fractions and each fraction subjected to TiO_2-based phosphopeptide enrichment before the LC-MS/MS analysis. The protein analysis showed that the Rg3 protein were found to display an inhibitory effect against CXCR4 and targets the anti-inflammatory nuclear factor-kB signaling pathway [121]. These findings suggested that ginsenoside Rg3 have chemopreventive properties that may control metastasis of breast cancer cells.

Staging breast cancer is another interesting aspect for proteomic study. In the lab of Lobo et. al., they investigated using **DIA label-free MS method** to analyze protein expression from patient blood samples at different breast cancer stages. Three specific proteins were found to have varied expressions at different stages of breast cancer. These proteins are **clusterin, apolipoprotein A-II (APOA2),** and **apolipoprotein C-III (APOC3).** These proteins play important roles in alterations related to protein glycosylation, progression and invasiveness of breast cancer [122]. Aside from tissues and blood samples, some studies made use of urine to detect biomarkers in label-free MS analysis. In the study of Beretov et al., successfully identified 13 proteins from urine for its potential utility to detect preinvasive breast cancer stage to metastatic breast cancer stage among DCIS patients [123]. The novel proteins they discovered from the tissue samples include **cytosolic non-specific dipeptidase (PEPA), multimerin-2 (MMRN2), neuronal growth regulator 1 (NEGR1), leucine-rich repeat-containing protein 36 (LRC36), microtubule-associated serine/threonine-protein kinase 4 (MAST4), keratin, type I cytoskeletal 10 (K1C10), uncharacterized protein C9orf131 (CI131), uncharacterized protein C4orf14 (CD014), filaggrin, dynein heavy chain 8, axonemal (DYH8), hemoglobin subunit alpha (HBA), AGRIN, and fibrinogen alpha chain (FIBA).** From this list, they tested their hypothesis only against MAST4 to validate the protein profile in both tissue and urine samples. The result of their study was able to detect the protein biomarkers in both body derivatives.

These are just some of the many on-going research that are exploring the application of various MS techniques and strategies that will lead to the discovery of novel protein tumor biomarkers for breast cancer. It is important to note that at this rate, mass spectrometry-based research is going very fast in integrating label-free MS methods together with library free computational tools approach to drive it workable utility in the clinical setting.

CONCLUSIONS AND FUTURE DIRECTIONS

The high throughput proteomic methods applied in cancer research have been evolving dramatically at an increasing pace. The majority of these protein profiling studies exhibit promising diagnostic, prognostic or predictive values in controlling and mitigating breast cancer. More study is in place to unravel breast cancer's nature, molecular features, and tumor biology to explain its uncontrolled growth and development. By using mass spectrometry and its computational tools for identification and quantitation it would not be long that soon this method would completely be used for clinical applications.

At present, breast cancer intrinsic protein biomarkers are routinely measured primarily by IHC using antibody-based techniques or by genomic-based test. Although both mRNA extraction kits require highly specific antibodies, genetic tests are also costly and are often effective only to patients expressing hormone receptors. IHC may also see some issues wherein the immunoreactivity can be compromised because of post-translational modifications of the proteins. Furthermore, without quantifiable biomarkers like in the case of the TNBC subtype, it would be challenging to come up with an exact regimen for medical treatment.

Today, mass spectrometry has seriously been reengineered to full capability to analyze down to

nanoparticle size samples at high throughput. The instrument and software developed for database search and analysis has remarkably enabled us to investigate in-depth breast cancer more effectively. It was also noted earlier, that there are different proteins involved in the development of the mammary gland, but very few are published about the proteomic landscape of breast cancer patients from young, premenopausal, to postmenopausal adults. Also, staging has been noted to be well studied even in mass spectrometry, but the proteomic profile accounting for breast cancer grades has not been well studied using this approach. These missing links in the biology of breast cancer may soon unfold especially with updated databases uploaded in the human proteome atlas (HPA) [124]. This open-access online platform is an advantage for researchers to explore the properties and functions of every protein. The HPA database includes a list of about 17000 unique proteins, 26000 antibodies, protein cellular locations, functions, cellular processes, and disease associated proteins including those involved in breast cancer [124]. The information in this database was mostly acquired from mass spectrometry-based proteomics data acquisition, other omics-based technologies and a variety of immuno-based methods. With new information that MS technology can derived, more novel proteins are to be found and mapped to networks and pathways associated to breast cancer development and use them for effective drug design and therapy.

.

REFERENCES

[1] WHO, *Who Report on Cancer*. 2020.
[2] NCI, "NCI Center to Reduce Cancer Health Disparities," 2019. [Online]. Available: https://www.cancer.gov/about-nci/organization/crchd. [Accessed: 28-May-2019].
[3] A.Cooper, *On the anatomy of the breast*. Jefferson, 1840.
[4] M. L.Marinovich *et al.*, "Agreement between MRI and pathologic breast tumor size after neoadjuvant chemotherapy, and comparison with alternative tests: individual patient data meta-analysis," *BMC Cancer*, vol. 15, p. 662, Oct.2015.
[5] M.Mayo Clinic, "Breast Cancer," *2019*, 2019. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/breast-cancer/diagnosis-treatment/drc-20352475.
[6] C. J.Watson andW. T.Khaled, "Mammary development in the embryo and adult: a journey of morphogenesis and commitment," *Development*, vol. 135, no. 6, pp. 995 LP – 1003, Mar.2008.
[7] J. H.Krege *et al.*, "Generation and reproductive phenotypes of mice lacking estrogen receptor beta," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 26, pp. 15677–15682, Dec.1998.
[8] W.Ruan andD. L.Kleinberg, "Insulin-Like Growth Factor I Is Essential for Terminal End Bud Formation and Ductal Morphogenesis during Mammary Development1," *Endocrinology*, vol. 140, no. 11, pp. 5075–5081, Nov.1999.
[9] J.McBryan, J.Howlin, S.Napoletano, andF.Martin, "Amphiregulin: Role in Mammary Gland Development and Breast Cancer," *J. Mammary Gland Biol. Neoplasia*, vol. 13, pp. 159–169, Jul.2008.
[10] S. R.Oakes, H. N.Hilton, andC. J.Ormandy, "The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium," *Breast Cancer Res.*, vol. 8, no. 2, p. 207, 2006.
[11] T. R.Milanese *et al.*, "Age-related lobular involution and risk of breast cancer.," *J. Natl. Cancer Inst.*, vol. 98, no. 22, pp. 1600–1607, Nov.2006.
[12] V.Sangwan andM.Park, "Receptor tyrosine kinases: role in cancer progression," *Curr. Oncol.*, vol. 13, no. 5, pp. 191–193, Oct.2006.
[13] S. P.Chiang, G. S.Karagiannis, J. S.Condeelis, andJ. E.Segall, "Abstract 1844: Investigating the role of amphiregulin in breast cancer," *Cancer Res.*, vol. 77, no. 13 Supplement, pp. 1844 LP – 1844, Jul.2017.
[14] S.Lee *et al.*, "Alterations of gene expression in the development of early hyperplastic precursors of breast cancer.," *Am. J. Pathol.*, vol. 171, no. 1, pp. 252–262, Jul.2007.
[15] S.Mao *et al.*, "Abstract 5114: The role of amphiregulin in mammary gland development and breast cancer," *Cancer Res.*, vol. 78, no. 13 Supplement, pp. 5114 LP – 5114, Jul.2018.
[16] A.Baillo, C.Giroux, andS. P.Ethier, "Knock-down of amphiregulin inhibits cellular invasion in inflammatory breast cancer.," *J. Cell. Physiol.*, vol. 226, no. 10, pp. 2691–2701, Oct.2011.
[17] S.Rao, S. J. F.Cronin, V.Sigl, andJ. M.Penninger, "RANKL and RANK: From Mammalian Physiology to Cancer Treatment," *Trends Cell Biol.*, vol. 28, no. 3, pp. 213–223, 2018.
[18] E.Gonzalez-Suarez *et al.*, "RANK ligand mediates progestin-induced mammary epithelial proliferation and carcinogenesis.," *Nature*, vol. 468, no. 7320, pp. 103–107, Nov.2010.
[19] D.Schramek *et al.*, "Osteoclast differentiation factor RANKL controls development of progestin-driven mammary cancer.," *Nature*, vol. 468, no. 7320, pp. 98–102, Nov.2010.
[20] D. C.Radisky andL. C.Hartmann, "Mammary involution and breast cancer risk: transgenic models and clinical studies," *J. Mammary Gland Biol. Neoplasia*, vol. 14, no. 2, pp. 181–191, Jun.2009.
[21] R. W. E.Clarkson andC. J.Watson, "Microarray analysis of the involution switch.," *J. Mammary Gland Biol. Neoplasia*, vol. 8, no. 3, pp. 309–319, Jul.2003.
[22] L. T.Bemis andP.Schedin, "Reproductive state of rat mammary gland stroma modulates human breast cancer cell migration and invasion.," *Cancer Res.*, vol. 60, no. 13, pp. 3414–3418, Jul.2000.
[23] M.Xue *et al.*, "Regulation of estrogen signaling and breast cancer proliferation by an ubiquitin ligase TRIM56," *Oncogenesis*, vol. 8, no. 5, p. 30, 2019.
[24] J.Wang andB.Xu, "Targeted therapeutic options and future perspectives for HER2-positive breast

cancer," *Signal Transduct. Target. Ther.*, vol. 4, no. 1, p. 34, 2019.
[25] L.Xia *et al.*, "Role of the NFκB-signaling pathway in cancer," *Onco. Targets. Ther.*, vol. 11, pp. 2063–2073, Apr.2018.
[26] T. H. B.Gomig *et al.*, "Quantitative label-free mass spectrometry using contralateral and adjacent breast tissues reveal differentially expressed proteins and their predicted impacts on pathways and cellular functions in breast cancer," *J. Proteomics*, vol. 199, no. August 2018, pp. 1–14, 2019.
[27] E.Presti, Daniele and Quaquarini, "The PI3K/AKT/mTOR and CDK4/6 Pathways in Endocrine Resistant HR+/HER2− Metastatic Breast Cancer: Biological Mechanisms and New Treatments," *Cancers (Basel).*, vol. 11, no. 9, pp. 1–20, 2019.
[28] B. M.Krishna *et al.*, "Notch signaling in breast cancer: From pathway analysis to therapy," *Cancer Lett.*, vol. 461, pp. 123–131, 2019.
[29] M. A.Gillette *et al.*, "Proteogenomics connects somatic mutations to signalling in breast cancer," *Nature*, pp. 1–19, 2016.
[30] H. J.Burstein, "The distinctive nature of HER2-positive breast cancers.," *N. Engl. J. Med.*, vol. 353, no. 16, pp. 1652–1654, Oct.2005.
[31] NIH, "Breast Cancer," 2019. [Online]. Available: https://ghr.nlm.nih.gov/condition/breast-cancer#resources.
[32] K. A.Metcalfe *et al.*, "The risk of breast cancer in BRCA1 and BRCA2 mutation carriers without a first-degree relative with breast cancer.," *Clin. Genet.*, vol. 93, no. 5, pp. 1063–1068, May2018.
[33] Z.Mitri, T.Constantine, andR.O'Regan, "The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy," *Chemother. Res. Pract.*, vol. 2012, p. 743193, 2012.
[34] M. C.Figueroa-Magalhães, D.Jelovac, R.Connolly, andA. C.Wolff, "Treatment of HER2-positive breast cancer," *Breast*, vol. 23, no. 2, pp. 128–136, Apr.2014.
[35] M.Dowsett *et al.*, "Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group," *J. Natl. Cancer Inst.*, vol. 103, no. 22, pp. 1656–1664, Nov.2011.
[36] M.Abubakar *et al.*, "Combined quantitative measures of ER , PR , HER2 , and KI67 provide more prognostic information than categorical combinations in luminal breast cancer," *Mod. Pathol.*, pp. 1244–1256, 2019.
[37] F. J.Candido *et al.*, "An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation," pp. 1–13, 2017.
[38] A.Vaishnavi, A. T.Le, andR. C.Doebele, "TRKing down an old oncogene in a new era of targeted therapy.," *Cancer Discov.*, vol. 5, no. 1, pp. 25–34, Jan.2015.
[39] J. J.Lee, K.Loh, andY.-S.Yap, "PI3K/Akt/mTOR inhibitors in breast cancer," *Cancer Biol. Med.*, vol. 12, no. 4, pp. 342–354, Dec.2015.
[40] O.Willis *et al.*, "PIK3CA gene aberrancy and role in targeted therapy of solid malignancies," *Cancer Gene Ther.*, 2020.
[41] Y. F.Hou, M.F., Chen, Y.L., Tseng, T.F., Lin, C.M., Chen, M.S., Huang, C.J., Huang, Y.S., Hsieh, J.S., Huang, T.J., Jong, S.B., Huang, "Evaluation of serum CA27.29, CA15-3 and CEA in patients with breast cancer," *Kaohsiung J. Med. Sci.*, 1999.
[42] C.VanPoznak, L. N.Harris, andM. R.Somerfield, "Use of Biomarkers to Guide Decisions on Systemic Therapy for Women With Metastatic Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline," *J. Oncol. Pract.*, vol. 11, no. 6, pp. 514–516, Jul.2015.
[43] A. M.Kabel, "Tumor markers of breast cancer: New prospectives," *J. Oncol. Sci.*, vol. 3, no. 1, pp. 5–11, 2017.
[44] X.Zhe, M. L.Cher, andR. D.Bonfil, "Circulating tumor cells: finding the needle in the haystack," *Am. J. Cancer Res.*, vol. 1, no. 6, pp. 740–751, 2011.
[45] W. A.Osta *et al.*, "EpCAM Is Overexpressed in Breast Cancer and Is a Potential Target for Breast Cancer Gene Therapy," *Cancer Res.*, vol. 64, no. 16, pp. 5818 LP – 5824, Aug.2004.
[46] S.Cotterill, "Cancer Genetics Web," 2019. [Online]. Available: http://www.cancer-genetics.org/PTPRC.htm. [Accessed: 10-Jul-2020].
[47] E.Jarasch, R. A. Y. B.Nagle, M.Kaufmann, C.Maurer, andW. J.Bocker, "Differential Diagnosis of Benign Epithelial Proliferations and Carcinomas of the Breast Using Antibodies to Cytokeratins," 1987.
[48] Y.Jing, J.Zhang, S.Waxman, andR.Mira-y-Lopez, "Upregulation of cytokeratins 8 and 18 in human breast cancer T47D cells is retinoid-specific and retinoic acid receptor-dependent.," *Differentiation.*, vol. 60, no. 2, pp. 109–117, May1996.
[49] M. B.Lustberg *et al.*, "Heterogeneous atypical cell populations are present in blood of metastatic breast cancer patients.," *Breast Cancer Res.*, vol. 16, no. 2, p. R23, Mar.2014.
[50]C. P .Leo, C.Leo, andT. D.Szucs, "Breast cancer drug approvals by the US FDA from 1949 to 2018," *Nat. Rev. Drug Discov.*, vol. 19, no. 1, p. 11, 2020.
[51] A.Garcia-Diaz *et al.*, "Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression.," *Cell Rep.*, vol. 19, no. 6, pp. 1189–1201, May2017.
[52] Y.Han, D.Liu, andL.Li, "PD-1/PD-L1 pathway: current researches in cancer," *Am. J. Cancer Res.*, vol. 10, no. 3, pp. 727–742, Mar.2020.
[53] F.Schütz, S.Stefanovic, L.Mayer, A.VonAu, C.Domschke, andC.Sohn, "PD-1/PD-L1 Pathway in Breast Cancer," *Oncol. Res. Treat.*, vol. 40, no. 5, pp. 294–297, 2017.
[54] A.Illiano, G.Pinto, C.Melchiorre, A.Carpentieri, V.Faraco, andA.Amoresano, "Protein Glycosylation Investigated by Mass Spectrometry: An Overview," *Cells*, vol. 9, no. 9, pp. 1–22, 2020.
[55] K. A.Resing andN. G.Ahn, "Proteomics strategies for protein identification," vol. 579, pp. 885–889, 2005.
[56] G.Büyükköroğlu, D. D.Dora, F.Özdemir, andC.Hızel, "Chapter 15 - Techniques for Protein Analysis," D.Barh andV. B. T.-O. T. and B.-E.Azevedo, Eds.Academic Press, 2018, pp. 317–351.

[57] T. J.Griffin andR.Aebersold, “Advances in Proteome Analysis by Mass Spectrometry,” *J. Biol. Chem.*, vol. 276, no. 49, pp. 45497–45500, 2001.
[58] S. B.Nukala, G.Baron, G.Aldini, M.Carini, andA.D’Amato, “Mass Spectrometry-based Label-free Quantitative Proteomics To Study the Effect of 3PO Drug at Cellular Level.,” *ACS Med. Chem. Lett.*, vol. 10, no. 4, pp. 577–583, Apr.2019.
[59] M. W.Duncan andS. W.Hunsucker, “Proteomics as a tool for clinically relevant biomarker discovery and validation.,” *Exp. Biol. Med. (Maywood).*, vol. 230, no. 11, pp. 808–817, Dec.2005.
[60] F. M.White, “Quantitative phosphoproteomic analysis of signaling network dynamics.,” *Curr. Opin. Biotechnol.*, vol. 19, no. 4, pp. 404–409, Aug.2008.
[61] A. P.Frei *et al.*, “Direct identification of ligand-receptor interactions on living cells and tissues.,” *Nat. Biotechnol.*, vol. 30, no. 10, pp. 997–1001, Oct.2012.
[62]R.Aebersold andM.Mann, “Mass spectrometry-based proteo mics,” *Nature*, vol. 422, no. 6928, pp. 198–207, 2003.
[63] F.Desiere *et al.*, “The PeptideAtlas project.,” *Nucleic Acids Res.*, vol. 34, no. Database issue, pp. 655–658, 2006.
[64] G.Rosenberger *et al.*, “A repository of assays to quantify 10,000 human proteins by SWATH-MS,” *Sci. Data*, vol. 1, no. 1, p. 140031, 2014.
[65] P.Palmowski *et al.*, “The Generation of a Comprehensive Spectral Library for the Analysis of the Guinea Pig Proteome by SWATH-MS,” *Proteomics*, vol. 19, no. 15, pp. 1–5, 2019.
[66] Y.Lin, “In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics,” *Nat. Commun.*, pp. 1–11, 2020.
[67] E.Hoedt, G.Zhang, andT. A.Neubert, “Stable isotope labeling by amino acids in cell culture (SILAC) for quantitative proteomics.,” *Adv. Exp. Med. Biol.*, vol. 806, pp. 93–106, 2014.
[68] S.Wiese, K. A.Reidegeld, H. E.Meyer, andB.Warscheid, “Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research.,” *Proteomics*, vol. 7, no. 3, pp. 340–350, Feb.2007.
[69] J.Muntel *et al.*, “Comparison of Protein Quanti fi cation in a Complex Background by DIA and TMT Work fl ows with Fixed Instrument Time,” 2019.
[70] L. R.Zieske, “A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies.,” *J. Exp. Bot.*, vol. 57, no. 7, pp. 1501–1508, 2006.
[71] L.Zhang andJ. E.Elias, “Relative Protein Quantification Using Tandem Mass Tag Mass Spectrometry BT - Proteomics: Methods and Protocols,” L.Comai, J. E.Katz, andP.Mallick, Eds.New York, NY: Springer New York, 2017, pp. 185–198.
[72] D.Virág *et al.*, “Current Trends in the Analysis of Post-translational Modifications,” *Chromatographia*, vol. 83, no. 1, pp. 1–10, 2020.
[73] B.Manadas, V. M.Mendes, J.English, andM. J.Dunn, “Peptide fractionation in proteomics approaches.,” *Expert Rev. Proteomics*, vol. 7, no. 5, pp. 655–663, Oct.2010.
[74] E. R.Sauter, “Reliable Biomarkers to Identify New and Recurrent Cancer,” *Eur. J. breast Heal.*, vol. 13, no. 4, pp. 162–167, Oct.2017.
[75] E. M.Woo, D.Fenyo, B. H.Kwok, H.Funabiki, andB. T.Chait, “Efficient Identification of Phosphorylation by Mass Spectrometric Phosphopeptide Fingerprinting,” *Anal. Chem.*, vol. 80, no. 7, pp. 2419–2425, Apr.2008.
[76] S. A.Beausoleil *et al.*, “Large-scale characterization of HeLa cell nuclear phosphoproteins,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 33, pp. 12130–12135, 2004.
[77] C.Ludwig, L.Gillet, G.Rosenberger, S.Amon, B. C.Collins, andR.Aebersold, “Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial,” *Mol. Syst. Biol.*, vol. 14, no. 8, p. e8126, Aug.2018.
[78] Y. S.Ting *et al.*, “Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data.,” *Mol. Cell. Proteomics*, vol. 14, no. 9, pp. 2301–2307, Sep.2015.
[79] S.Tyanova, T.Temu, andJ.Cox, “The MaxQuant computational platform for mass spectrometry-based shotgun proteomics,” *Nat. Protoc.*, vol. 11, p. 2301, Oct.2016.
[80] O.Bernhardt *et al.*, *Spectronaut: a fast and efficient algorithm for MRM-like processing of data independent acquisition (SWATH-MS) data*. 2014.
[81] R. C.Poulos *et al.*, “Strategies to enable large-scale proteomics for reproducible research,” *Nat. Commun.*, vol. 11, no. 1, pp. 1–13, 2020.
[82] Institute for Systems Biology, “Spectrum Library Central at PeptideAtlas,” 2015. [Online]. Available: http://www.peptideatlas.org/speclib/. [Accessed: 08-Oct-2019].
[83] H. L.Röst *et al.*, “OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data.,” *Nature biotechnology*, vol. 32, no. 3. United States, pp. 219–223, Mar-2014.
[84] J.Wang *et al.*, “MSPLIT-DIA: sensitive peptide identification for data-independent acquisition.,” *Nature methods*, vol. 12, no. 12. pp. 1106–1108, Dec-2015.
[85] Y.Yang, X.Liu, C.Shen, Y.Lin, P.Yang, andL.Qiao, “In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics,” *Nat. Commun.*, vol. 11, no. 1, p. 146, Jan.2020.
[86] S.Gessulat *et al.*, “Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning,” *Nat. Methods*, vol. 16, no. 6, pp. 509–518, 2019.
[87] N. H.Tran *et al.*, “Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry.,” *Nat. Methods*, vol. 16, no. 1, pp. 63–66, Jan.2019.
[88] A. M.Frank, M. M.Savitski, M. L.Nielsen, R. A.Zubarev, andP. A.Pevzner, “De novo peptide sequencing and identification with precision mass spectrometry.,” *J. Proteome Res.*, vol. 6, no. 1, pp. 114–123, Jan.2007.

[89] R. S.Johnson andJ. A.Taylor, “Searching sequence databases via De novo peptide sequencing by tandem mass spectrometry,” *Mol. Biotechnol.*, vol. 22, no. 3, pp. 301–315, 2002.
[90] B.Ma *et al.*, “PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.,” *Rapid Commun. Mass Spectrom.*, vol. 17, no. 20, pp. 2337–2342, 2003.
[91] Y.Li *et al.*, “Group-DIA: analyzing multiple data-independent acquisition mass spectrometry data files.,” *Nature methods*, vol. 12, no. 12. United States, pp. 1105–1106, Dec-2015.
[92] Y. S.Ting *et al.*, “PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data.,” *Nat. Methods*, vol. 14, no. 9, pp. 903–908, Sep.2017.
[93] C.-C.Tsou *et al.*, “DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics.,” *Nat. Methods*, vol. 12, no. 3, pp. 258–64, 7 p following 264, Mar.2015.
[94] B. C.Searle *et al.*, “Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry,” *Nat. Commun.*, vol. 9, no. 1, p. 5128, 2018.
[95] P.Pandit, R.Patil, V.Palwe, S.Gandhe, R.Patil, andR.Nagarkar, “Prevalence of Molecular Subtypes of Breast Cancer: A Single Institutional Experience of 2062 Patients,” *Eur. J. breast Heal.*, vol. 16, no. 1, pp. 39–43, Nov.2019.
[96] C. K. Y.Ng, A. M.Schultheis, F.-C.Bidard, B.Weigelt, andJ. S.Reis-Filho, “Breast cancer genomics from microarrays to massively parallel sequencing: paradigms and new insights.,” *J. Natl. Cancer Inst.*, vol. 107, no. 5, Feb.2015.
[97] Cancer Genome Atlas Network, “Comprehensive molecular portraits of human breast tumours.,” *Nature*, vol. 490, no. 7418, pp. 61–70, Oct.2012.
[98] K.Rezaul *et al.*, “Differential protein expression profiles in estrogen receptor-positive and -negative breast cancer tissues using label-free quantitative proteomics,” *Genes and Cancer*, vol. 1, no. 3, pp. 251–271, 2010.
[99] B. L. K.S. Cha, M.B. Imielinski, T. Rejtar, E.A. Richardson, D. Thakur, D.C. Sgroi, “In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology Mol Cell Proteomics, 9 (2010), pp. 2529-2544,” 2010.
[100] N. Q.Liu *et al.*, “Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer,” *J. Natl. Cancer Inst.*, vol. 106, no. 2, 2014.
[101] A. U.T. De Marchi, N.Q. Liu, C. Stingl, M.A. Timmermans, M. Smid, M.P. Look, M. Tjoa, R.B. Braakman, M. Opdam, S.C. Linn, F.C. Sweep, P.N. Span, M. Kliffen, T.M. Luider, J.A. Foekens, J.W. Martens, “4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer Mol Oncol, 10 (2016), pp. 24-39,” 2016.
[102] S.Tyanova, R.Albrechtsen, P.Kronqvist, J.Cox, M.Mann, andT.Geiger, “Proteomic maps of breast cancer subtypes,” *Nat. Commun.*, vol. 7, no. 1, p. 10259, 2016.
[103] P. Mertins, D.R. Mani, K.V. Ruggles, M.A. Gillette, K.R. Clauser, P. Wang, “Proteogenomics connects somatic mutations to signalling in breast cancer Nature, 534 (2016), pp. 55-62,” 2016.
[104] H. J.Johansson *et al.*, “Breast cancer quantitative proteome and proteogenomic landscape,” *Nat. Commun.*, vol. 10, no. 1, p. 1600, 2019.
[105] P.Bouchal *et al.*, “Breast Cancer Classification Based on Proteotypes Obtained by SWATH Mass Spectrometry,” *Cell Rep.*, vol. 28, no. 3, pp. 832-843.e7, Jul.2019.
[106] T.Nagao, T.Kinoshita, T.Hojo, H.Tsuda, K.Tamura, andY.Fujiwara, “The differences in the histological types of breast cancer and the response to neoadjuvant chemotherapy: the relationship between the outcome and the clinicopathological characteristics.,” *Breast*, vol. 21, no. 3, pp. 289–295, Jun.2012.
[107] M. V.Dieci, E.Orvieto, M.Dominici, P.Conte, andV.Guarneri, “Rare Breast Cancer Subtypes : Histological , Molecular , and Clinical Peculiarities,” pp. 805–813, 2014.
[108] H.Zou *et al.*, “P4HB and PDIA3 are associated with tumor progression and therapeutic outcome of diffuse gliomas,” *Oncol. Rep.*, vol. 39, no. 2, pp. 501–510, Feb.2018.
[109] M.Song *et al.*, “Proteomic Analysis of Breast Cancer Tissues to Identify Biomarker Candidates by Gel-Assisted Digestion and Label-Free Quantification Methods Using LC-MS / MS,” vol. 35, no. 10, pp. 1839–1847, 2012.
[110] L. T.Li, G.Jiang, Q.Chen, andJ. N.Zheng, “Predic Ki67 is a promising molecular target in the diagnosis of cancer (Review),” *Mol. Med. Rep.*, vol. 11, no. 3, pp. 1566–1572, 2015.
[111] F.Mantovani, L.Collavin, andG.DelSal, “Mutant p53 as a guardian of the cancer cell,” *Cell Death Differ.*, vol. 26, no. 2, pp. 199–212, 2019.
[112] A. D.Sorrell, C. R.Espenschied, J. O.Culver, and J. N.Weitzel, “Tumor protein p53 (TP53) testing and Li-Fraumeni syndrome : current status of clinical applications and future directions,” *Mol. Diagn. Ther.*, vol. 17, no. 1, pp. 31–47, Feb.2013.
[113] A. S.Al-Wajeeh *et al.*, “Comparative proteomic analysis of different stages of breast cancer tissues using ultra high performance liquid chromatography tandem mass spectrometer,” *PLoS One*, vol. 15, no. 1, pp. e0227404–e0227404, Jan.2020.
[114] J.-E.Lee *et al.*, “Identification of EDIL3 on extracellular vesicles involved in breast cancer cell invasion,” *J. Proteomics*, vol. 131, pp. 17–28, 2016.
[115] I.-H.Chen *et al.*, “Phosphoproteins in extracellular vesicles as candidate markers for breast cancer,” *Proc. Natl. Acad. Sci.*, vol. 114, no. 12, pp. 3175 LP – 3180, Mar.2017.
[116] H.Zhang *et al.*, “SILAC-based phosphoproteomics reveals an inhibitory role of KSR1 in p53 transcriptional activity via modulation of DBC1,”

Br. J. Cancer, vol. 109, no. 10, pp. 2675–2684, Nov.2013.
[117] D.Liu andK.Zhou, “BRAF/MEK Pathway is Associated With Breast Cancer in ER-dependent Mode and Improves ER Status-based Cancer Recurrence Prediction,” *Clin. Breast Cancer*, vol. 20, no. 1, pp. 41-50.e8, 2020.
[118] H.-L.Jiang *et al.*, “Loss of RAB1B promotes triple-negative breast cancer metastasis by activating TGF-beta/SMAD signaling.,” *Oncotarget*, vol. 6, no. 18, pp. 16352–16365, Jun.2015.
[119] B.Tang *et al.*, “TGF-beta switches from tumor suppressor to prometastatic factor in a model of breast cancer progression,” *J. Clin. Invest.*, vol. 112, no. 7, pp. 1116–1124, Oct.2003.
[120] X. C. L.Qian, H.Jiang, andJ.Chen, “Ginsenoside Rg3 inhibits CXCR 4 expression and related migrations in a breast cancer cell line,” pp. 519–523, 2011.
[121] M.Zou, J.Wang, J.Gao, H.Han, andY.Fang, “Phosphoproteomic analysis of the antitumor effects of ginsenoside Rg3 in human breast cancer cells,” *Oncol. Lett.*, vol. 15, no. 3, pp. 2889–2898, Mar.2018.
[122] M. D. P.Lobo, F. B. M. B.Moreno, G. H. M. F.Souza, S. M. M. L.Verde, R. de A.Moreira, andA. C. de O.Monteiro-Moreira, “Label-free proteome analysis of plasma from patients with breast cancer: Stage-specific protein expression,” *Front. Oncol.*, vol. 7, no. FEB, pp. 1–12, 2017.
[123] J.Beretov, V. C.Wasinger, E. K. A.Millar, P.Schwartz, P. H.Graham, andY.Li, “Proteomic analysis of urine to identify breast cancer biomarker candidates using a label-free LC-MS/MS approach,” *PLoS One*, vol. 10, no. 11, 2015.
[124] “Human Proteome Atlas,” *2020*. [Online]. Available: https://www.proteinatlas.org/.