

**A PROTEOMIC ANALYSIS OF NEOPLASTIC PROGRESSION IN BREAST CANCER**

by

**Nicholas William Bateman**

B.S. Biological Sciences. SUNY at Buffalo. 2002

Submitted to the Graduate Faculty of  
Pharmacology and Chemical Biology in partial fulfillment  
of the requirements for the degree of Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

School of Medicine – Pharmacology and Chemical Biology

This dissertation was presented

by

Nicholas William Bateman

It was defended on

November 23<sup>rd</sup>, 2010

and approved by

Christopher Bakkenist, PhD, Department of Pharmacology and Chemical Biology

Rohit Bhargava, MD, Department of Pathology

Bruce A. Freeman, PhD, Department of Pharmacology and Chemical Biology

Saleem A. Khan, PhD, Department of Microbiology and Molecular Genetics

Advisor: Thomas P. Conrads, PhD, Department of Pharmacology and Chemical Biology

**A PROTEOMIC ANALYSIS OF NEOPLASTIC PROGRESSION IN BREAST  
CANCER**

Nicholas W. Bateman, BS

University of Pittsburgh, 2010

Copyright © by Nicholas W. Bateman

2010

## **ABSTRACT.**

The utilization of high-throughput -omics strategies, such as proteomics, in the analysis of breast cancer will function to define central molecular characteristics across a disease that is associated with a high degree of molecular heterogeneity. Data reported herein details the investigation of key subjects in breast cancer biology focused on the characterization of endogenous and experimentally-induced disease biology characteristics utilizing the application of LC-MS based proteomic analyses of both *in vitro* models of breast cancer as well as primary clinical samples. Results include a combined global and functional proteomic strategy to identify governing functional roles for mutually, differentially abundant proteins observed across three divergent cell line models of breast cancer. Further, evidence is presented which provides insights into the regulatory activity of the breast cancer-associated microRNA (miR-145) in several cell line models of breast cancer in which expression of this microRNA has been restored. Lastly, robust analyses are detailed focused on the identification of differential protein characteristics indicative of disease stage as well as of recurrent disease in breast cancer derived from proteomic analysis of formalin-fixed, paraffin embedded (FFPE) clinical samples. These studies contribute to the field of proteomics in the form of 1) providing robust experimental workflows directed towards investigation of functional themes and associated functional targets in large protein data sets 2) detailing strategies for navigating the application of proteomic analysis to microRNA target discovery and 3) further development and utilization of methodologies towards the proteomic analysis of clinical, FFPE tissue samples. Furthermore, these studies benefit the breast cancer community on several fronts including 1) the elucidation of provocative protein candidates which warrant further investigation for their role in regulating disease mechanisms underlying

breast cancer biology and 2) through the discovery of diagnostic markers indicative of discrete subtypes and stages of disease progression in breast cancer. The results reported herein detail disease-specific protein abundance characteristics associated with neoplastic progression in breast cancer that will benefit further expansion of the basic biological understanding of this disease and describes novel proteins for further evaluation as biomarker candidates for the diagnosis of breast cancer.

## TABLE OF CONTENTS

ABSTRACT.....	iv-v
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix-xi
1.0 INTRODUCTION Proteomics of Breast Cancer: Biomarker Discovery and Functional Analysis.....	1
1.1 Breast Cancer – Background.....	1
1.2 Proteomics – Background.....	8
1.3 Proteomics of Breast Cancer – Biomarker Discovery.....	15
1.4 Proteomics of Breast Cancer – Functional Proteomics.....	25
1.5 Introduction to Data Chapters.....	29
2.0 CHAPTER 1: Defining Central Themes in Breast Cancer Biology by Differential Proteomics: Conserved Regulation of Cell Spreading and Focal Adhesion Kinase.....	30
3.0 CHAPTER 2: Quantitative proteomic analysis of microRNA-145 in breast cancer.....	58

4.0	CHAPTER 3: Differential Proteomic Analysis of Late-Stage and Recurrent Breast Cancer from formalin-fixed, paraffin embedded tissues.....	79
5.0	CHAPTER 4: Differential Proteomics Analysis of Lymph-Node Metastasis in Estrogen-Receptor positive Breast Cancer from formalin-fixed, paraffin embedded Tissues.....	102
6.0	SUMMARY AND CONCLUSIONS .....	113
7.0	BIBLIOGRAPHY.....	117

## LIST OF TABLES

<b>Table 1.</b> Breast Cancer Subtypes – Hormone receptor and growth factor status.....	5
<b>Table 2:</b> Mutually, Differentially Abundant Proteins Observed Across Breast Cancer Cells.....	46
<b>Table 3:</b> Summary of proteomic analyses and miR-145 specific effects.....	70
<b>Table 4:</b> Mutually-observed miR-145 target candidates.....	71
<b>Table 5:</b> Vital statistics of 25 patients presenting with discrete stages of disease.....	85
<b>Table 6:</b> Significant proteins differentiating ER+/LN+ and ER+/LN- breast cancer tissues.....	111

## LIST OF FIGURES

<b>Figure 1.</b> Anatomy of the Breast.....	2
<b>Figure 2:</b> Neoplastic Progression of Breast Cancer.....	3
<b>Figure 3:</b> Liquid Chromatography-Mass Spectrometry (LC-MS), Shotgun Proteomics Workflow.....	9
<b>Figure 4:</b> Dynamic range of mRNA and protein abundance in cDNA versus LC-MS proteomic analyses.....	13
<b>Figure 5:</b> Analytical workflow utilized for differential and comparative proteomic analysis four breast cell lines analyzed.....	37
<b>Figure 6:</b> Protein identification characteristics across four breast cell lines analyzed.....	38
<b>Figure 7:</b> Identification characteristics for significantly, differentially abundant proteins in breast cancer cell lines.....	40
<b>Figure 8:</b> Identification of functional characteristics across proteins significantly, differentially abundant in breast cancer cell lines.....	42
<b>Figure 9A:</b> Protein identification characteristics for proteins that were mutually, differentially abundant across breast cancer cell lines.....	48
<b>Figure 9B:</b> Protein identification characteristics for proteins that were mutually, differentially abundant across breast cancer cell lines.....	50

<b>Figure 10:</b> Network analysis of mutually, differentially abundant proteins observed across breast cancer cell lines.....	51
<b>Figure 11:</b> Verification of a subset of mutually, differentially abundant proteins observed across all breast cancer cell lines.....	53
<b>Figure 12:</b> microRNA Biogenesis.....	59
<b>Figure 13:</b> miRNA Target Discovery workflow.....	68
<b>Figure 14:</b> Quantitative PCR analysis of miR-145 expression in MDA-MB-231 and SK-BR-3 cells.....	69
<b>Figure 15:</b> Growth curve of MDA-145 vs. MDA-NEG cells.....	73
<b>Figure 16:</b> Cell cycle analysis of MDA-145 vs. MDA-NEG cells.....	75
<b>Figure 17:</b> Wound healing analysis of MDA-145 vs. MDA-NEG cells.....	76
<b>Figure 18:</b> Analytical workflow utilized for differential proteomic analysis of formalin-fixed paraffin embedded (FFPE) primary breast cancer tissues.....	85
<b>Figure 19:</b> Significant, differentially abundant proteins identified in comparisons of Stage 0 and Stage III breast cancer patient tissue samples.....	87
<b>Figure 20:</b> Differential abundance profile for thrombospondin-1 (TSP-1) by spectral count peptides (A) and immunohistochemical verification (B) from Stage 0 and Stage III breast cancer patient tissue samples.....	88
<b>Figure 21:</b> Representative immunohistochemical (IHC) analysis of thrombospondin-1 (TSP-1) in Stage 0 (A & B) versus Stage III (C & D) patient tissues.....	90
<b>Figure 22:</b> Significant, differentially abundant proteins identified between Stage 0 and Stage III breast cancer patient tissue samples.....	92

**Figure 23:** Significant, differentially abundant proteins and subsequent functional characteristics identified between Stage II NR and Stage II R breast cancer patient tissue samples.....95

**Figure 24:** Differential abundance profile for protein DJ-1 (PARK7) by spectral count peptides (A) and immunohistochemical verification (B) from Stage II NR and Stage II R breast cancer patient tissue samples.....96

**Figure 25:** Representative immunohistochemical (IHC) analysis of protein DJ-1 (PARK7) in stage II NR (A & B) versus stage II R (C & D) patient tissues.....97

**Figure 26:** Significant, differentially abundant proteins and subsequent functional characteristics identified between Stage II NR and Stage II R breast cancer patient tissue samples.....100

**Figure 27:** Analytical workflow utilized for differential proteomic analysis of formalin-fixed paraffin embedded (FFPE) primary breast cancer tissues.....107

**Figure 28:** Metric of LC-MS performance during analysis of 48 patient samples.....108

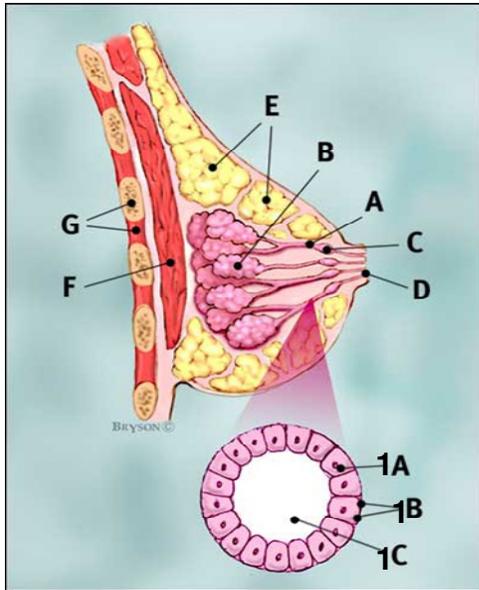
**Figure 29:** Hierarchical supervised cluster analysis of proteomic analysis derived from 35 ER+ patient tissue samples ± LN metastasis.....109

## **1.0 INTRODUCTION**

### **1.1: Breast Cancer - Background**

Breast cancer is the second leading cause of cancer related death in North American women, with greater than 200,000 new cases diagnosed and ~40,000 deaths estimated to occur in the US in 2010.<sup>1</sup> The onset of breast cancer can occur as a product of genetic pre-disposition or sporadically and is associated with a variety of risk factors that include smoking, obesity, post-menopausal hormone replacement therapy as well as evidence of increased breast density upon mammographic examination.<sup>2, 3</sup> Inherited forms of breast cancer are commonly associated with germline mutations in key tumor suppressor genes, such as the canonical tumor suppressors tumor protein 53 (p53) and the phosphatase and tensin homolog (PTEN) as well as in the DNA damage sensors breast cancer, early onset 1 or 2 (BRCA1 or BRCA2), with mutations in this latter subset of genes increasing breast cancer risk by as much as 10-20%.<sup>3, 4</sup> Sporadic breast cancers can similarly occur due to tumor suppressor gene inactivation as well as due to overexpression of various oncogenes, such as the cell cycle regulatory factor cyclin D1, the proto-oncogene c-myc or the epidermal growth factor receptors ERBB1 (EGFR) or ERBB2 (HER2/*neu*).<sup>3, 5</sup> These genetic alterations underlie the basis of mammary epithelial cell transformation, with the acquisition of further genetic instabilities as nascent tumor cell populations expand, resulting in the production of tumor cells which display more aggressive

characteristics, such as those capable of localized invasion into adjacent breast tissues and metastasis to distal sites (Figure 2).<sup>5-7</sup>

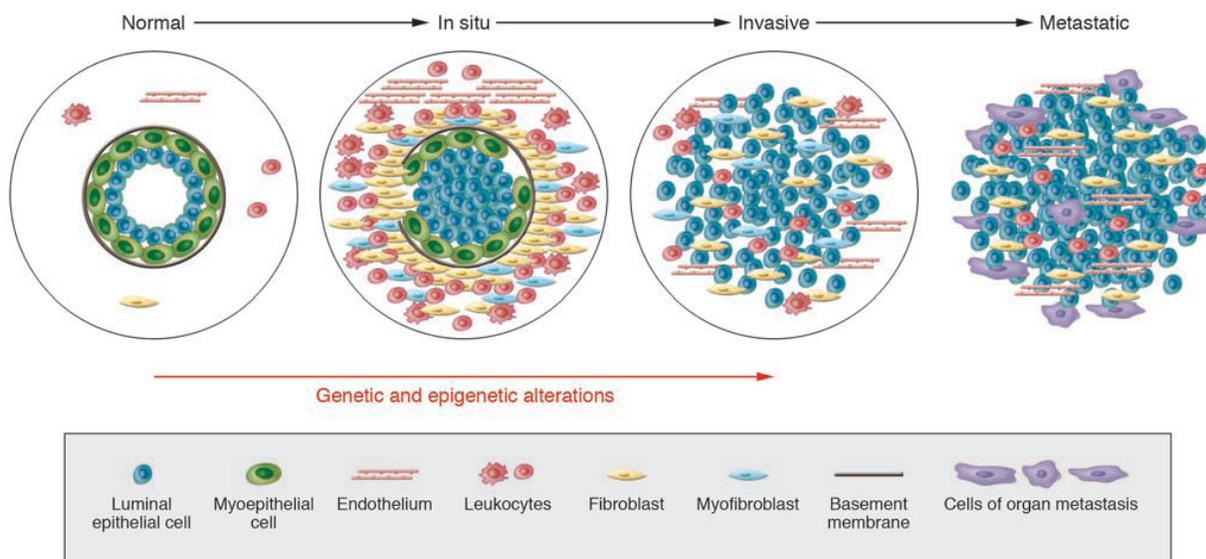


**Figure 1. Anatomy of the Breast.** (A) – Ducts, (B) – Lobules, (C) – Milk engorged duct, (D) – Nipple, (E) – Fat, (F) – Pectoralis muscle, (G) – Chest wall, (A1) – **Ductal epithelium.** (B1) – **Basement membrane.** (C1) – **Ductal Lumen.** (Adapted from <http://www.breastcancer.org>)

The adult human mammary gland is organized into systems of ductal structures which bridge lobular units, the sites of milk production, and the nipple, functioning as conduits for the transport of milk and nutrients (Figure 1).<sup>8</sup> Ductal structures are lined by epithelial cells that face the luminal space of the duct, termed luminal epithelium, and are further enveloped by myoepithelial cells that interface with an adjoining basement membrane (BM), a protein-rich structure separating the mammary epithelium from adjacent connective, or stromal tissues (Figures 1 & 2).<sup>5, 8</sup> Mammary gland tissue maturation and functional activities are predominantly regulated by the cyclical modulation of two steroid hormones, estrogen and

progesterone, that drive the growth and differentiation of these tissues.<sup>8</sup> These steroid hormones bind to discrete nuclear hormone receptors, estrogen (ER) and progesterone (PR) receptors, resulting in nuclear translocation and binding of hormone receptor complexes to ER/PR promoter response elements in target genes resulting in their transcriptional activation.<sup>7</sup> Breast cancer arises as a product of cellular transformation which is characterized by a series of discrete phenotypic changes that occur in mammary cells and include a failure to respond to growth inhibitory cues resulting in increased proliferative capacity and cell survival.<sup>9, 10</sup> Transformed

mammary cells commonly originate from luminal epithelial cell types, less so from myoepithelial cells and further, recent evidence has revealed these populations may also arise from stem-cell like mammary progenitor cells. (Figure 2, normal duct panel).<sup>5, 11-14</sup> The majority of breast cancers begin as premalignant lesions comprised of populations of hyperplastic epithelial cells that are confined to the intraductal space, termed atypical ductal hyperplasia (ADH). These lesions then progress to more aggressive localized disease, termed ductal carcinoma in situ (DCIS), the most common subtype of *in situ* disease, typified by increases in



**Figure 2. Neoplastic Progression of Breast Cancer:** Figure details a cross-section of the breast ductal epithelium and the progression from nascent through metastatic disease. (Adapted from Polyak, K. 2007)

cell abundance and the acquisition of more aberrant cell morphologies being displayed by hyperplastic populations.<sup>6, 15</sup> Disease progresses by invasion of tumor cells into adjacent tissues through degradation of the BM, a subsequent phase in progression of DCIS termed invasive, or infiltrating, ductal carcinoma (IDC) (Figure 2, *in situ* and invasive panels).<sup>5, 6, 11, 12, 16</sup> The BM is predominantly comprised of large macromolecular complexes that include extracellular matrix

(ECM) proteins, such as collagens, laminins and fibronectin which are actively secreted by both epithelial as well as stromal fibroblast cells.<sup>6, 11, 12, 16</sup> Normal mammary epithelial cell survival is dependent upon interactions with the BM as these contacts provide anti-apoptotic and survival cues and is an event termed anchorage-dependence.<sup>11, 16, 17</sup> Conversely, mammary tumor cells are capable of surviving in the absence of these interactions, so-called anchorage-independence, which is a phenotypic hallmark of cellular transformation.<sup>11, 16, 17</sup> Degradation of the BM by tumor cells is achieved through modulation of ECM protein expression and interactions and as well as due to the aberrant expression and activity of various cellular proteases, such as the matrix metalloproteinase (MMP) family of endopeptidases.<sup>11, 16, 18</sup> Invasion is further accompanied by disorganization and decreases in myoepithelial cell populations and increased numbers and activation of both stromal and immune cell types, precipitating an increase in growth factor and cytokine secretion, including the growth regulatory protein tumor growth factor beta (TGF- $\beta$ ) and the pro-angiogenic protein platelet-derived growth factor (PDGF).<sup>5, 16, 18</sup> This cellular context produces a microenvironment that further drives tumor progression through enhancement of cellular signaling which supports pro-growth and invasive cellular behaviors, such as the conversion of epithelial tumor cells to mesenchymal-like cell types (EMT)<sup>19</sup>, which embodies tumor cells that exhibit more aggressive invasive potential.<sup>5, 16, 18</sup> The evolution of tumor cells within this microenvironment results in the formation of hypoxic and acidic conditions due to restricted access to vasculature and a greater dependence of tumor cells on glycolysis versus oxidative metabolism for ATP production, a metabolic pathway which results in the conversion of pyruvate to lactate and lactic acid.<sup>20-22</sup> These selective pressures promote activation of cell survival and angiogenic signaling resulting in locoregional invasion of tumor cells into adjacent lymphovascular tissues, such as to axillary or sentinel lymph nodes (SLN)

surrounding the breast, ultimately leading to the formation of metastasis which are distal from the primary tumor site, commonly occurring in the bones, lungs and livers of breast cancer patients.<sup>11, 23</sup>

Breast Cancer Subtype	Receptor Status (Characteristics)
Luminal A	ER+/PR+/HER2- (High ER levels)
Luminal B	ER+/PR+/HER2 (+/-) (Lower ER, Higher PR) (Luminal C? – ER+/PR+/HER2+)
HER2+	ER-/PR-/HER2+ (High HER2 levels)
Basal-like (Triple-negative)	ER-/PR-/HER2-
Normal-Like	Low expression of luminal-type genes, high expression of basal-epithelial genes.

**Table 1. Molecular Subtypes of Breast Cancer – Hormone receptor and growth factor status.** ER – Estrogen Receptor, PR – Progesterone Receptor, ERBB2/HER2 (Human Epidermal growth factor Receptor 2)  
23, 26, 28-31

The inherent genetic instabilities of mammary tumor cell populations and the dynamic microenvironments and selective pressures encountered by these cells *in situ* during neoplastic progression produces disease which exhibits high inter- and intra-patient molecular heterogeneity. These characteristics thus confound disease classification, prediction of disease prognosis and definition of optimal patient treatment options.<sup>24-28</sup> The variety of disease etiologies and variability in therapeutic responses encountered in the treatment of breast cancer during the pre-genomic era underscored the basis for this disease to be one of the first solid

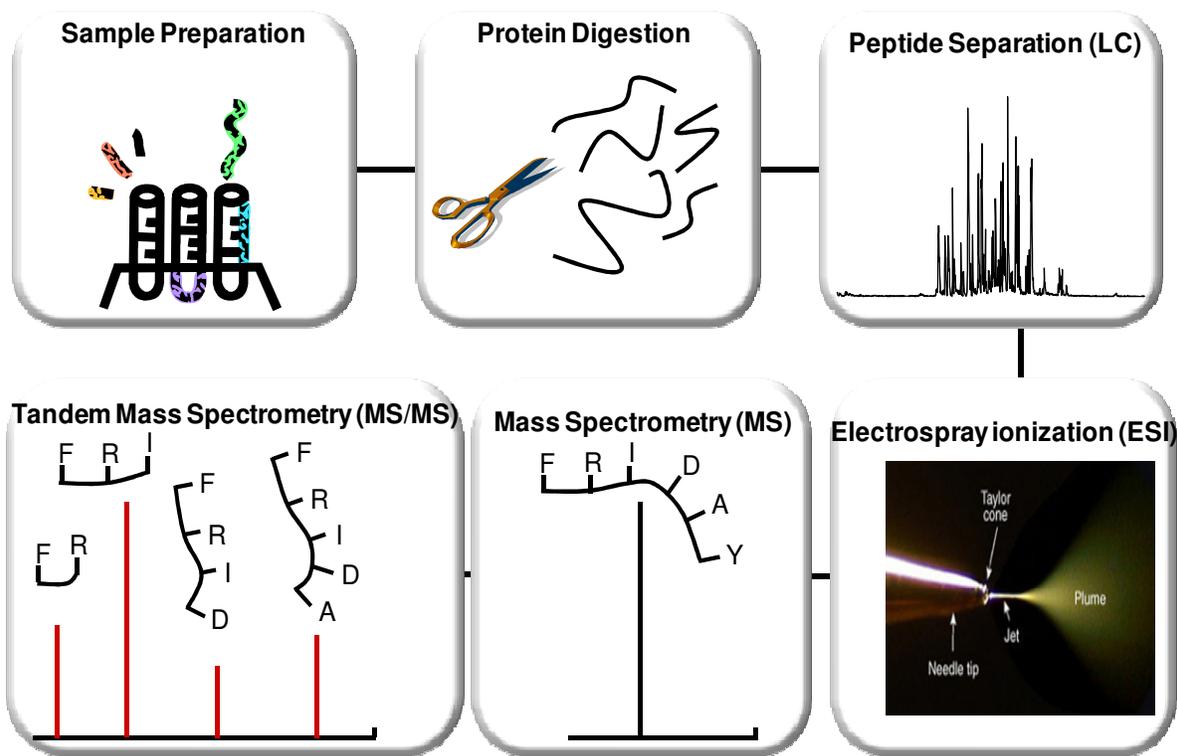
tumor cancers to be analyzed by global gene expression profiling utilizing complimentary DNA (cDNA) microarray analyses.<sup>25, 26, 28, 29</sup> Seminal studies published by Perou *et al.*<sup>25</sup> encompassing microarray analyses of 85 tissue samples from 42 breast cancer patients as well as cell line models of normal and cancerous breast cells resulted in the identification of 5 predominant molecular subtypes of breast cancer (Table 1).<sup>25, 28, 30-33</sup> These subtypes indicated that key molecular signatures are conserved amongst breast cancers which bear specific hormone and growth factor receptor expression profiles, i.e. expression of ER and PR and the human epidermal growth factor 2 receptor (HER2).<sup>25, 28, 30-33</sup> Assessment of hormone receptor and HER2 expression status has been a facet of pathological analyses of breast cancer for over 20 years and is typically achieved by immunohistochemical (IHC) analysis of biopsied patient tissue sections.<sup>27, 34-36</sup> Of the five subtypes detailed in Table 1, three predominant subtypes are encountered clinically 1) the luminal subtype which is based on ER positivity, with the further sub-classification of luminal subtypes which are also HER2+ as being either luminal-b or luminal-c as currently being in debate by the community<sup>33, 37</sup>, 2) HER2+ or 3) basal-like disease (Table 1).<sup>31, 32, 38</sup> The majority of breast cancers diagnosed are of the luminal subtype, ~70%, and are associated with significantly better 5-year survival rates than the less commonly diagnosed HER2+ or basal-like subtypes.<sup>26, 39</sup> In addition to insights into disease prognosis, assessment of hormone and growth factor receptor expression can further guide the selection of patient treatment options.<sup>31</sup> Examples include the treatment of luminal breast cancers with estrogen receptor-specific antagonists, or selective estrogen receptor modulators (SERMS), such as Tamoxifen, as well as usage of a monoclonal antibody-based therapeutic, Trastuzumab/Herceptin, which is a HER2 growth factor receptor antagonist used in the treatment of HER2+ disease.<sup>31</sup>

In conjunction with assessment of hormone and growth factor receptor status, a variety of clinicopathological analyses are utilized for the classification and determination of disease prognosis following an initial diagnosis of breast cancer. These include focused imaging studies, such as supplemental mammography with application of Breast Imaging Reporting and Data System (BI-RADS) assessment criteria, which provides insights into the likelihood that a mammographic finding may be normal, benign or malignant, as well as pathological analysis of patient tissue biopsies obtained by fine-needle aspiration cytology, stereotactic core needle biopsy or during surgical intervention.<sup>40, 41</sup> Aside from determination of hormone receptor and HER2+ status, standard pathological analyses performed in breast cancer commonly encompass determination of disease stage as well as tumor grade.<sup>42-44</sup> The criteria for staging breast cancer is summarized by the tumor node metastasis (or TNM) classification system that encompasses staging designations ranging from evidence of a carcinoma *in situ* lacking lymph node or distal metastasis, i.e. stage 0 disease, to the presence of tumor with nodal involvement as well as evidence of distal metastasis, i.e. stage IV disease.<sup>44</sup> Determination of tumor grade (commonly the Nottingham grading system in the US), encompasses measures of mitotic rate, tubule formation, as well as the prevalence of nuclear polymorphisms in tumor cells.<sup>42, 43</sup> Less common pathological analyses utilized for disease prognostication include assessment of the tumor proliferation index, which is commonly achieved via Ki-67 expression profiling, as well as by profiling for specific cytokeratins, such as CK5, CK14 and CK17 and for other growth factor receptors, such as human epidermal growth factor 1, or EGFR/ERBB1, which is often performed in basal-like breast cancer subtypes.<sup>45</sup> These clinicopathological measures, with particular emphasis on the expression status of key hormone and growth factor receptors, comprise the

standard of care for breast cancer diagnosis and determination of disease prognosis.<sup>35, 46</sup> Thus, as measures of hormone and growth factor receptor panels are core techniques utilized in the clinicopathological diagnosis of breast cancer, “proteomic” analyses have thus been a long-standing facet of this field.

## **1.2: Proteomics - Background**

Proteomics is defined as the study of global protein expression produced by a genome, cell, tissue or organism.<sup>47</sup> This strategy is directed towards the high-throughput characterization of the full complement of protein species present within a complex protein mixture, enabling protein identification, the quantitation of protein abundance as well as characterization of post-translational modifications (PTM) of proteins, such as phosphorylation and ubiquitination status.<sup>48-51</sup> Early proteomic methods included two-dimensional gel electrophoresis (2-DE) which commonly entail separating a complex protein mixture by 1-D sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) followed by further separation by isoelectric focusing (IEF), resulting in dissemination of heterogeneous protein populations occupying similar molecular weights into discrete protein groups.<sup>52</sup> When comparing two conditions after staining, differential spot intensities provided relative quantitative information about protein abundance, and further mobility shifts in protein spots indicated the possibility of differential protein modifications between conditions.<sup>52</sup> However, protein spot identification was typically achieved by direct sequencing of proteins, such as via N-terminal Edman degradation, a method which has since been largely displaced by mass spectrometry-based strategies.<sup>52, 53</sup>



**Figure 3. Liquid Chromatography-Mass Spectrometry (LC-MS), Shotgun Proteomics Workflow** – Sample preparation entails cell or tissue lysis followed by protein digestion, such as with the proteolytic enzyme trypsin. Resulting peptides are then separated into a less complex mixture by liquid chromatography and are ionized via electrospray ionization, resulting in evaporation of mobile phase and liberation of charged peptide ions. Peptide ions are then introduced into a mass spectrometer resulting in initial measurement of their intact mass to charge ratio ( $m/z$ ), hypothetical peptide F-R-I-D-A-Y displayed in MS event, followed by a second, tandem MS event (MS/MS) in which an isolated peptide is fragmented by collision with Nobel gas molecules (termed collision induced dissociation (CID) followed by measurement of subsequent fragment ion  $m/z$  ratios.

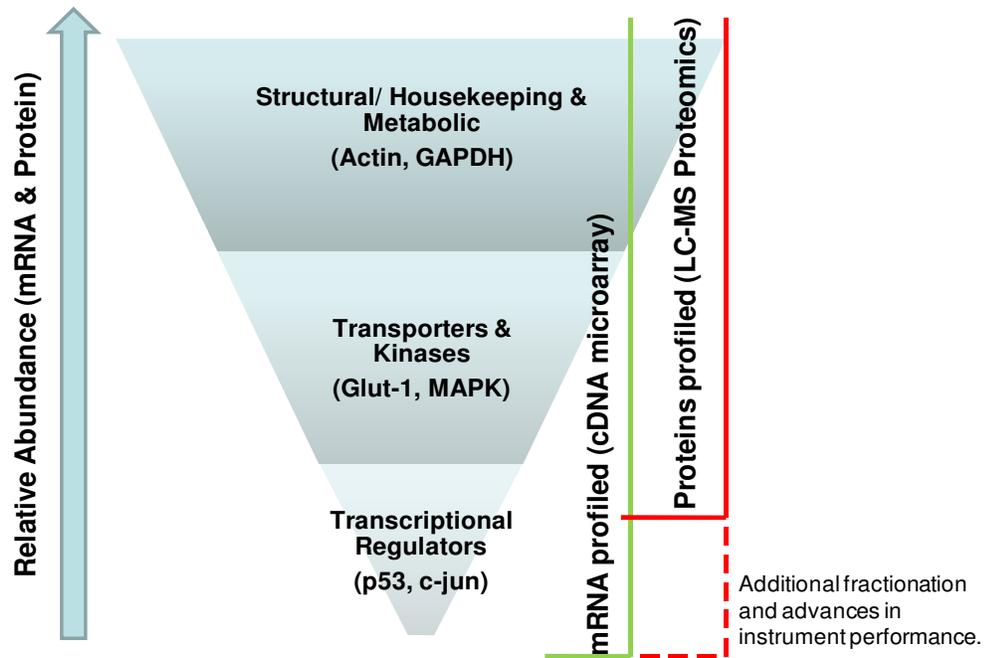
Proteomic strategies which utilize mass spectrometry and the application of so-called “shotgun” strategies directed towards the characterization of many proteins in a complex mixture, in

conjunction with various upstream fractionation methodologies, such as liquid-chromatography, i.e. LC-MS, enable more comprehensive characterization of proteins samples.<sup>49, 54</sup> LC-MS-based proteomics (Figure 3 – ESI LC-MS/MS workflow) describes a strategy in which a complex protein mixture is enzymatically digested into constituent peptides which are then resolved utilizing liquid chromatography. This is achieved by binding of peptides onto an analytical LC column packed with a stationary phase in which peptides directly bind, such as reversed- phase, e.g. C-18 bonded silica beads, or ion exchange resins, such as strong-cation exchange.<sup>54</sup> Peptides are progressively eluted via introduction of an increasing gradient of non-polar mobile phase, such as in the case of reverse-phase resins, over a defined period of time and eluted peptides are then introduced into a mass spectrometer via electrospray ionization (ESI) (Figure 3).<sup>51, 55,56</sup> ESI entails the application of an ambient voltage to an analytical LC column resulting in the vaporization of eluting peptide droplets followed by transfer of subsequent gas-phase peptide ions into an MS source which is further under vacuum.<sup>51</sup> Intact peptide ion precursor  $m/z$  and CID spectra are then compared computationally to an *in silico* digested peptide library constructed from a species-specific protein database where CID spectra are scored based on their degree of concordance with predicted peptide masses and associated theoretical fragment ions masses. The highest scoring comparisons result in the identification of a peptide sequence followed by assignment of that peptide to a corresponding parent protein.<sup>57-59</sup> Peptide identification accuracy is dependent upon the ability of a mass spectrometer to resolve unique peptide ion spectra, which aids in the determination of peptide charge state, and to perform highly accurate mass measurements of peptide precursors and fragment ions, which affords the application of more stringent criteria in peptide searches, thus, assuring more confident peptide identifications are obtained.<sup>51</sup> Examples of high resolution, high mass accuracy mass

spectrometers include the Fourier transform-mass spectrometer (FT-MS) as well the Orbitrap MS which are capable of achieving resolutions in excess of 100,000 and mass accuracies of < 2 parts per million (ppm). Lower resolution instruments include ion trap or TOF instruments which achieve average resolutions of  $4.0\text{-}8.0 \times 10^3$  and mass accuracies >100ppm.<sup>51, 60</sup> The application of further search criteria specifying peptide characteristics that are imposed by experimental design, such as the expectation that peptides generated by digestion of proteins with trypsin, a proteolytic enzyme which cleaves peptides bonds after lysine or arginine residues, will terminate in K or R, further aids to strengthen the validity of peptide identifications observed.<sup>61, 62</sup> Assembly of peptide identifications indicates the degree of protein coverage observed, with the general consensus in the proteomics community of a protein being identified by at least two peptides as being considered a confident identification.<sup>63</sup>

Bottom-up, shotgun proteomic strategies, in which peptides produced by protein digestion are analyzed by LC-MS and then compiled to identify corresponding parent proteins, facilitates the identification and quantitation of many peptides/ protein species within a given sample.<sup>48, 54, 64</sup> Comparisons of peptide identification rates (or spectral counts), peptide ion peak areas and intensities (such as in SILAC-analysis) or the peak intensities of fragmentation spectra derived from chemical tags covalently added to peptide species prior to MS analysis (such as in iTRAQ-technologies), facilitates the quantitative comparison of global peptide species and thus protein abundance between multiple experimental conditions.<sup>65-67</sup> Like other non-MS based protein abundance quantitation strategies, such as western and enzyme-linked immune absorbent assay (ELISA) analyses, comparative, quantitative analyses utilizing LC-MS-based proteomics begins with equivalent amounts of protein/ peptide concentrations. Spectral count-based quantitative

proteome analyses is achieved by comparison of equivalent peptide populations from multiple conditions where total numbers of peptides identified for a given protein are quantitatively compared across analyses to achieve a measure of differential protein abundance, where a greater number of peptides identified for a protein in sample 1 versus 2 would indicate relatively greater abundance in sample 1.<sup>65</sup> To minimize the contribution of analytical variability imposed by differential sample processing or instrument performance, LC-MS analysis of isotopically-labeled (SILAC) or chemically-tagged peptides (iTRAQ) affords the analysis of multiple experimental conditions simultaneously.<sup>66 67</sup> For example, the SILAC MS approach affords concurrent analysis of up to three different complex protein mixtures as a product of metabolic labeling of proteins with either single or multiple amino acids that are labeled with stable isotope atoms (<sup>13</sup>C, <sup>2</sup>H, <sup>15</sup>N, or <sup>18</sup>O) or 'light' (<sup>12</sup>C, etc.), being achieved via supplementation in tissue culture media *in vitro* or by dietary supplementation in *in vivo* animal studies.<sup>66</sup> The result of metabolic labeling utilizing 'heavy' isotope amino acids is their incorporation into proteins that are identical to those from native conditions in all respects with the exception that they possess a greater mass according to the number of 'heavy' labeled amino acids present in the sequence.<sup>66</sup> Differentially labeled peptide populations are mixed in equivalent ratios and subjected to LC-MS/MS in which peak areas for identical peptide isotopomers, which exhibit identical LC retention times differing only by the mass deflections contributed by the stable isotope-labeled amino acids, are thus compared providing quantitative information about the abundances of a given peptide/ protein across multiple conditions in a single LC-MS analysis.<sup>66, 68</sup>



**Figure 4: Dynamic range of mRNA and protein abundance in cDNA versus LC-MS proteomic analyses:** The dynamic range of expressed cellular genes and proteins vary from those ubiquitously expressed (such as structural/housekeeping factors) to factors present at lower abundances (such as kinases, transcription factors). As LC-MS, global proteomic analyses depends on *de novo* peptide/protein identification versus more targeted approaches utilized in gene expression / cDNA microarray analyses, the dynamic range of proteins poses a challenge to the comprehensive analysis of the full complement of proteins present in a cell lysate. The integration of further peptide/protein fractionation strategies paired with advances in MS instrument performance have benefited this analytical hurdle. GAPDH: Glyceraldehyde 3-phosphate dehydrogenase, Glut-1: glucose transporter 1, MAPK: mitogen-activated protein kinase, p53: tumor protein 53, c-jun: AP-1 early response transcription factor subunit.

The application of proteomic analyses to assess global protein abundance characteristics provides insights into the biology at play in a given tissue sample or cell population as it is a measure of the end-product of epigenetic, post-transcriptional and post-translational regulatory events.<sup>69, 70</sup> However, the complexity and expansive dynamic range of the cellular proteome poses a challenge to the characterization of the full complement of proteins in a complex sample by proteomic analysis as this technique depends on the *de novo* characterization of proteins and peptides. This is in contrast to global gene expression profiling, such as cDNA microarray analyses, which affords greater specificity comparatively as observations are derived from the interaction of cDNA with pre-determined targets (Figure 4).<sup>71</sup> Specifically, peptides corresponding to cellular proteins which are in high abundance, such as cytoskeletal or so-called “housekeeping” proteins, are over-represented in population of peptides derived from a total cellular proteome digest relative to peptides derived from more low abundant proteins, such as those which are often of interest in disease-related biological processes, e.g. protein kinases or transcription factors. Strategies to decrease the complexity of the cellular proteome prior to MS analysis have proven useful in achieving greater depths of proteome coverage, such as the application of in-gel digestion, involving separation of a complex protein sample by 1D SDS-PAGE followed by dissection of gel spots containing protein populations at differing molecular weights and subsequent digestion and extraction of peptides. Further strategies which combine multiple orthogonal fractionation techniques, such as the multidimensional protein identification technology (or MudPIT) strategy, comprising an in-line LC technique which separates peptides firstly by strong cation exchange and then by reverse-phase chromatography have also increased proteome coverage.<sup>72 53</sup> Further, advances in MS instrumentation, such as the recent release of the Velos series of mass spectrometers by Thermo Scientific, have resulted in technological

capabilities that have increased peptide identification rates by as much as 128% over previous technologies.<sup>73</sup> With these strategies in place, protein identification capabilities and the high throughput nature of modern proteomic workflows have proven fruitful for the purposes of discovering biomarkers and elucidating functional biological roles of protein groups in a variety of disease-related pathologies including breast cancer.

### **1.3: Proteomics of Breast Cancer – Biomarker Discovery**

The molecular heterogeneity of breast cancer underscores the benefits to be gained from analysis of this disease utilizing high-throughput gene and protein expression profiling. The main focus of these studies is to discover gene and protein expression characteristics that are conserved across breast cancer with the goal of identifying biomarkers which will aid in disease diagnosis by better defining disease subtypes and responsiveness to therapeutics as well as towards identifying systems of genes and proteins that underlie and drive breast cancer disease biology.<sup>74-76</sup> The advantages afforded by proteomic analyses are that the targets of these studies are the predominant functional mediators of biology, which readily facilitates translation of identified proteins of interest to subsequent investigations towards discerning the utility of these candidates as therapeutic targets or as factors central to molecular disease mechanisms.<sup>74</sup> Protein identification is the end product of post-transcriptional regulatory events which are mechanisms that can impact the translation of a gene target of interest; a lesson recently underscored by evidence which has emerged from the burgeoning field of microRNA-mediated regulation of gene translation.<sup>77-79</sup> This point further underscores the value of performing multi-platform -omics analyses of breast cancer in which genomic, differential mRNA, microRNA as well as

protein abundance characteristics are integrated at a systems level to provide insights into the regulatory dynamics underlying the gain or loss of factors associated with disease biology.<sup>80, 81</sup>

As a product of the variety of sample types suited to proteomic analyses, such as the ability to analyze clinical tissue samples and biofluids as well as to perform high-throughput analyses of large sample sets, proteomic research in breast cancer has been largely directed towards the discovery of disease-specific protein characteristics that can be directly translated to clinical and therapeutic applications.<sup>74</sup> The identification of biomarkers in breast cancer via proteomics has been approached utilizing models systems ranging from cell lines<sup>69, 82-91</sup>, animal models of breast cancer, including transgenic murine models bearing HER2 gain-of-function characteristics<sup>92</sup> and xenograft models of human breast cancer<sup>93-96</sup>, as well as in clinical samples ranging from blood serum<sup>97-102</sup>, cerebrospinal fluid (CSF)<sup>103</sup> and nipple aspirate fluid (NAF)<sup>104-114</sup> to clinical breast cancer tissues<sup>115-118</sup> and breast cancer tissue effluent, or breast tissue interstitial fluid (TIF)<sup>119-121</sup>.

Investigations of breast cancer-specific protein abundance characteristics derived from studies of cell line models of breast cancer, *in vivo* breast cancer tissue models and clinical tissue samples have been driven by the logic that these sample sources will possess the most robust molecular characteristics representative of breast cancer disease biology. For example, a global, proteomic analysis utilizing a MALDI-TOF MS/MS approach of pooled membrane proteins derived from several human cell line models of breast cancer, namely the ER+ cell lines, MCF7 and T-47D, as well as the EGFR-positive cell lines MDA-MB-468 and BT-474, resulted in the identification of three novel proteins, i.e. BCMP11, BCMP84 and BCMP101. These proteins were found to be specific for cancerous versus normal breast epithelium in subsequent large cohort IHC analyses;

with BCMP11 in particular as being observed at high abundances in 43 versus 58 breast cancer patients analyzed.<sup>82</sup> Further, recent analysis of an isogenic model of breast cancer disease progression, termed MCF10AT, which comprises four cell lines representative of discrete stages of disease progression, i.e. normal-like breast cells, pre-neoplastic cells, and low and high-grade breast cancer cells, initially established from clonal isolates originating from a squamous carcinoma derived from a xenograft of a RAS-transformed variant of a cell line model of normal human mammary epithelium, MCF10A.<sup>93, 95</sup> Global proteomic analysis of this system was achieved via a quantitative ESI LC-MS/MS approach in which total peptides derived from each cell line model were labeled with iTRAQ-reporter tags, thus facilitating the coordinate determination of abundance for a given peptide identification across all conditions.<sup>93</sup> Results yielded a total of 1200 protein identifications and revealed protein abundance characteristics indicative of disease progression in breast cancer, such as the observed progressive increase in vimentin abundance with more aggressive cell types, a marker of mesenchymal cells and indicative of breast cells which have undergone EMT.<sup>93</sup> Further, these studies resulted in the identification of differentially abundant protein profiles which had not been previously implicated in breast cancer, namely adenylate kinase-1 (AK1), copper transport protein (ATOX1) and histone H2B type 1-M (HIST1H2BM); which were validated in a larger cohort of 26 matched normal and breast cancer tissue samples by IHC analysis.<sup>93</sup>

Proteomic analysis of clinical breast cancer tissues have revealed that discrete tissue compartments, i.e. tumor epithelium and adjacent stromal tissues, bear unique proteomic signatures<sup>122</sup>, a finding which underscores the value of obtaining homogenous cellular regions utilizing laser microdissection (LM) prior to performing proteomic analysis to ensure cellular

populations of interest are enriched for.<sup>117, 122</sup> A recent global, proteomic analysis utilizing a label-free (spectral count) LC-MS method of LM-captured tumor regions derived from fresh, frozen breast tissue specimens obtained from one normal patient and three with IDC resulted in the identification of a 60 member protein panel specific for IDC tissues.<sup>115</sup> The authors further validated the differential abundance observed for one factor, nucleoside diphosphate kinase A (NDKA), in a subset of patient tissue samples by western and IHC analysis.<sup>115</sup> Clinical samples analyzed by -omics technologies are often obtained by biopsy or surgical intervention and as these samples are often too precious or exhausted through subsequent clinicopathological analyses, it is often difficult to amass a large number of fresh clinical tissue specimens to perform analyses with sufficient power. One resource of clinical breast cancer specimens which addresses this issue are formalin-fixed, paraffin embedded (FFPE) tissues, a common tissue preservation technique utilized for the preparation of clinical samples for pathological analysis.<sup>80, 123</sup> Wide-scale usage of FFPE tissue preservation has resulted in the assembly of vast archives of clinical breast cancer tissue that represent excellent resources for conducting retrospective proteomic analyses.<sup>80</sup> Proteomic versus genomic analyses of FFPE tissues have further revealed that the impact of formalin fixation on tissues, which produces protein-protein (lysines, backbone amides), protein-DNA and DNA-DNA cross-links, more greatly influences the quality of RNA that can be obtained from these tissues than protein.<sup>124 123, 125-127</sup> Global protein profiling of FFPE tissue samples utilizing various proteomic strategies, such as reverse-phase protein arrays and LC-MS methodologies, further when paired with LM, have proven robust.<sup>123, 126-131</sup> One proteomic analysis of FFPE breast cancer tissues has been described<sup>132</sup> and encompassed a proof of concept analysis towards optimizing protein/peptide recovery from FFPE tissues for utility in a MALDI-TOF Imaging MS strategy. Results revealed peptide spectra were optimally obtained

after subjecting samples to a heat-induced antigen retrieval and subsequent trypsin digestion regimen.<sup>132</sup> Further, data presented in chapters 3 and 4 of this thesis describe two LC-MS-based, global proteomic analyses utilizing FFPE clinical breast cancer tissues towards the characterization of disease specific protein characteristics.

The primary focus of proteomic research in breast cancer has been directed towards the discovery of single or multiple biomarker panels which can accurately diagnose disease and which can be profiled from sample sources obtained by minimally invasive means, such as via a blood sampling, utilizing commonly employed clinical techniques, such as immunohistochemical-based analyses.<sup>74, 133, 134</sup> Therefore, many discovery efforts have focused on proteins which are actively secreted by tumor cells as protein candidates which may enter into lymphovascular circulation, such as “secretome” analyses of spent tissue culture media containing proteins secreted from cell line models of human breast cancer cells, as well as studies of clinical biofluid samples, such as serum, CSF, and fluids proximal to primary tumor sites, such as NAF and breast TIF.<sup>84, 91, 96-121</sup> Biomarker discovery efforts which have investigated serum derived from breast cancer patients for disease specific protein patterns have been confounded by the dynamic range bias imposed by abundant proteins present in serum and thus often employ a depletion strategy to remove these high abundant species prior to LC-MS analysis, resulting in enrichment of lower abundant proteins which may be of interest to disease biology.<sup>97, 99, 135, 136</sup> In the case of a recent global proteomic analysis utilizing a MudPIT LC-MS/MS strategy of serum samples derived from a cohort of patients with benign breast disease or presenting with IDC with or without lymph-node metastasis, serum samples were depleted of albumin utilizing an EtOH precipitation method.<sup>97</sup> Results yielded a total of 2,078 proteins being

identified by at least two peptides across all samples analyzed and differentially abundant protein populations by hierarchical cluster analysis of spectral count peptides revealed 41 proteins ( $\pm$  ~16) which significantly differentiated these various patient groups.<sup>97</sup> Further analyses of one protein, tenascin-XB (TNXB), which was decreased in IDC patient serum with lymph node involvement, in a large patient cohort of 131 serum samples revealed this expression profile was specific for this subgroup of patients.<sup>97</sup> As dynamic range biases imposed by high abundant proteins are factors to consider prior to proteomic analysis of both serum and CSF samples, alternative strategies towards characterizing protein populations which are more enriched and specific for breast tumor cells have been contrived.<sup>84, 91, 104-114, 119-121</sup> *In vitro* global, proteomic analyses of conditioned media derived from cultured breast cancer cell line models has proven a robust strategy for the characterization of proteins that are actively shed from breast cancer cells.<sup>84, 85, 134, 137, 138</sup> One such study utilized an LC-MS/MS approach to characterize conditioned media derived from 3 human breast cell lines, i.e. MCF10A and two cancerous cell lines, BT474 and MDA-MB-468 cells, which were grown to confluency and cultured in serum-free media, resulted in the identification of ~600 proteins across these media conditions.<sup>84</sup> One issue with this strategy is the concern that proteins present in conditioned media may contain carryover serum proteins, as cells are typically established in serum containing media prior to being transitioned to serum-free conditions, as well as intracellular proteins released by cells which have undergone lysis.<sup>84, 134</sup> Characterization of the known cellular localization status via bioinformatic analyses for these protein identifications did not indicate that specific enrichment of known secreted or ECM-localized proteins was achieved and rather showed that proteins were identified from various cellular compartments.<sup>84</sup> To increase the confidence that those proteins identified from conditioned media were actually the product of shedding by tumor cells, authors

performed parallel analyses of corresponding total cell lysates; through comparisons of proteomic data sets obtained from lysed cells and conditioned media, it was possible to identify secreted protein candidates with higher confidence.<sup>84</sup> Interestingly, comparative analysis of secreted protein candidates from these studies with previously published proteomic analyses of NAF and breast TIF derived from tumor samples (described below) revealed a high degree of concordance between conditioned media proteins and those observed in proximal fluid analyses.<sup>84</sup> Further, these authors inspected the levels of one secreted protein candidate, the secreted epithelial proteinase inhibitor elafin, by ELISA analysis of serum samples from three groups of breast cancer patients reflective of increasing levels of the cancer antigen CA 15-3 as a cohort of patients exhibiting progressively more aggressive disease characteristics.<sup>84</sup> CA 15-3 is a derivative of the cell-surface associated protein, MUC-1 which has been proposed as a serum biomarker of breast early breast cancer.<sup>76, 84</sup> Though initial observations revealed a decrease in elafin levels in conditioned media derived from tumorigenic versus normal breast cells, no significant pattern of elafin levels in serum from the patient cohort was observed.<sup>84</sup>

Proteomic analyses of clinical fluids immediately proximal to a primary breast tumor, such as NAF, have also yielded provocative breast cancer biomarker candidates.<sup>104-114, 119-121</sup> NAF, or breast ductal fluid, is produced by the alveolar-ductal system of the breast and has been an attractive resource for breast cancer biomarker discovery as it is a breast-specific proximal fluid that is attainable by non-invasive means.<sup>139-141</sup> Further, profiling of NAF for carcinoembryonic antigen (CEA) and HER2 in abnormal nipple discharge has been approved for the diagnosis of breast cancer in Japan.<sup>139, 142, 143</sup> Several proteomic analyses of NAF have been reported since 2001 which have focused on optimizing strategies for analysis of NAF,<sup>104-114</sup> characterizing

inter-patient variability of NAF proteome characteristics<sup>144</sup> and establishing proteomic profiles and identifying biomarkers in NAF which are specific for breast cancer.<sup>107-114</sup> A recent LC-MS analysis of the human NAF proteome which compared samples derived from normal and breast cancer patients by spectral count revealed this proximal fluid to contain as many as 896 proteins with an average of 37% identified as being localized to the extracellular matrix and 27% to the plasma membrane of cells with 41% further overlapping with proteins which have previously been observed in normal blood serum.<sup>108</sup> It was further shown that the NAF proteome of breast cancer patients exhibits significant heterogeneity relative to NAF samples derived from healthy individuals and reported the identification of several proteins which have been proposed as biomarkers for breast cancer, i.e. urokinase-type plasminogen activator (uPA),<sup>76, 145</sup> cathepsin-D,<sup>146</sup> cancer antigen (CA) 15.3<sup>147</sup> and tissue plasminogen activator (tPA).<sup>148</sup>

An additional proximal fluid source for analysis of secreted biomarkers from breast tumor cells has been breast TIF (tissue interstitial fluid or tumor interstitial fluid).<sup>127-129</sup> Breast TIF is obtained from fresh breast tissues acquired immediately post-surgical resection where regions of normal and tumor-containing tissues are diced and incubated in phosphate-buffered saline (PBS) at 37° C to facilitate the efflux of secreted proteins from cell populations.<sup>127-129, 149</sup> The logic underlying this strategy is that proteins effluxed from tumor tissues will contain high concentrations of secreted, tumor-specific proteins that may become diluted in serum and may thus be analyzable by more targeted assays in blood.<sup>121, 149</sup> One possible issue arising from this strategy are the potential biases imposed by incubation of live tissues in PBS due to this solution not being representative of the *in situ* tissue environment, proffering similar concerns as tissue culture-based secretome strategies utilize cells cultured in serum-free media. However, recent

assessment of various buffer conditions ranging from PBS and serum free tissue culture media (DMEM) to tissue preservation solutions utilized for organ transplantation, such as Celsior solution S and histidinetryptophan-ketoglutarate (HTK) solution, revealed no specific buffer condition significantly modulated the array of proteins identified by cellular localization in recursive proteomic analyses of TIF samples harvested from ovarian and kidney tissues incubated in these various buffers, indicating PBS as a suitable buffer to utilize for TIF acquisitions.<sup>149</sup> The seminal proteomic analysis of breast TIF utilized a 2DE, MALDI-MS strategy which analyzed TIF obtained from 16 patients with IDC resulting in the identification of 267 proteins which predominantly comprised proteins previously known to be localized to the cellular cytoplasm as well as present in serum.<sup>121</sup> Later analyses by this group using identical proteomic workflows focused on comparisons of matched normal and breast tumor tissue-derived TIF from 69 breast cancer patients.<sup>150</sup> Initial analyses were performed utilizing a matched pair of normal and tumor TIF derived from a single patient which yielded 110 proteins candidates that were robustly increased in tumor versus normal TIF. These initial 110 candidates then functioned as a training set for the assessment of 2DE gels profiles of tumor TIF acquired from the remaining 68 patients with the goal of identifying a subset of proteins exhibiting similar expression trends in at least 90% or more of the patients analyzed.<sup>150</sup> Results revealed a panel of 26 proteins which were observed at increased abundances in all 69 tumor TIF patient samples, a subset of 9 was further validated in a larger cohort of 70 malignant breast carcinomas by tissue microarray analysis.<sup>150</sup> The authors further propose future analyses which will test the feasibility of profiling expression of this subset of proteins as biomarkers of breast cancer by simple blood tests.<sup>150</sup>

Though evidence has been provided indicating that a variety of compelling breast cancer biomarker candidates have emerged from proteomic-based discovery efforts, thus far, a lack of sufficient vetting of these factors, particularly in large patient cohort studies, has limited the translation of these biomarkers to the clinic.<sup>76, 133</sup> Further, when metrics which classically define a “good” clinical biomarker, such as those detailing that a marker be highly sensitive for identifying individuals with disease and produce a low false-positive rate, i.e. be highly specific for these individuals, biomarker candidate performance often falters due to significant inter-patient variability as well as due to inconsistencies in marker profiling strategies utilized.<sup>76, 133, 134, 151</sup>

In the context of –omics technologies actively utilized as clinical breast cancer tools, gene-based biomarker discovery efforts utilizing cDNA microarray analysis have yielded two gene expression profiling strategies, the Oncotype DX and the FDA-approved Mammaprint platforms, which are now commonly utilized by the clinical community, further having expanded the repertoire of prognostic strategies available to clinicians.<sup>26, 28, 46, 76, 125, 152-154</sup> For example, the Oncotype DX assay comprises a 21-gene panel designed for assessment of disease prognosis in breast cancer patients with early-stage ER-positive (ER+) disease lacking nodal involvement (pN0) which has further been shown to accurately predict 10-year disease recurrence rates in this subgroup of patients.<sup>125, 154</sup> Further, Oncotype DX assay results are provided as a composite recurrence score that has been shown to be effective in determining higher risk patients that may benefit from extended hormonal therapies.<sup>125, 154</sup> These diagnostic platforms underscore the value of assessing disease-associated biomarker panels rather than single factors as these strategies increase the sensitivity and specificity of the diagnostic endpoint in question.<sup>155</sup>

#### **1.4: Proteomics of Breast Cancer – Functional Proteomics**

The utility of data derived from global proteomic analyses not only functions as a resource for the discovery of disease-specific protein signatures which can be directly utilized in clinical and diagnostic applications, but further, when combined with bioinformatic analyses of functional protein databases, such as gene ontology (GO) identifiers, protein-protein interaction and biological pathway networks, the potential to provide significant information regarding functional biological characteristics associated with a disease subtype or biological event becomes possible.<sup>156, 157</sup> These analyses facilitate the ability to investigate novel questions regarding protein functionality and activity in breast cancer, yielding insights which can provide the basis for further hypothesis-driven research to elucidate the role of factors of interest. For example, in the instance of a single proteomic analysis, consideration of trends in functional protein subtypes through comparative assessment with protein databases comprising known protein functions or peptide sequences with archives of conserved protein domains known to mediate specific protein functions, such as protein kinase or DNA-binding domains, can provide insights into predominant systems level biological activities.<sup>157</sup> In the context of considering protein functionality along with protein abundance, if proteins which exhibit common functional characteristics occupy similar regions within the dynamic range of the protein abundances observed, such as multiple proteins at high abundances associated with, for example, regulation of endosomal transport or with specific cellular signaling cascades, this would support the hypothesis that the known functional roles associated with these proteins may play key roles in the biological system analyzed. Further, as protein abundance does not directly correspond to

protein activity, consideration of protein groups which share similar functionalities not only by abundance characteristics, but further in relation to their post-translational modification status, such as phosphorylation-state, adds further dimensions to assessment of whether a protein group may significantly contribute to the underlying biology of a system.<sup>157</sup> In addition, these types of observations are strengthened by analysis of a large number of samples, being that consideration of protein functionality across larger sample sets enables more focused discovery of central themes conserved across similar systems, further aiding to highlight characteristics which are unique to subtypes that diverge from central system features. This course of logic is particularly powerful towards the characterization of conserved and divergent functional protein characteristics in cancer biology, particularly in the case of breast cancer, due to the high degree of molecular heterogeneity associated with these diseases.

Methods to perform these analyses are diverse and are focused on statistical measurements of confidence determining whether a given protein or protein group and corresponding functional roles are significantly overrepresented within a given data set versus the possibility that this observation is occurring due to random chance, i.e. the null hypothesis.<sup>157-161</sup> The most commonly utilized strategy to assess functional characteristics within large gene or protein data sets utilize GO terms, which are hierarchical categories of specific biological functions that correspond to three main functional sub-groups, those associated with 1) biological process 2) molecular function and 3) cellular component, and a given gene or protein is assigned to a category based on published findings and correlative functional protein domains.<sup>157, 162</sup> Various tools, such as NIH DAVID<sup>163</sup> and the CytoScape plug-in ClueGO<sup>161</sup>, facilitate calculation of enriched GO terms via referencing user-defined data sets against GO databases and provide

measures of significance as to whether a given GO category is overrepresented in a data set through occupancy of experimental candidates in that GO category.<sup>157</sup> Further, proprietary functional analysis tools, such as Ingenuity Pathway Analysis (IPA Systems, [www.ingenuity.com](http://www.ingenuity.com)), afford the comparison of multiple functional databases to a user data set to discern significantly enriched biological functions, further enabling direct assessment of the primary findings in which a given association between a gene/protein and function was initially derived, enabling users tools to determine whether an observation is made in a similar species, disease subtype or biological system.<sup>157</sup> Aside from assessment of enriched functional characteristics within proteomic data sets, several tools, including CytoScape and IPA, are also focused towards the identification of protein interaction networks derived from comparisons with established protein interaction databases compiled from large-scale yeast two-hybrid and synthetic lethal screens as well as maps of immunoprecipitated protein complexes and protein-DNA interaction networks to enable characterization of signal transduction and protein interaction networks.<sup>158, 159, 164</sup>

In the context of breast cancer, several global proteomic analyses of human breast cancer cell lines have been published which have focused on elucidation of functional protein characteristics associated with breast cancer pathogenesis.<sup>69, 89, 90</sup> One such study detailing an LC-MS based proteomic analysis of four breast cell lines, one representative of normal mammary epithelium (HMEC) and three breast tumor cells (MCF7, MDA-MB-231 and BT-474) resulted in the identification of 2235 proteins across all cell lines, 234 of which were mutually observed in all conditions.<sup>90</sup> Mutually observed proteins were then 1) compared to microarray data previously published for these cell lines, 2) analyzed for the prevalence of underlying protein interaction

networks within this group and 3) investigated further for potential PTMs.<sup>90</sup> Results revealed nine proteins which exhibited similar protein and gene expression trends across the cell line conditions which further clustered in discrete protein interaction networks via assessment of these factors against the Biomolecular Interactions Database.<sup>90</sup> These results further revealed the modulation of several integrin isoforms (proteins associated with intercellular interactions) as being decreased across tumor cell lines as well as factors which have been shown to be key mediators of tumorigenicity in breast cancer cells, such as epidermal growth factor receptor (EGFR).<sup>90</sup> Assessment of PTMs within this subset of proteins via variable modification searches considering phosphorylation, acetylation, O-GlcNAc, palmitoylation, C-mannosylation, hydroxylation, glucosylation, and S-nitrosylation PTM events revealed cancer cell line specific modifications, such as phosphorylation of alpha catenin (CTNNA1), a key member of cellular adherens junctions, as well as O-GlcNAc modification of coronin-1B, a cytoskeletal protein associated with cell motility and cytokinesis.<sup>90</sup> Additionally, an inferred protein-network analysis comparing microarray data generated from two large cohort studies of primary breast cancer patients with and without metastatic disease with previously established protein-protein interaction databases resulted in the identification of metastasis-specific protein sub-networks.<sup>156</sup> Specifically, the authors identified factors previously associated with breast cancer pathogenesis, such as p53 and HER2, as well as interacting network proteins which exhibited discrete expression trends in metastatic versus non-metastatic groups.<sup>156</sup> Data was further provided showing that metastasis-specific sub-networks identified from one microarray analyses could accurately predict individuals with metastatic disease in the corresponding data set ~50% of the time.<sup>156</sup> An additional analysis derived from microarray studies characterizing key oncogenic signaling factors in breast cancer further revealed provocative evidence that assessment of

pathway activation state can provide another dimension to standard disease subtype classification conventions that extend beyond ER, PR and HER2 profiling alone.<sup>165</sup> These findings were further proposed as a strategy to identify higher risk patients in these subgroups towards the development of more tailored therapeutic strategies.<sup>165</sup> Though the utility of functional proteomic as well as genomic data are just beginning to be realized, the correlation of a given gene or protein group factors of interest with previously published observations associating a factor with a specific functional biological role may aid in selecting targets for future investigations for their role in a disease state, further expanding the knowledge-base surrounding both the factor of interest and the associated disease process.

## **1.5: Introduction to Data Chapters**

The following thesis comprises studies focused on the application of proteomics and subsequent bioinformatics towards discerning differential protein abundance and associated functional characteristics which underlie tumorigenesis and neoplastic progression in breast cancer. The data detailed in chapters one<sup>166, 167</sup> and two are focused on *in vitro* analyses of conserved protein abundance and functional characteristics across divergent molecular subtypes of breast cancer as well as towards the identification of proteins regulated by the breast cancer associated microRNA, miR-145. Data chapters three<sup>168</sup> and four are focused on the proteomic analysis of formalin-fixed paraffin embedded clinical breast cancer tissues towards the characterization of protein abundance characteristics which are indicative of disease stage, stage-specific characteristics, such as lymph-node involvement, as well as of disease recurrence in breast cancer.

## **2.0. CHAPTER 1: DEFINING CENTRAL THEMES IN BREAST CANCER BIOLOGY BY DIFFERENTIAL PROTEOMICS: CONSERVED REGULATION OF CELL SPREADING AND FOCAL ADHESION KINASE**

The molecular heterogeneity encountered in breast cancer underscores the importance of better defining the gene and protein abundance patterns that are conserved across breast cancer cells with the goal of defining context-specific characteristics that will aid in the development of more targeted therapeutic strategies. The combination of global proteomic analyses with bioinformatic analyses of functional protein databases can provide insights into the overarching biological characteristics present in a sample analyzed. The definition of protein abundance and subsequent functional biological characteristics that are conserved across breast cancer cells will clarify the central themes underlying neoplastic progression and further aid to discern disease subtype-specific characteristics.

Although several groups have reported global, proteomic analyses of various breast cancer cell line models<sup>82-88</sup>, few focused analyses of the functional biological characteristics associated with differentially abundant proteins across breast cancer cell types have been described.<sup>69, 89, 90</sup> We have therefore undertaken a comparative global, mass spectrometry (MS)-based proteomic analysis of three divergent cell line models of breast cancer reflective of the most common clinical breast cancer subtypes encountered<sup>25, 27, 28, 30</sup> and a model of normal mammary epithelial cells with the goal of better defining protein abundance and functional biological characteristics which are mutually shared across breast cancer cells. The three breast cancer cell lines investigated include: MCF7, an estrogen receptor (ER) positive breast cancer cell line that exhibits luminal-like characteristics and is non-invasive, yet tumorigenic in the presence of

exogenous estrogen<sup>169-173</sup>; SK-BR-3, a luminal-like, HER2/*neu* positive breast cancer cell line that is non-invasive, yet tumorigenic; and MDA-MB-231, a basal/mesenchymal-like ER-/PR-/HER2- (“triple negative”), highly invasive and aggressively tumorigenic breast cancer cell line.<sup>169-172, 174</sup> The proteomic analyses of these three breast cancer cell lines were compared with a model of normal mammary epithelial cells: MCF10A, a non-tumorigenic, spontaneously immortalized breast cell line reflective of normal mammary epithelium that exhibits basal-like characteristics.<sup>169, 171, 172, 175</sup>

A systems analysis utilizing Ingenuity Pathway Analysis (IPA) revealed significant modulation of extracellular matrix (ECM), plasma membrane and nuclear-localized proteins across tumor cell lines relative to MCF10A cells. These characteristics have been previously found to be indicative of breast cancer cells, with ECM modulation specifically as being shown to be an integral event underlying breast cancer pathogenesis.<sup>11, 12, 176</sup> Functional analyses of differentially abundant proteins across breast cancer cell lines revealed modulation of proteins associated with regulation of cell morphology, specifically with losses of proteins previously shown to promote “cell spreading”. These observations are supported by the anchorage-independent phenotypes maintained by the breast cancer cell lines analyzed.<sup>170, 174</sup> The present investigation further resulted in the identification of 82 proteins which are mutually, differentially abundant across three divergent cell line models of breast cancer. This subset included proteins which have previously been shown to regulate the focal adhesion complex protein focal adhesion kinase (FAK), a fundamental regulator of cell spreading and a key player in breast cancer pathogenesis.<sup>177</sup> Differentially abundant protein populations reveal that modulation of ECM and nuclear protein composition, as well as proteins associated with regulation of cell spreading and

focal adhesion kinase expression levels and activity are characteristics which are mutually conserved across breast cancer cells.

## **Materials and Methods**

### **Cell Lines and Lysate Preparation**

MCF10A, MCF7, SK-BR-3 and MDA-MB-231 cell lines were purchased from ATCC (American Type Culture Collection) and cultured in DMEM/F12K media supplemented with 10% fetal bovine serum (Hyclone, Logan, UT). Lysates were prepared from ~70% confluent plates by scraping cells into 150  $\mu$ l of 1X SDS buffer (10% SDS and 10 mM Tris-HCl, pH 7.4). Samples were sonicated twice for 10 s at an energy of 1.5 W utilizing a micro-tip sonicator (Misonix, Indianapolis, IN) followed by centrifugation at 14,000  $\times g$  for 10 min. Protein concentration of the supernatants was determined by the bicinchoninic assay (BCA) (Pierce, Rockford, IL).

### **Sample Preparation for LC-MS/MS**

Thirty-five  $\mu$ g of total cell lysate were resolved in two consecutive gel lanes via 1D SDS-PAGE on a 4-12% Bis-Tris gel (NuPAGE, Invitrogen, Carlsbad, CA). Gels were stained with Coomassie blue (SimplyBlue SafeStain, Invitrogen) and duplicate lanes were cut into ten equivalently sized slices. Gel slices were de-stained in 50% acetonitrile (ACN) and 50 mM ammonium bicarbonate (AMB) at ambient temperature for 3 h followed by dehydration in 100% ACN. Reactive cysteine residues were reduced via re-hydration of gel spots in 10 mM DTT, 25 mM AMB followed by incubation at 56  $^{\circ}$ C for 30 min and alkylated via incubation in 55 mM

iodoacetamide, 25 mM AMB for 30 min at ambient temperature in darkness. Gel slices were then washed with 25 mM AMB, dehydrated in 100% ACN and re-hydrated in 25 mM AMB containing 20  $\mu\text{g/mL}$  porcine sequencing grade modified trypsin (Promega, Madison, WI) on ice for 30 min. Digestions were incubated for 16 h at 37°C. Tryptic peptides were extracted with 70% ACN, 5% formic acid (FA), dried by vacuum centrifugation and re-suspended in 40  $\mu\text{l}$  of 0.1% trifluoroacetic acid (TFA). In addition, 10 fmol/ $\mu\text{L}$  of three internal standard peptides, VTIASLP<sup>13</sup>C<sub>10</sub>R, FLVGPDPGIPIMox<sup>13</sup>C<sub>10</sub>R and TGISPALI<sup>13</sup>C<sub>6</sub>K, were added to each sample digest to monitor LC-MS/MS analytical performance parameters.

### **Liquid chromatography-tandem mass spectrometry analyses**

Peptide digests were resolved by nanoflow reverse-phase liquid chromatography (Ultimate 3000, Dionex Inc., Sunnyvale, CA) coupled online via electrospray ionization to a hybrid linear ion trap-Orbitrap mass spectrometer (LTQ-Orbitrap, ThermoFisher Scientific, Inc., San Jose, CA). Duplicate injections of 5  $\mu\text{L}$  of peptide extracts were resolved on 100  $\mu\text{m}$  i.d. by 360  $\mu\text{m}$  o.d. by 200 mm long fused silica capillary columns (Polymicro Technologies, Phoenix, AZ) slurry-packed in-house with 5  $\mu\text{m}$ , 300 Å pore size C-18 silica-bonded stationary phase (Jupiter, Phenomenex, Torrance, CA). After sample injection, peptides were eluted from the column using a linear gradient of 2% mobile phase B (0.1% FA in ACN) to 40% mobile phase B over 125 min at a constant flow rate of 200 nL/min followed by a column wash consisting of 95% B for an additional 30 min at a constant flow rate of 400 nL/min. The LTQ-Orbitrap MS was configured to collect high resolution ( $R=60,000$  at  $m/z$  400) broadband mass spectra ( $m/z$  375-1800) from which the seven-most abundant peptide molecular ions dynamically determined from the MS scan were selected for tandem MS using a relative CID energy of 30%. Dynamic exclusion was

utilized to minimize redundant selection of peptides for CID. Analytical metrics of LC-MS/MS performance (mass measurement accuracy and retention time) were monitored for each sample analyzed using three stable isotope, internal standard peptides by calculating reconstructed mass chromatogram areas for each standard peptide using QuanBrowser (XCalibur v2.1, ThermoScientific). QuanBrowser processing method parameters tolerated  $\pm 15$  ppm mass accuracy and  $\pm 5.0$  min in retention time relative to monoisotopic masses and retention times estimated from manual inspection of internal standard peptides from representative raw data.

### **Peptide Identification and Spectral Count Analysis**

Peptide identifications were obtained by searching the LC-MS/MS data utilizing SEQUEST (BioWorks, v3.2, ThermoScientific) on a 72-node Beowulf cluster against a UniProt-derived human proteome database (version 10/08, 56,301 protein entries) obtained from the European Bioinformatics Institute (EBI) using the following parameters: trypsin (KR); full enzymatic-cleavage; two missed cleavages sites; 20 ppm peptide mass tolerance peptide tolerance, 0.5 amu fragment ion tolerance; and variable modifications for methionine oxidation ( $m/z$  15.99492) and cysteine carboxyamidomethylation ( $m/z$  57.02146); data analysis revealed 10.7% ( $\pm 1.3\%$ ) of all cysteine-containing peptides were non-reacted. Resulting peptide identifications were filtered according to specific SEQUEST scoring criteria: delta correlation ( $\Delta C_n$ )  $\geq 0.08$  and charge state dependent cross correlation (Xcorr)  $\geq 1.9$  for  $[M+H]^+$ ,  $\geq 2.2$  for  $[M+2H]^{2+}$ , and  $\geq 3.5$  for  $[M+3H]^{3+}$ . These criteria resulted in a false discovery rate of 0.4% for all peptides identified as determined by searching the data set against a decoy human database where the protein sequences were reversed.<sup>178</sup> Differences in protein abundance between the samples were derived by spectral counting (SC) and peptides whose sequence mapped to multiple protein isoforms

were grouped as per the principle of parsimony.<sup>179</sup> Equivalency in peptide load was also determined by comparison of the spectral count values for the protein actin, cytoplasmic 1 (ACTB) across all samples.<sup>180</sup> Differential protein abundance was calculated by normalizing total peptide spectral counts observed in the LC-MS/MS analysis of MCF7, SK-BR-3 and MDA-MB-231 conditions against total spectral counts from MCF10A. A value of 0.5 was added to each spectral count value prior to  $\log_2$  transformation to enable ratio values to be calculated for proteins identified in one cell line, but not another.<sup>181</sup> Proteins which exhibited a  $> 95\%$  confidence interval from the mean for each comparison performed were considered statistically significant.

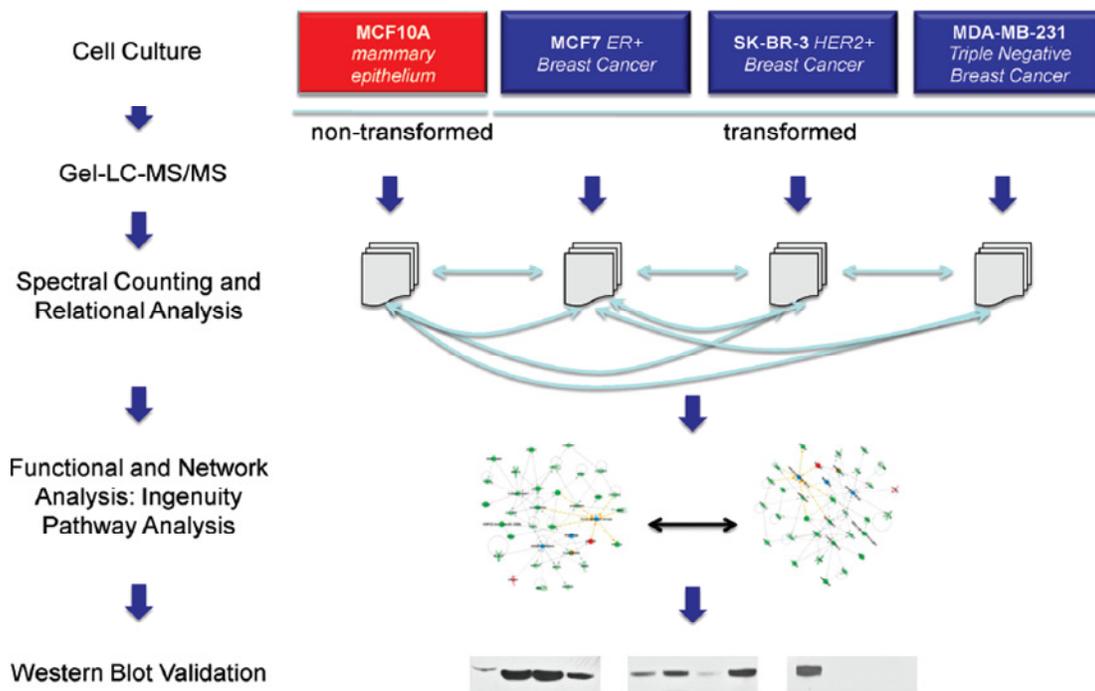
### **Bioinformatic Analyses**

Uniprot accessions corresponding to proteins identified by at least two peptides were mapped to HUGO (HGNC) gene symbols utilizing Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems, [www.ingenuity.com](http://www.ingenuity.com)). Accessions which failed to map were converted to IPI identifiers with the mapping utility available at [www.uniprot.org](http://www.uniprot.org) and re-mapped to IPA to maximize protein identifications available for downstream bioinformatic analyses. Protein localization and subtype assignments were derived from IPA-mapped data sets. Statistical analyses for the enrichment of specific protein sub-types amongst differentially-expressed proteins were performed utilizing a two proportions–z test, with a significance level set to  $p \leq 0.05$ . Functional analysis of significant protein lists were performed utilizing the “Core Analysis” function in IPA using default parameters. Significant ( $p < 0.05$ ) biofunctions containing a minimum of two associated proteins were correlated from each “Core Analysis” by functional annotation across global significant protein lists. Network and protein interaction analyses were also performed

utilizing IPA for mutually, differentially abundant proteins in which a maximum of 35 proteins for network assignment were allowed.

### **Western Blot Analyses**

Twenty-five micrograms of total cell lysate derived from biological replicates of cell line conditions analyzed was resolved via 1D SDS-PAGE and resulting gels were transferred to a PVDF membrane at a constant 30V / 160 mA for 2 h. Blots were blocked for 1 h at 37 °C in 1X blotto (5% dehydrated milk, 0.1% Tween-20 and 1X TBS) followed by incubation in primary antibody diluted in 1X blotto overnight at 4 °C. Blots were washed in 1X TBST (1X TBS, 0.1% Tween-20 (Sigma Aldrich, St. Louis, MO) followed by incubation in secondary antibody for 1 h at ambient temperature. Blots were washed again in 1X TBST followed by incubation in chemiluminescent substrate (SuperSignal West Pico or Dura, Rockford, IL) for 5 min at ambient temperature. Primary and secondary antibody conditions utilized were as follows: goat anti-rabbit IgG-HRP (1:10,000) (Abcam, Cambridge, MA: ab6721), goat anti-mouse IgG-HRP (1:50,000) (Pierce: 31432), goat, anti-Rat IgG HRP (1:10,000) (RND Systems, Minneapolis, MN: HAF005), anti-beta actin (1:10,000) (Abcam: ab6276), anti-Cytokeratin 19 (1:1000) (Abcam: ab15463), anti-CD10 (MME/Neprilysin) (1:1000) (Abcam: ab951), anti-integrin alpha 11 (1:250) (RND Systems: MAB4235), anti-TGFBI (1:1000) (Abcam: ab89062), anti-FAK (1:1000) (Abcam: ab40794), anti-Hic5 (1:1000) (TGFBI1) (BD Biosciences, Franklin Lakes, NJ: BDB611164), anti-TNC (1:200) (RND Systems: MAB2138).

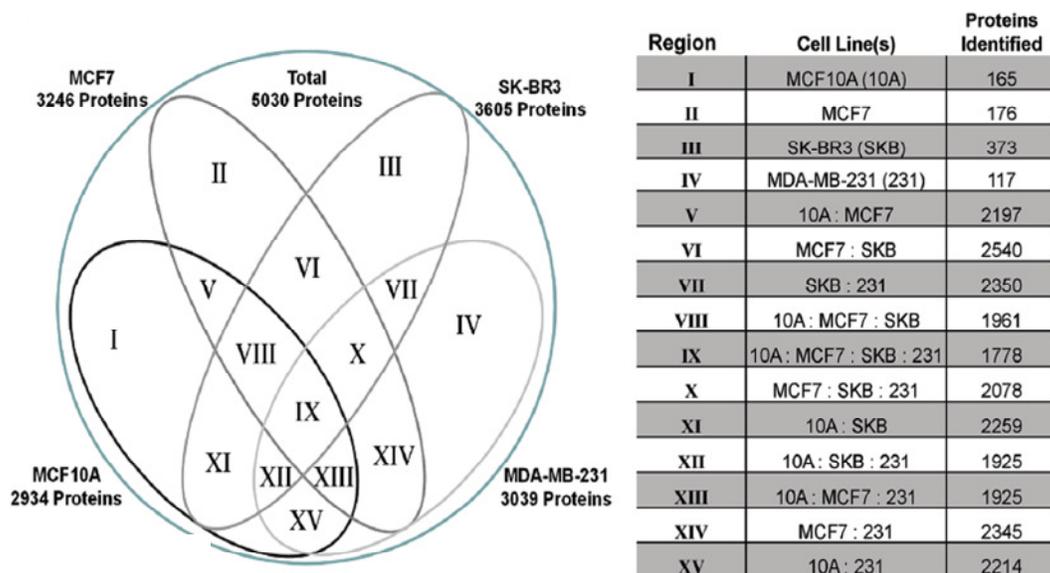


**Figure 5: Analytical workflow utilized for differential and comparative proteomic analysis four breast cell lines analyzed.**

## Results and Discussion

### Global proteomic analysis of human breast cell line models

Global proteomic analyses of three breast cancer cell line models (MCF7, SK-BR-3 and MDA-MB-231) reflective of the most common clinical breast cancer subtypes encountered were compared with a model of normal mammary epithelial cells MCF10A to identify central themes in protein abundance and functional characteristics underlying the transformed phenotypes of these key model systems of breast cancer (Fig. 5 - Workflow). Duplicate, LC-MS/MS analyses of tryptic peptides corresponding to 70  $\mu$ g of total protein from each cell line condition on an LTQ-Orbitrap resulted in a total of 39,158 peptide identifications and 5,030 proteins identified



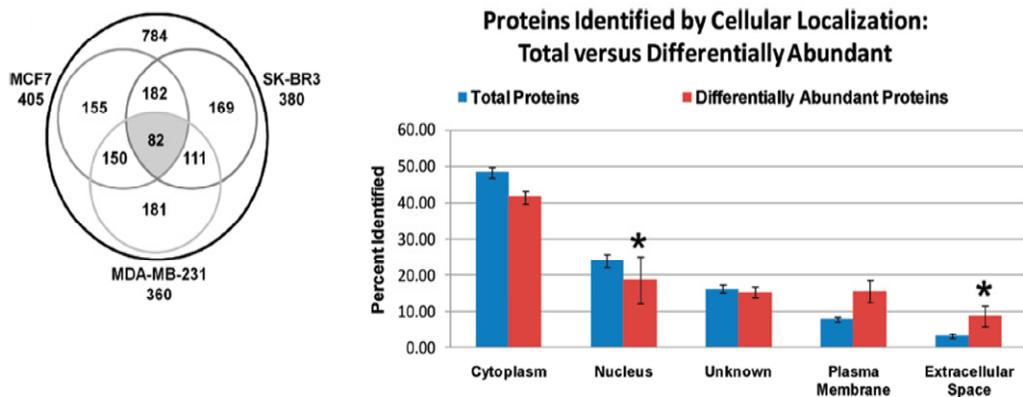
**Figure 6: Protein identification characteristics across four breast cell lines analyzed.** A four-way Venn diagram detailing the total number of proteins identified and co-identified in each breast cell line analyzed. Proteins reported were derived from two technical LC-MS/MS replicates of each cell line and were identified by at least two peptides.

by at least two peptides (Fig. 6). Equivalency of protein digest input for each cell line was determined by comparison of the total number of peptides identified across all four cell lines analyzed by LC-MS/MS, revealing a relative standard deviation (RSD) of 12.3%, with total mean RSD peptide identification rates between replicate sample injections of 2.7% ( $\pm 2.7\%$ ). Equivalent peptide digest input was further confirmed by comparison of total peptides identified for beta-actin (ACTB, Fig. 11) commonly utilized as a loading control for western blot analysis,<sup>180</sup> which revealed an RSD value of 23.1% across all cell lines analyzed. Metrics of analytical performance were further monitored by comparison of retention time and mass chromatogram area variance for three internal standard peptides that were added to each sample

prior to LC-MS/MS analysis. These results show an aggregate peak area RSD of 25.1% ( $\pm$  9.6%) for the three control peptides across the 80 LC-MS/MS analyses that were conducted. Taken together, these results provide strong evidence that the global protein digest input into the LC-MS/MS analyses was equivalent across all cell lines analyzed. This observation critically underpins determination of significant, differentially abundant proteins by spectral counting without the need for normalization. These analyses resulted in the coincident identification of 1,778 proteins across all cell lines (Fig. 6, region IX). Proteins identified in both MCF7 and SK-BR-3 (2,540 proteins, Fig. 6, region VI) revealed the greatest overlap in a qualitative two-way comparison and between MCF7, SK-BR-3 and MDA-MB-231 (2,078) in a three-way comparison (Fig. 6, region X). Interestingly, the greater overlap of proteins identified between MCF7 and SK-BR-3 cells is consistent with these cell lines both exhibiting characteristics of luminal-type breast cancer cells.<sup>169, 171, 172</sup>

Uniprot accession numbers corresponding to proteins identified were mapped to HUGO gene symbols (HGNC) utilizing IPA, yielding a final list of 4,859 proteins that were included in subsequent bioinformatic analyses. Comparison of known cellular localization and protein subtype characteristics for final mapped data sets revealed that proteins were consistently identified from similar cellular compartments and functional sub-types between the four cell lines analyzed.

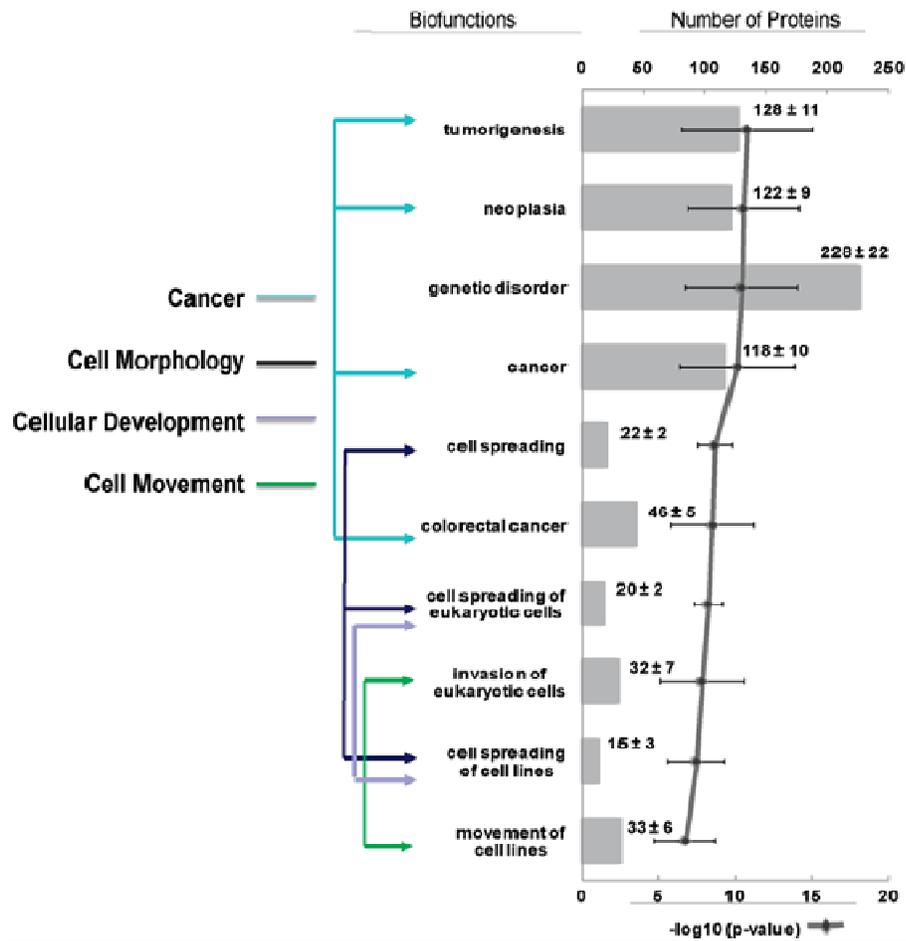
**Comparative analysis of proteins differentially abundant in individual comparisons of breast cancer cell lines versus non-transformed MCF10A cells.**



**Figure 7: Identification characteristics for significantly, differentially abundant proteins in breast cancer cell lines. A.** A three-way Venn diagram comparing the number of significantly, differentially abundant proteins observed across breast cancer cell lines relative to MCF10A cells. The central “grayed” region designates the 82 proteins which were observed as being mutually, differentially abundant across all breast cancer cell lines relative to MCF10A cells. **B.** Comparison of proteins identified by cellular localization across global proteins identified versus significantly, differentially abundant proteins observed in breast cancer cell lines. Significant, differential enrichment is designated by \*,  $p < 0.05$ .

Spectral count ratios of proteins identified in the breast cancer cell lines MCF7, SK-BR-3 and MDA-MB-231 exhibiting a >95% confidence interval relative to proteins identified in MCF10A cells were considered as significantly differentially abundant, with the average cut-off for increased abundance being  $\log_2(3.48 \pm 0.48)$  and decreased abundance  $\log_2(-3.04 \pm 0.14)$ . The results revealed a total of 784 unique protein groups as differentially abundant across all breast cancer cell lines analyzed, with 379 of these (mean of  $163 \pm 5$ ) as increased and 405 (mean of  $219 \pm 26$ ) as decreased in abundance (Fig. 7A). Significant, differentially abundant proteins amongst the breast cancer cell lines include hallmark candidates that are well established in the

literature, such as increased abundance of E-cadherin (CDH1) observed in MCF7 cells ( $\log_2$  4.6 fold-change), high levels of HER2/*neu* (ERBB2) in SK-BR-3 cells ( $\log_2$  5.4) and high EGFR levels in MDA-MB-231 cells ( $\log_2$  4.2).<sup>169, 172, 182</sup> Significant increases in the number of ECM localized proteins and decreases in nuclear-localized proteins were observed amongst differentially abundant proteins across breast cancer cell lines (Fig. 7B). This profile is supported by the notion that modulation of nuclear and ECM protein composition are characteristic of breast cancer cells and occurs during breast tumorigenesis.<sup>10-12, 16, 176</sup> To define overarching functional characteristics amongst differentially abundant proteins significant ( $p < 0.05$ ) biological functions (biofunctions) observed via “Core Analysis” (IPA) of protein lists from each tumor cell line were correlated. A total of 10 significant biofunctions overlapped between proteins with increased abundance and 147 between proteins decreased in abundance across all tumor cell line conditions. These results indicate that proteins decreased in abundance across the breast cancer cell lines share greater functional similarities in comparison with proteins increased in abundance relative to MCF10A. The top ten scoring overlapping biofunctions between differentially abundant proteins from all breast cancer cell lines (Fig. 8) were associated with two predominant functional categories, cancer (4/10 total) and “cell morphology” (3/10), and included specific sub-categories, such as tumorigenesis, as well as those relating to cell spreading, respectively. In comparing these results with the top ten scoring biofunctions observed for proteins exhibiting increased or decreased abundance, coordinate losses and gains in proteins associated with the functional category of cancer are observed across all breast cancer cell lines (Fig. 8). Proteins associated with the biofunction category “cell morphology” were predominantly decreased in abundance across tumor cell lines, however, with the majority being associated with the specific sub-category of cell spreading (Fig. 8).



**Figure 8: Identification of functional characteristics across proteins significantly, differentially abundant in breast cancer cell lines.** Top 10 significant biofunctions observed across differentially abundant protein lists identified from each breast cancer cell line. Figure reports average p-values observed and mean number of proteins which clustered with a given biofunction.

Cell spreading describes the transition a cell undergoes from a rounded to a more “flattened” morphology during anchoring to a substratum.<sup>183, 184</sup> This event is mediated via reorganization of the cellular actin cytoskeleton and through the establishment of focal adhesion complexes,

predominantly comprising cellular integrins, that mediate communication between the intracellular and extracellular environments through interactions with ECM proteins.<sup>12, 183, 184</sup> ECM anchoring is required for normal epithelial cell survival, which is an event termed anchorage-dependence, with survival in the absence of these interactions, or anchorage-independence, being a hallmark characteristic of cellular transformation<sup>17, 184</sup> In the normal mammary epithelium, ECM interactions are maintained with the collagen- and laminin-rich basement membrane (BM) separating the epithelial compartment from adjacent stromal tissues and these interactions provide cues that function to inhibit apoptosis and promote cell survival.<sup>16, 185, 186</sup> Anchorage-dependent MCF10A cells have been shown to actively secrete laminins in culture, specifically laminin-5, the expression of which has been further shown to be characteristic of non-tumorigenic mammary epithelial cells.<sup>12, 185-187</sup> Assessment of laminin-5 abundance in our data revealed high abundances in MCF10A cells and significantly decreased levels in MDA-MB-231 cells, as well as low levels in MCF7 and SK-BR-3 conditions. Further, MCF10A cells exhibit morphological characteristics of normal mammary epithelial cells in culture, having a flattened appearance and forming defined epithelial sheets.<sup>182, 188, 189</sup> MCF7 and SK-BR-3 cells deviate from these characteristics in that they form tight colonies of polygonal cells in culture, with SK-BR-3 cells exhibiting even greater morphological heterogeneity that includes semi-adherent, spheroidal cell populations, and are thus considered less differentiated than MCF7 cells.<sup>182</sup> Lastly, MDA-MB-231 cells exhibit a mesenchymal or fibroblastic-like morphology in culture as they have undergone epithelial to mesenchymal transition, forming erratic networks of spindle shaped cells.<sup>182</sup> Comparison of cell spreading and subsequent cell survival in normal versus H-RAS-transformed fibroblasts has revealed a direct correlation between the degree of cell spreading and survival in normal cells that is not observed in

transformed cell types.<sup>189</sup> These findings clarify the observed loss of proteins associated with cell spreading in the breast cancer cell lines analyzed as, unlike MCF10A cells, all three breast cancer cell lines analyzed are capable of sustaining anchorage-independent growth.<sup>170, 174</sup> Therefore, as MCF7, SK-BR-3 and MDA-MB-231 cell lines represent a range of more to less differentiated breast cancer cell types,<sup>182</sup> respectively, and as a significant loss of proteins associated with the regulation of cell spreading is observed across all conditions, these alterations may represent early events in the transformation of mammary epithelial cells.

### **Comparative analysis of mutually, differentially abundant proteins observed across breast cancer cell lines versus non-transformed MCF10A cells.**

A comparative analysis of mutually, differentially abundant proteins revealed a total of 82 proteins that were identified in all three breast cancer cell lines relative to MCF10A cells (Table 2). The majority of these proteins were decreased in abundance (n=68) in all breast cancer cell lines versus a small subset which were increased (n=13). A single exception to this was filamin C gamma (FLNC), a protein commonly increased in mesenchymal cell types,<sup>169</sup> which was significantly increased in MDA-MB-231 cells ( $\log_2$  4.1) though decreased in MCF7 and SK-BR-3 cells ( $\log_2$  -4.2 ( $\pm$  1.1)). Unlike the other breast cancer cell lines analyzed, this observation is consistent with MDA-MB-231 cells having undergone epithelial to mesenchymal transition and may be a novel characteristic of this event. Consistent with previous reports comparing normal and breast cancer cells and tissues, we observed increased levels of keratin-19 (KRT19) (Fig. 11)<sup>169, 190</sup>, glutathione s-transferase mu 3 (GSTM3)<sup>85</sup>, as well as decreases in levels of tropomyosin 1.<sup>83</sup> Analysis of protein localization amongst this mutually dysregulated protein subset revealed

significant increases in both plasma membrane and ECM proteins and decreases in nuclear proteins relative to total, differentially abundant protein populations identified across all breast cancer cell lines (Figure 9A). This observation further underscores the modulation of ECM and nuclear protein composition as central themes of breast cell transformation, further adding modulation of plasma membrane protein composition as a characteristic conserved across breast cancer cells.<sup>10, 12, 16</sup> Evaluation of this subset of 82 differentially abundant proteins revealed decreases in several canonical basement membrane proteins, such as the proteoglycan heparin sulfate proteoglycan 2 (HSPG2) as well as several collagen isoforms, including two type 4 collagens, alpha 1 (COL4A1) and alpha 2 (COL4A2) (Table 2). HSPG2 and type 4 collagens are well known constituents of the basement membrane in normal mammary epithelium and loss of expression is commonly observed during breast cancer pathogenesis.<sup>191-193</sup> Interestingly, a 26-kDa region of COL4A1, termed arresten, has been shown to possess tumor suppressor and anti-angiogenic properties in murine models of tumor metastasis.<sup>194-196</sup> Further, a 24-kDa region of COL4A2, termed canstantin, has also been similarly shown to inhibit tumor growth and discourage angiogenesis in xenograft models of tumorigenesis and in *in vitro* analyses of endothelial cells, respectively.<sup>194, 196, 197</sup> Indeed, coordinate losses of various basement membrane proteins, several of which have been shown to possess tumor suppressor functions, across breast cancer cell lines analyzed underscores their anchorage-independent phenotypes and is consistent with their tumorigenic phenotypes.

**Table 2: Mutually, Differentially Abundant Proteins Observed Across B.Cancer Cells**

HGNC symbol	Entrez gene name	average (log <sub>2</sub> -fold)	standard deviation	cellular localization	functional type
ABCC3	ATP-binding cassette, subfamily C (CFTR/MRP), member 3	-4.52	N/A	Plasma Membrane	transporter
ADNP	activity-dependent neuroprotector homeobox	4.15	0.85	Nucleus	transcription regulator
AIDA	axin interactor, dorsalization associated	-3.17	N/A	unknown	other
ALDH1L2	aldehyde dehydrogenase 1 family, member L2	-3.70	N/A	unknown	enzyme
ANPEP	alanyl (membrane) aminopeptidase	-8.13	N/A	Plasma Membrane	peptidase
ANTXR1 <sup>a</sup>	anthrax toxin receptor 1	-3.70	N/A	Plasma Membrane	other
APAF1	apoptotic peptidase activating factor 1	-3.17	N/A	Cytoplasm	other
ASPH	aspartate beta-hydroxylase	-5.29	1.34	Cytoplasm	enzyme
BIN1	bridging integrator 1	-3.46	N/A	Nucleus	other
BIRC6	baculoviral IAP repeat-containing 6	4.03	0.75	Cytoplasm	enzyme
C10ORF58	chromosome 10 open reading frame 58	4.27	0.76	Extracellular Space	other
CD109	CD109 molecule	-5.20	0.92	Plasma Membrane	other
CDH2	cadherin 2, type 1, N-cadherin (neuronal)	-4.25	N/A	Plasma Membrane	other
CNN1	calponin 1, basic, smooth muscle	-4.95	N/A	Cytoplasm	other
CNTNAP1	contactin associated protein 1	-4.25	N/A	Plasma Membrane	other
COL12A1	collagen, type XII, alpha 1	-7.94	1.34	Extracellular Space	other
COL1A1	collagen, type I, alpha 1	-7.22	0.92	Extracellular Space	other
COL3A1	collagen, type III, alpha 1	-3.17	N/A	Extracellular Space	other
COL4A1	collagen, type IV, alpha 1	-3.17	N/A	Extracellular Space	other
COL4A2	collagen, type IV, alpha 2	-3.91	N/A	Extracellular Space	other
COL7A1	collagen, type VII, alpha 1	-4.09	N/A	Extracellular Space	other
COL8A1	collagen, type VIII, alpha 1	-4.25	N/A	Extracellular Space	other
CRADD	CASP2 and RIPK1 domain containing adaptor with death domain	-3.17	N/A	Cytoplasm	other
CRYAB	Crystallin, alpha B	-3.46	N/A	Nucleus	other
DFNA5	deafness, autosomal dominant 5	-3.46	N/A	unknown	other
ELAC2	elaC homologue 2 ( <i>E. coli</i> )	4.49	0.57	Nucleus	enzyme
ENG	endoglin	-3.46	N/A	Plasma Membrane	other
FAP	fibroblast activation protein, alpha	-4.25	N/A	Cytoplasm	peptidase
FBLIM1	filamin binding LIM protein 1	-4.25	N/A	Plasma Membrane	other
FBN1 <sup>a</sup>	fibrillin 1	-5.14	0.92	Extracellular Space	other
FGF2 <sup>a</sup>	fibroblast growth factor 2 (basic)	-3.17	N/A	Extracellular Space	growth factor
FLNC <sup>a</sup>	filamin C, gamma	-1.42	4.82	Cytoplasm	other
FRMD6	FERM domain containing 6	-3.46	N/A	unknown	other
FSCN1	fascin homologue 1, actin-bundling protein (Strongylocentrotus purpuratus)	-5.73	N/A	Cytoplasm	other
GALNT2	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)	-4.23	0.92	Cytoplasm	enzyme
GBP2	guanylate binding protein 2, interferon-inducible	-3.91	N/A	Cytoplasm	enzyme
GPC1	glypican 1	-5.20	0.92	Plasma Membrane	transmembrane receptor
GSTM3	glutathione S-transferase mu 3 (brain)	6.07	0.41	Cytoplasm	enzyme
HLA-A	major histocompatibility complex, class I, A	-3.96	N/A	Plasma Membrane	transmembrane receptor
HSPG2	heparan sulfate proteoglycan 2	-7.37	N/A	Plasma Membrane	other
HTRA1	HtrA serine peptidase 1	-3.17	N/A	Extracellular Space	peptidase
ITGA11 <sup>a</sup>	integrin, alpha 11	-4.52	0.92	Plasma Membrane	other
IVNS1ABP	influenza virus NS1A binding protein	-3.17	N/A	Nucleus	other
JPH2	junctophilin 2	-3.46	N/A	Cytoplasm	enzyme
KRT17	keratin 17	-4.99	2.71	Cytoplasm	other
KRT19	keratin 19	8.40	1.74	Cytoplasm	other
KRT80	keratin 80	4.79	0.97	unknown	other
LMOD1	leiomodlin 1 (smooth muscle)	-3.70	0.00	Cytoplasm	other
LPGAT1	lysophosphatidylglycerol acyltransferase 1	-3.17	N/A	Cytoplasm	other
LRBA	LPS-responsive vesicle trafficking, beach and anchor containing	5.04	1.43	Cytoplasm	other
LRP1	low density lipoprotein-related protein 1 (alpha-2-macroglobulin receptor)	-6.01	0.92	Plasma Membrane	transmembrane receptor
MAGI3	membrane associated guanylate kinase, WW and PDZ domain containing 3	-3.70	N/A	Cytoplasm	kinase
MALT1	mucosa associated lymphoid tissue lymphoma translocation gene 1	-3.46	N/A	Cytoplasm	peptidase
MCAM	melanoma cell adhesion molecule	-5.78	N/A	Plasma Membrane	other
MME	membrane metallo-endopeptidase	-3.46	N/A	Plasma Membrane	peptidase
MRC2	mannose receptor, C type 2	-4.52	0.92	Plasma Membrane	transmembrane receptor
MYLK	myosin light chain kinase	-4.23	0.92	Cytoplasm	kinase
NDUFV3	NADH dehydrogenase (ubiquinone) flavoprotein 3, 10 kDa	-3.17	N/A	Cytoplasm	enzyme
NEK7	NIMA (never in mitosis gene a)-related kinase 7	-3.17	N/A	Nucleus	kinase
P4HA2	prolyl 4-hydroxylase, alpha polypeptide II	-5.93	N/A	Cytoplasm	enzyme
PALLD <sup>a</sup>	palladin, cytoskeletal associated protein	-4.58	1.76	Cytoplasm	other
PAPD4	PAP associated domain containing 4	-4.25	N/A	unknown	enzyme
PC	pyruvate carboxylase	5.37	1.33	Cytoplasm	enzyme

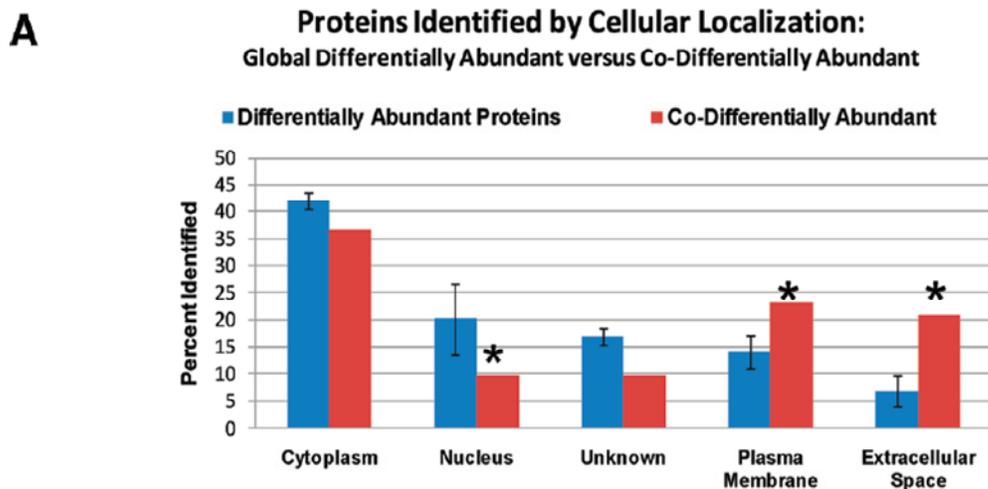
PDLIM7	PDZ and LIM domain 7 (enigma)	-4.50	1.62	Cytoplasm	other
PHLDB1	pleckstrin homology-like domain, family B, member 1	-3.70	0.00	unknown	other
PNPT1	polyribonucleotide nucleotidyltransferase 1	4.23	0.28	Cytoplasm	enzyme
POSTN <sup>a</sup>	periostin, osteoblast specific factor	-6.10	0.92	Extracellular Space	other
PYCR1	pyrroline-5-carboxylate reductase 1	4.20	0.87	unknown	enzyme
SEC24B	SEC24 family, member B ( <i>S. cerevisiae</i> )	3.88	0.81	Cytoplasm	transporter
SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5	-6.15	N/A	Extracellular Space	other
SLC26A2	solute carrier family 26 (sulfate transporter), member 2	-3.70	N/A	Plasma Membrane	transporter
SLC44A1	solute carrier family 44, member 1	-3.46	N/A	Plasma Membrane	transporter
SYNPO	synaptopodin	-3.91	N/A	Cytoplasm	other
TAGLN	transgelin	-6.06	1.87	Cytoplasm	other
TF	transferrin	-3.17	N/A	Extracellular Space	transporter
TGFB111 <sup>b</sup>	transforming growth factor beta 1 induced transcript 1	-4.95	N/A	Nucleus	transcription regulator
TGFBI <sup>a</sup>	transforming growth factor, beta-induced, 68 kDa	-7.74	N/A	Extracellular Space	other
TNC <sup>b</sup>	tenascin C	-8.29	0.92	Extracellular Space	other
TNS1	tensin 1	-3.91	N/A	Plasma Membrane	other
TPM1 <sup>a</sup>	tropomyosin 1 (alpha)	-3.17	N/A	Cytoplasm	other
URB1	URB1 ribosome biogenesis 1 homologue ( <i>S. cerevisiae</i> )	4.38	0.82	Nucleus	other
VCAN	versican	-6.19	N/A	Extracellular Space	other

<sup>a</sup> Proteins associated with promotion of cell spreading. <sup>b</sup> Proteins associated with inhibition or differential regulation of cell spreading.

## Modulation of proteins associated with regulation of cell spreading is conserved across breast cancer cell lines.

Further analysis of this protein subset revealed that the top ten significant biofunctions amongst this group clustered into two predominant functional categories, cell development (4/10) and cell morphology (4/10), including specific biofunctions associated with cell adhesion (adhesion of fibroblast cell lines) and overlap between these functional categories and various subcategories associated with the regulation of cell spreading (Fig. 9B). The significance of cell spreading amongst this mutually dysregulated protein subset underscores this cellular event as being a central characteristic of breast cancer cells. The subcategories associated with cell spreading encompass proteins derived from *in vitro* characterization of this event in eukaryotic and fibroblast cell types.<sup>198-213</sup> A total of eleven proteins were associated with all four cell spreading sub-categories combined (Fig. 9B) and each were significantly decreased in abundance across all tumor cell lines relative to MCF10A cells, with the exception of FLNC which was observed at

increased levels in MDA cells. Of these 11 proteins, nine have been shown to directly promote cell spreading (Table 1, HGNC symbols designated as (a),<sup>198, 199, 201-205, 207, 209-211, 213</sup> with the remaining two having been found to either decrease or variably impact cell spreading (Table 1, HGNC symbols designated as (b)).<sup>200, 204, 206-208, 212</sup>

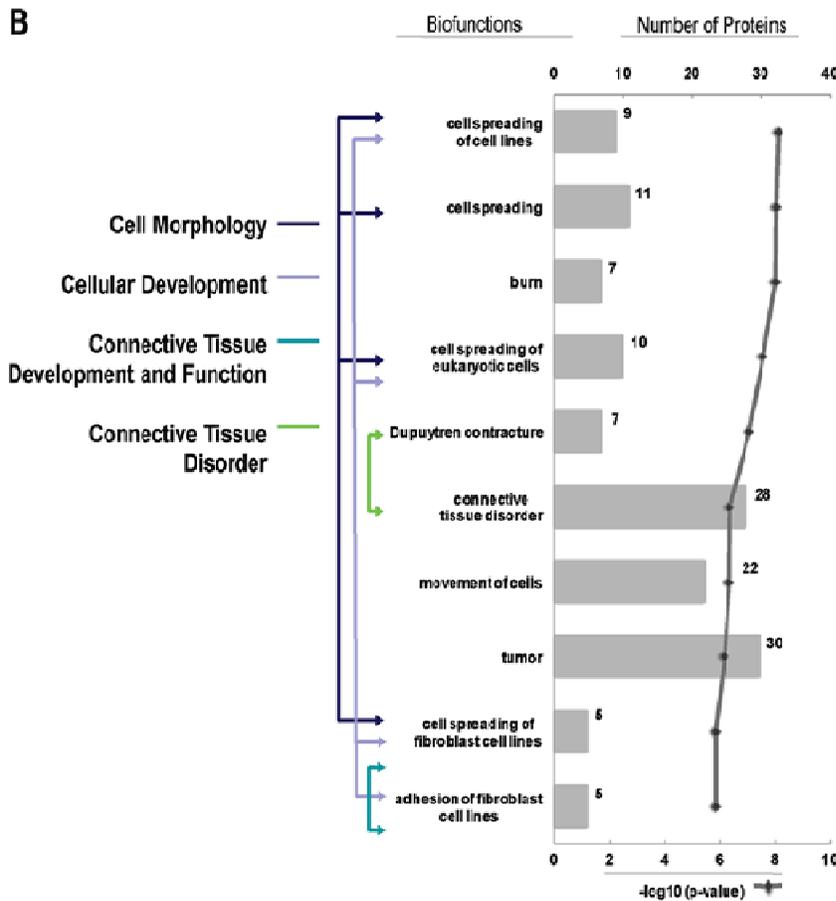


**Figure 9A: Protein identification characteristics for proteins that were mutually, differentially abundant across breast cancer cell lines. A:** Comparison of proteins identified by cellular localization across global significantly, differentially abundant proteins versus mutually (co-differentially) abundant proteins observed in breast cancer cell lines. Significant, differential enrichment is designated by \*,  $p < 0.05$ .

Several of these proteins that promote cell spreading have been shown to be mutated or lost with disease progression in breast cancer and include fibrillin (FBN1), fibroblast growth factor 2 (FGF2), transforming growth factor, beta-induced (TGFBI) and tropomyosin 1 alpha (TPM1).<sup>83, 214-216</sup> In breast cancer, single nucleotide polymorphisms have been observed in the gene encoding the structural microfibrillar, ECM glycoprotein FBN1.<sup>216</sup> Expression of the ECM

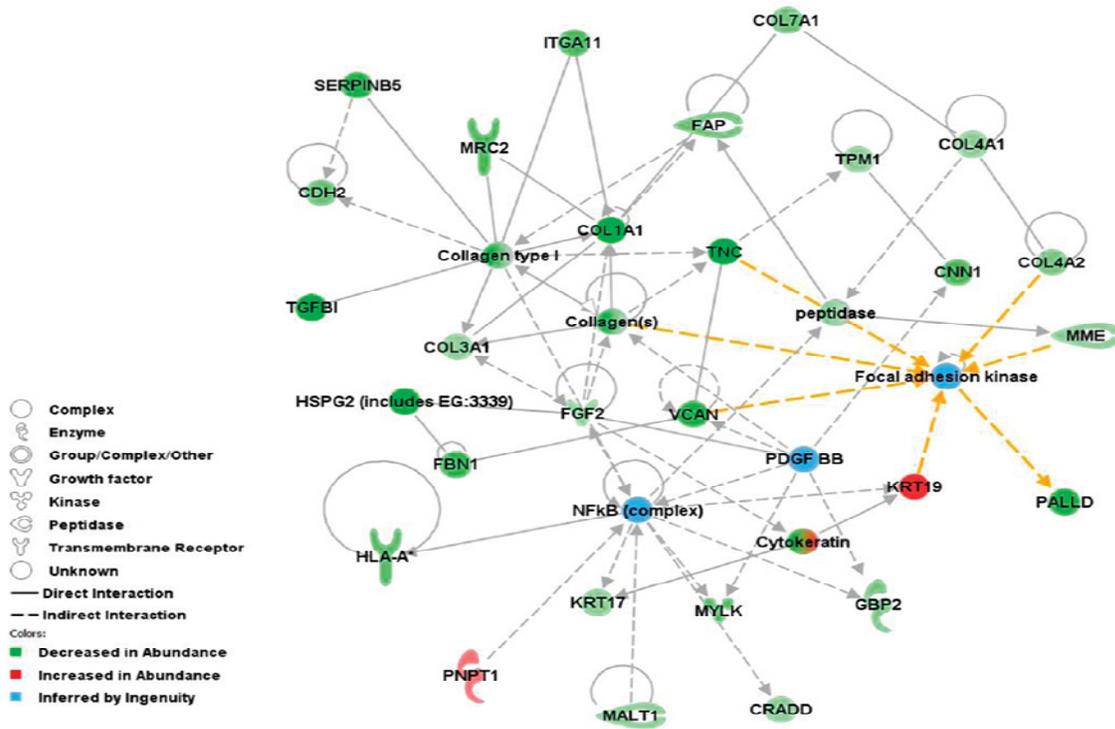
protein TGFBI, also known as BigH3 (Fig. 11), the cytoskeletal protein TPM1 as well as the growth factor FGF2 have all been shown to be lost with neoplastic progression in breast cancer and in cancerous versus normal breast cell types.<sup>83, 214, 215</sup> Conversely, three proteins previously shown to promote cell spreading, the type 1 transmembrane protein anthrax toxin receptor 1 (ANTXR1), integrin alpha 11 (ITGA11, Fig. 11) and the putative cell adhesion protein periostin, osteoblast-specific factor (POSTN), have been found to be increased in breast cancer.<sup>217-219</sup> Expression of ANTXR1, also known as Tem8, has been shown to be increased in breast cancer tissues, and further to be a marker of microvasculature in the tumor endothelium of breast cancers.<sup>217</sup> Further, POSTN has been shown to be increased in expression at both the gene and protein levels in breast cancer tissues.<sup>218, 219</sup>

The remaining two proteins observed in this subset of 11 cell spreading associated proteins have been shown to either inhibit or variably impact cell spreading, one being the transcriptional regulator transforming growth factor beta 1 induced transcript 1 (TGFB1I1, Fig. 11) and the other the ECM protein tenascin-C (TNC, Fig. 11), respectively. TGFB1I1, or Hic-5, has been shown to decrease cell spreading in fibroblast cells and, in the context of breast cancer, TGFB1I1 has been shown to play an integral role in mediating glucocorticoid signaling in breast cancer cells and capable of promoting epithelial to mesenchymal transition when ectopically expressed in MCF10A cells.<sup>206, 220, 221</sup> Additionally, TNC has been shown to variably impact cell spreading in neural and leukemia cancer cells<sup>200, 204, 212</sup> and, in the context of breast cancer, to exhibit increased levels in breast tumor stroma as well as in breast cancer cells, including MDA-MB-231 cells, further being shown to promote angiogenesis and metastasis in breast cancer through the activation of oncogenic signaling cascades such as mitogen-activated protein kinase



**Figure 9B: Protein identification characteristics for proteins that were mutually, differentially abundant across breast cancer cell lines. Top 10 significant biofunctions observed across mutually, differentially abundant proteins across breast cancer cell lines. Figure reports average, significant p-values observed and mean number of proteins observed which clustered with a given biofunction.**

(MAPK) and Wnt.<sup>222, 223</sup> TNC binds various ECM proteins such as HSPG2 as well as fibronectin, which we observe as being significantly abundant in MCF10A cells relative to the tumor cell lines analyzed. When taken together, the array of ECM proteins expressed in MCF10A cells relative to all breast cancer cell lines analyzed, combined with the propensity of



**Figure 10: Network analysis of mutually, differentially abundant proteins observed across breast cancer cell lines.** Top protein network derived from “Core Analysis” of significant co-identified proteins. Orange arrows denote those proteins which clustered specifically with FAK.

TNC to directly interact with these ECM factors, may produce an ECM context discouraging induction of aberrant signaling by TNC. The predominant loss of proteins that function to promote cell spreading across tumor cell lines is supported by the aberrant morphologies and anchorage-independent phenotypes maintained by these cell types.<sup>182, 224</sup> Further, as the majority of these factors have been shown to be lost with disease progression in breast as well as other forms of cancer,<sup>83, 214-216, 225, 226</sup> this subset of proteins may play an integral role in the onset of key events that support mammary epithelial cell transformation, such as promoting a loss of contact inhibition as well as anchorage-independent growth.

**Network analysis reveals several mutually, differentially abundant proteins cluster with focal adhesion kinase, (FAK) a key regulator of cell spreading.**

Investigation of cell signaling networks associated with this subset of differentially expressed proteins revealed the top scoring network to include 28 out of the 82 (34%) differentially expressed proteins observed across all breast cancer cell lines (Fig. 10). This 28- member network comprises predominantly ECM, plasma membrane and cytoplasmic proteins and includes several basement membrane proteins, such as HSPG2 and several collagen isoforms, in addition to seven proteins associated with regulation of cell spreading, e.g. FBN1, FGF2, ITGA11, PALLD, TGFBI, TNC and TPM1. Proteins involved in this network as inferred by IPA include the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB), platelet-derived growth factor, BB dimer variant (PDGF-BB) and focal adhesion kinase (FAK). When considering this network in the context of the recurring functional theme of cell spreading observed amongst differentially abundant proteins across the breast cancer cell lines analyzed, the assignment of FAK specifically to this network is noteworthy as this protein plays a central role in regulating the event of cell spreading in epithelial, endothelial and fibroblast cell types.<sup>177,</sup>  
<sup>227</sup> In the context of breast cancer, FAK has been shown to play key roles in disease pathogenesis, further being found to function as an obligate co-factor underlying the malignant transformation of neoplastic breast cells.<sup>177, 228, 229</sup> FAK is a non-receptor tyrosine kinase that localizes to focal adhesion complexes and is typically activated in response to integrin-ECM protein interactions, thus functioning to translate extracellular, growth factor and integrin-mediated signaling events to the intracellular environment, typically via PI3K and RAS-MAPK signaling cascades.<sup>177, 230</sup> Activation of FAK has further been shown to occur via epidermal and



Interestingly, in non-transformed cells, the activation state of FAK has been shown to be directly related to cell spreading as studies of FAK activation in cultured primary endothelial and fibroblast cells have revealed that less “spread” cells exhibit decreased levels of activated FAK relative to cells which are more “spread” in culture.<sup>227</sup> As FAK has been shown to often be overexpressed and constitutively active in transformed cells,<sup>177, 231</sup> it is reasonable to hypothesize that typical regulation of FAK activation during cell spreading via integrin-ECM interactions shifts towards alternate activation mechanisms as transformed cells display greater anchorage-independent characteristics. This hypothesis is supported by the losses observed in the present investigation in proteins across tumor cell lines that have previously been shown to promote cell spreading in other cell types. Furthermore, the transformed phenotypes maintained by the breast cancer cell lines analyzed have been shown by the literature to be driven largely in part by abnormal growth factor signaling, such as through aberrant estrogen responses in MCF7 cells and constitutively active EGF signaling in SK-BR-3 cells due to HER2 amplification and MDA-MB-231 cells due to overexpression of EGFR or through mutant KRAS, which may thus produce a signaling context that promotes activation of FAK in the absence of anchorage-dependent mechanisms.<sup>169, 171-174, 232</sup> In support of this notion, silencing of FAK expression in breast cancer cell lines, including MDA-MB-231 cells, exhibiting aberrant RAS signaling results in growth arrest of these cell types and in HER2 positive breast cancer cell types, including SK-BR-3 cells, results in diminished growth capacity, underscoring the integral role FAK plays in maintaining the transformed phenotype of breast cancer cells.<sup>232</sup>

Assessment of differentially abundant proteins observed across the breast cell lines analyzed that clustered specifically with FAK (Fig. 10) included several factors that have been shown to

impact expression levels or activation of FAK, namely the membrane metallo-endopeptidase (MME, also known as neprilysin), TNC, KRT19, the ECM proteoglycan versican (VCAN), and canstantin, the 24-kDa region of COL4A2.<sup>190, 204, 235-237</sup> Proteins exhibiting decreased abundances across these breast cancer cell lines that have been shown to inhibit FAK activation include MME, TNC (Fig. 8) as well as canstantin.<sup>204, 235, 236</sup> Ectopic expression of MME in prostate cancer cells has been shown to decrease activation of FAK, with basal MME expression being further shown to be inversely correlated with FAK levels in several prostate cancer cell line models.<sup>236</sup> In breast cancer, MME expression has been shown to be decreased in breast tumor tissues with this loss correlating with a decrease in overall disease-free survival.<sup>238</sup> However, MME expression has also been shown to be increased in breast tumor stroma and this observation has further been shown to correlate with an increase in disease recurrence.<sup>239</sup> TNC has been shown to decrease activation of FAK in glioblastoma cells grown on a matrix consisting of fibronectin and TNC versus fibronectin alone.<sup>63</sup> Further, co-treatment of human umbilical vein endothelial cells with canstantin (the 24-kDa region of COL4A2) and serum resulted in decreased activation of FAK which was induced by serum alone.<sup>235</sup>

Conversely, KRT19 and VCAN have been shown to either decrease expression or increase activation of FAK, respectively, and exhibited abundance profiles across breast cancer cell lines that are contrary to supporting constitutive activation of FAK in these cell lines.<sup>190, 237</sup> Restoration of KRT19 in the KRT19-negative breast cancer cell line BT-549 has been shown to decrease expression of FAK and to promote cell dormancy, which is consistent with the high levels of KRT19 observed in dormant, circulating tumor cells.<sup>190</sup> This group further showed that FAK expression is increased in MDA-MB-231 cells by silencing of KRT19 expression.<sup>190</sup>

Interestingly, KRT19 has recently been shown to be actively released by MCF7, SK-BR-3 and MDA-MB-231 cells and evaluation of KRT19 expression in metastatic tumor cells from the bone marrow of breast cancer patients has revealed high levels of KRT19 to correlate with disease occurrence, progression and an overall decrease in survival.<sup>240</sup> This apparent inverse relationship between KRT19 and FAK expression levels in breast cancer cells may be explained by context-specific roles for these factors, being that KRT19-mediated cell dormancy may be required to survive distal metastasis, whereas the role of FAK in maintaining the tumorigenic potential of breast cancer cells may require a cellular or tissue context.

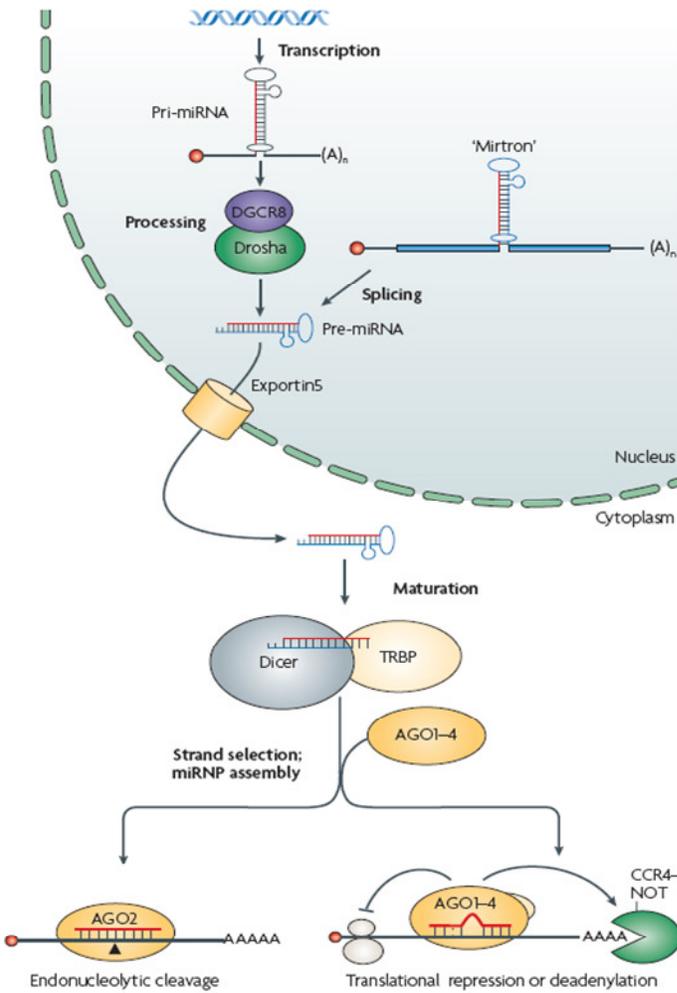
Analysis of integrin beta-1-dependent cell adhesion in U87 glioblastoma cells revealed ectopic expression of a c-terminal domain variant of VCAN results in increased cell adhesion and activation of FAK.<sup>237</sup> Wu and colleagues further showed that MDA-MB-231 cells grown in spent media produced by U87 cells overexpressing this variant of VCAN results in increased cell adhesion as well, but the subsequent impact on FAK activation in these cells was not noted.<sup>237</sup> Versican has been shown to be highly expressed in the stroma of breast cancer patients with this further correlating with a decrease in overall disease-free survival.<sup>241</sup> It is reasonable to hypothesize that high levels of VCAN in breast tumor stroma may drive FAK activation in *in situ* disease. However, this VCAN variant effect was found to be dependent upon interaction with integrin beta-1, which in the present investigation was observed as decreased in abundance across all breast cancer cell lines versus MCF10A cells. These findings indicate a subset of proteins that have been previously shown to regulate FAK expression and activity in other cell types and disease pathologies that therefore warrant further investigation for their roles in regulating FAK activation in breast cancer pathogenesis.

## **Conclusion**

A comparative, global proteomic analysis and subsequent functional systems analysis utilizing IPA software of proteins differentially abundant across three human breast cancer cell line models (MCF7, SK-BR-3 and MDA-MB-231) reflective of the most common clinical disease subtypes encountered relative to a cell line model of normal, human mammary epithelial cells (MCF10A) revealed modulation of ECM, plasma membrane and nuclear protein composition is conserved across breast cancer cells. Systems biology analysis revealed predominant losses in proteins associated with the regulation of cell spreading and modulation of proteins that have previously been shown to impact FAK expression levels and activity in various cell types and disease pathologies, such as MME, TNC, COL4A2, KRT19 and VCAN. These analyses thus provide a subset of candidates that warrant further investigation for their roles in regulating cell spreading and FAK expression and activity in breast cancer. In conclusion, these data provide insights into protein abundance and subsequent functional characteristics that are conserved across three divergent models of breast cancer cells relative to a model of normal mammary epithelial cells, revealing characteristics that represent molecular events central to the pathogenesis of breast cancer.

### **3.0 CHAPTER 2: QUANTITATIVE PROTEOMIC ANALYSIS OF MICRORNA-145 IN BREAST CANCER.**

Recent efforts in the field of RNA-mediated interference has revealed an endogenous gene expression regulatory mechanism that is mediated by small, non-coding RNA segments termed microRNAs (miRNA).<sup>242</sup> MicroRNAs disrupt gene expression post-transcriptionally by binding to discrete, yet potentially multiple target messenger RNA (mRNA) transcripts resulting in their degradation or translational repression (Figure 12). miRNA biogenesis begins with transcription of a primary miRNA transcript (pri-miRNA) by RNA polymerase II that is then cleaved by the nuclear RNase III complex Drosha-DGCR8 or by processing of miRNA-bearing, intronic sequences (Mirtrons) by mRNA splicing machinery to produce ~75 nucleotide (nt) stem-loop structures termed precursor miRNAs (pre-miRNAs). Pre-miRNAs are then exported to the cellular cytoplasm via exportin-5 where they are further cleaved by the RNase III-type Dicer and TAR RNA binding protein (TRBP) enzyme complex into mature miRNAs; ~22 nt double-stranded RNA (dsRNA) containing ~2 nt, 3 prime (3') - end overhangs. These mature miRNAs are then incorporated into a miRNA-induced silencing complex (miRISC) which consists of various ribonucleoproteins, such as Dicer and the Argonaute (AGO) protein family which mediate the catalytic functions of the miRISC.<sup>242-245</sup> This complex is then typically directed to target sites within the 3' untranslated regions (UTR) of mRNAs resulting in the degradation or translational repression of gene targets.<sup>242-245</sup> The extent to which either mRNA degradation or translational repression occurs is dependent on the degree of mismatch between a miRNA and its corresponding mRNA target sequence; as greater mismatch results in translational repression and lesser mismatch, and therefore greater complementarity, in degradation.<sup>243, 246</sup>



**Figure 12: microRNA Biogenesis.** Adapted from Filipowich, Wet et al 2008.

Recent evidence supports the role of miRNAs in oncogenesis, such as in cases where miRNAs target tumor suppressor genes or conversely when the expression of miRNAs responsible for targeting proto-oncogenes are lost.<sup>247</sup> In the context of breast cancer, comparison of miRNA profiles between normal and tumorigenic breast cancer cell lines and tissues has revealed aberrant expression of various miRNAs accompany breast tumorigenesis, including *microRNA-145 (miR-145)* which is commonly lost in transformed breast cancer cells.<sup>248-</sup>

<sup>250</sup> Further, decreased *miR-145*

expression has been correlated with the onset of a pre-neoplastic state in breast myoepithelial cell populations indicating that this loss may represent an early event in breast cancer pathogenesis and thus may play a central role in underlying disease processes.<sup>248</sup> Early *in vitro* investigations of miR-145 in breast cancer revealed expression of this miRNA to be induced by p53 via direct interaction with a putative promoter region upstream of the miR-145 gene sequence and that restoration of miR-145 in MCF7 breast cancer cells resulted in a growth inhibitory phenotype which was shown to be induced by direct targeting of the proto-oncogene c-myc<sup>251</sup> or the Rho-

interacting scaffolding protein Rhotekin (RTKN).<sup>252</sup> Further, restoration of miR-145 expression in divergent cell line models of normal and cancerous breast cells revealed this growth inhibitory phenotype was dependent on either the presence of wild-type p53 or estrogen receptor (ER) expression, eliciting a pro-apoptotic effect in these cell types and further being shown to directly target ER $\alpha$ .<sup>253</sup> Later analyses of miR-145 restoration in breast cancer cell lines, including MDA-MB-231 and SK-BR-3 cells, revealed this miRNA induces decreases in cell motility and invasive capacity as well as decreased tumor burden in a xenograft models of tumor metastasis.<sup>254, 255</sup> Decreased cell invasion and metastatic capability was shown to be due to miR-145 directly targeting the metastasis-associated gene mucin-1, junctional adhesion molecule A (JAM-A) and the actin filament cross-linking protein, fascin.<sup>254, 255</sup>

Currently, 1048 miRNAs have been identified in humans (Sanger miRBase, 10/2010), though very little is known regarding the complete array of mRNA targets that may be regulated by a specific miRNA. While this regulatory event is highly effective, it is rather imprecise due to an imperfect targeting mechanism that is based on the free energy of binding between the first eight nucleotides in the 5' region of a miRNA, which contains the so-called "seed" region, and the 3' UTR of its target mRNA; an interaction which can be confounded by possible G:U wobble base-pairing (see Figure 10: miRNA target discovery workflow Figure).<sup>256</sup> Current methodologies applied in the prediction of miRNA targets utilize computational algorithms (such as TargetScan<sup>257</sup>, MIRANDA<sup>258</sup> or PITA<sup>259</sup>) that score the receptiveness of an mRNA transcript as a target for a specific miRNA based on this 5'- end complementarity and the thermodynamic stability of the interaction between the remaining 3' region of the miRNA and the 3'UTR target sequence.<sup>256, 260</sup> An evaluation of putative miR-145 targets utilizing TargetScan, MIRANDA

(microcosm) and PITA indicates that 1923 ( $\pm$  2039) unique mRNA transcripts are targeted by this miRNA (10/2010), representing an impressive ensemble of targets that have not been experimentally observed. As miRNAs regulate their targets post-transcriptionally, the use of RT-PCR or mRNA microarray profiling to validate miRNA targets provides only limited insight into the global effects of this regulatory event as mRNAs translationally repressed by a miRNA would fail to be detected by these techniques.<sup>256</sup> Therefore, a miRNA target elucidation strategy which incorporates global, quantitative proteomic analysis provides a strategy in which to measure the impact of a miRNA on mRNA translation. The combination of quantitative proteomic and genomic analyses towards miRNA target discovery has confirmed this hypothesis, revealing that discrete populations of mRNAs are degraded and/or translationally repressed in miRNA loss or gain-of-function studies.<sup>77, 78, 261</sup> In the context of breast cancer, the utilization of an iTRAQ-based quantitative proteomic strategy to discover miRNA targets has been reported for miR-21, a pro-oncogenic miRNA, in which miR-21 expression was silenced in MCF-7 cells and proteins increased in abundance were investigated as potential targets of miR-21 via screening of corresponding 3'UTR sequences for miR-21 "seed" regions.<sup>79</sup>

We therefore undertook a miRNA target discovery approach utilizing an LC-MS, quantitative proteomics to elucidate targets of miR-145 in MDA-MB-231 and SK-BR-3 human breast cancer cells. Investigation of miR-145-specific effects was achieved via sequence analysis of 3'UTR's derived from genes corresponding to proteins which were decreased in abundance in MDA-MB-231 cells expressing miR-145 (MDA-145) and revealed 33 target candidates that contained a miR-145 "seed" motif. Differential proteomic analysis of SKB-145 cells revealed 280 target candidates that contained a miR-145 "seed" motif. Correlation of these of miR-145 target

candidate populations revealed 11 mutually overlapping targets between MDA-145 and SKB-145 cells. Furthermore, preliminary analyses of the phenotypic impact of miR-145 restoration on cell survival, cell cycle progression and cell migration in MDA-MB-231 cells are presented. Results revealed miR-145 did not modulate basal cell survival or cell cycle characteristics in MDA-MB-231 cells, ectopic expression however did produce disorganized cell monolayers in confluent cultures as well as modest impairment of cell migratory characteristics relative to MDA-NEG cells. These observed phenotypic effects in MDA-145 cells have been corroborated by other groups<sup>253-255</sup>, which, upon further examination, indicates that the degree of impact on miR-145 on cell migration in MDA-MB-231 cells is concentration dependent.<sup>254</sup>

## **Materials and Methods**

### **Generation of MDA-MB-231 and SKBR3 breast cancer cell lines stably expressing miR-145.**

Pri-miR-145 sequences [native hairpin-loop sequence (pre-miRNA, ~88nt) flanked by  $\pm$  80bp of genomic DNA]<sup>262, 263</sup> were commercially synthesized with 5' *Bam*HI and 3' *Xho*I sites (Integrated DNA Technologies, Coralville, IA) and cloned into the pcDNA6.2-GW/EmGFP-miR expression vector (Invitrogen, Carlsbad, CA) via *Bam*HI and *Xho*I restriction sites. MDA-MB-231 and SK-BR-3 cells (ATCC) cultured in DMEM/F-12 K media supplemented with L-glutamine and 10% fetal bovine serum were transfected with 5 $\mu$ g of the miR-145 containing vector or a negative control variant of pcDNA6.2-GW/EmGFP-miR encoding a scrambled pre-miRNA sequence. Forty-eight hours post-transfection, cells were selected for stable expression in medium containing Blasticidin (MDA-MB-231: 10  $\mu$ g/mL and SK-BR-3: 5 $\mu$ g/ml, Invitrogen)

for up to four weeks, being monitored for EmGFP expression by fluorescence microscopy throughout. Further, silencing of restored miR-145 expression in SK-BR-3 cells was achieved via transient transfection of 100nM of a miR-145-specific antagomiR (Ambion) followed by incubation for 72 hours. miR-145 expression was confirmed by quantitative PCR via purification of small RNA from stable expressing cell populations utilizing the mirVana miRNA Isolation Kit (Ambion, Austin TX) and a TaqMan microRNA quantitative PCR assay (ABI, Carlsbad, CA) specific for mature miR-145 and RNU43 (a small nucleolar RNA (C/D box 43) which functioned as a small RNA loading control) on an ABI 7900HT real-time PCR machine as per manufacturer's recommendations.

### **Sample preparation and liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses**

Cell lysates were prepared from ~70% confluent plates by scraping cells into 150  $\mu$ l of 1X SDS buffer, sonicating utilizing a micro-tip sonicator followed by centrifugation at 14,000 x g for 10 min. Protein concentration of the supernatants was determined by the bicinchoninic assay (BCA). Thirty-five  $\mu$ g of total cell lysate derived from each conditions was resolved via 1D SDS-PAGE on a 4-12% Bis-Tris gel. Gels were stained with Coomassie blue and gel lanes were cut into ten equivalently sized slices. Gel slices were de-stained in 50% acetonitrile (ACN) and 50 mM ammonium bicarbonate (AMB) at ambient temperature for 3 h followed by dehydration in 100% ACN. Reactive cysteine residues were reduced via re-hydration of gel spots in 10 mM DTT, 25 mM AMB followed by incubation at 56 °C for 30 min and alkylated via incubation in 55 mM iodoacetamide, 25 mM AMB for 30 min at ambient temperature in darkness. Gel slices were then washed with 25 mM AMB, dehydrated in 100% ACN and re-hydrated in 25 mM

AMB containing 20 µg/mL porcine sequencing grade modified trypsin on ice for 30 min. Digestions were incubated for 16 h at 37°C. Tryptic peptides were extracted with 70% ACN, 5% formic acid (FA), dried by vacuum centrifugation and re-suspended in 0.1% trifluoroacetic acid (TFA). Peptide digests were resolved by nanoflow reverse-phase liquid chromatography (Ultimate 3000, Dionex Inc., Sunnyvale, CA) coupled online via electrospray ionization to an LTQ-Orbitrap mass spectrometer for MDA-NEG and MDA-145 samples or to an LTQ-XL for SKB-NEG vs. SKB-145 and SKB-145 vs. SKB-145KD sample analyses. Duplicate injections of peptide extracts were resolved on 100 µm i.d. by 360 µm o.d. by 200 mm long fused silica capillary columns slurry-packed in-house with 5 µm, 300 Å pore size C-18 silica-bonded stationary phase. After sample injection, peptides were eluted from the column using a linear gradient of 2% mobile phase B (0.1% FA in ACN) to 40% mobile phase B over 125 min at a constant flow rate of 200 nL/min followed by a column wash consisting of 95% B for an additional 30 min at a constant flow rate of 400 nL/min. The LTQ MS was configured to collect broadband mass spectra [ $m/z$  375-1800; LTQ-Orbitrap at high resolution ( $R=60,000$ )] from which the seven-most abundant peptide molecular ions dynamically determined from the MS scan were selected for tandem MS using a relative CID energy of 30%. Dynamic exclusion was utilized to minimize redundant selection of peptides for CID.

### **Peptide Identification and Spectral Count Analysis**

Peptide identifications were obtained by searching the LC-MS/MS data utilizing SEQUEST (BioWorks, v3.2, ThermoScientific) on a 72-node Beowulf cluster against a UniProt-derived human proteome database (version 10/08, 56,301 protein entries) obtained from the European Bioinformatics Institute (EBI) using the following parameters: trypsin (KR); full enzymatic-

cleavage; two missed cleavages sites; 20 ppm peptide mass tolerance peptide tolerance, 0.5 amu fragment ion tolerance; and variable modifications for methionine oxidation ( $m/z$  15.99492) and cysteine carboxyamidomethylation ( $m/z$  57.02146); data analysis revealed 10.7% ( $\pm 1.3\%$ ) of all cysteine-containing peptides were non-reacted. Resulting peptide identifications were filtered according to specific SEQUEST scoring criteria: delta correlation ( $\Delta C_n$ )  $\geq 0.08$  and charge state dependent cross correlation (Xcorr)  $\geq 1.9$  for  $[M+H]^+$ ,  $\geq 2.2$  for  $[M+2H]^{2+}$ , and  $\geq 3.5$  for  $[M+3H]^{3+}$ . Differential protein abundance was calculated by normalizing total peptide spectral counts observed in miR-145 expressing versus negative control conditions (or in SKB-145KD vs. SKB-145 in this case). A value of 0.5 was added to each spectral count value prior to  $\log_2$  transformation to enable ratio values to be calculated for proteins identified in one cell line, but not another.<sup>181</sup> Proteins which were decreased below  $\log_2$  -0.25-fold change in MDA-145 vs. MDA-NEG and SKB-145 vs. SKB-NEG conditions (or increased  $\log_2$  +0.25 fold in SKB-145KD vs. SKB-145), as per previous analyses which observed miRNA-specific effects on protein abundance as occurring at this fold change<sup>77</sup>, were utilized for downstream analysis.

### **Identification of miR-145 target candidates:**

Uniprot accessions corresponding to proteins decreased in abundance in MDA-145 conditions (or increased in SKB-145KD) were converted to Ensemble gene (ENSG) ID's via biomart.org (Database: ENSEMBL GENES 59 (SANGER UK) Homo Sapiens (GRCh37) and subsequent 3'UTR sequences corresponding to converted ENSG ID's were similarly obtained. This workflow was arrived at empirically and yielded the greatest number of Uniprot accessions which mapped to corresponding 3'UTR sequences. Data sets were correlated utilizing Cytoscape<sup>164</sup> and 3'UTR gene sequences were mined for intact, 6-mer and 7-mer miR-145 target

“seed” motifs, i.e. 1) cuggac 2) acugga 3) aacugg 4) acuggac 5) aacugga 6) aaacugg utilizing a search algorithm constructed in-house. Proteins containing at least one 6-mer miR-145 “seed” motif were considered a putative miR-145 target candidate. miR-145 target candidates were correlated from independent analyses of MDA-145 and SKB-145 conditions to discern mutually conserved miR-145 targets. Targets of interest were manually interrogated for their potential role in breast cancer via assessment of published literature as well as utilizing Ingenuity Pathway Analysis software.

#### **MTT growth assay of stable, miR-145-expressing MDA-MB-231 cells:**

MDA-145 or MDA-NEG cells were plated in triplicate at  $5 \times 10^3$  cells per well in a 96-well plate on day 1. On days 2 thru 5, 10  $\mu$ l of MTT reagent (5 mg/ml 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide, Sigma, St. Louis, MO) was added to each well, plates were incubated for an additional 3 h at 37 °C followed by addition of 100  $\mu$ l of MTT solvent (0.1 N HCl in 2-propanol) and measurement of solubilized formazan crystals was performed at 570 nm (background at 690 nm) utilizing a microplate spectrophotometer (SpectraMAX, Molecular Devices, Sunnyvale, CA).

#### **Cell cycle analysis of stable, miR-145-expressing MDA-MB-231 cells:**

MDA-NEG or MDA-145 cells were plated at  $1 \times 10^5$  cells per well in a 6-well plate on day 1. On days 2 thru 5, cells were trypsinized, washed with 1X PBS, fixed with ice-cold 70% EtOH for 10 min followed by re-suspension in 1X PBS, 0.1 mg/ml propidium iodide and 0.04 mg/ml RNase A.<sup>264</sup> Stained cells were immediately analyzed on a BD LSR II flow cytometer (Franklin Lakes,

NJ) in which 10,000 FSC versus PE (562–588 nm emission) events were acquired. Data was analyzed using WinMDI 2.8 flow cytometry analysis software.

#### **Wound healing assay of stable, miR-145-expressing MDA-MB-231 cells:**

MDA-NEG or MDA-145 cells were plated at  $2.5 \times 10^5$  per well in a 6-well plate on day 1. On day 2, confluent monolayers were “wounded” utilizing a sterile 20  $\mu$ l micropipette tip and wounds were demarcated to facilitate monitoring of identical areas by microscopic imaging. Wound areas were imaged immediately upon wounding and at ~12-24 h intervals (data shown are 16 and 48 hour timepoints).

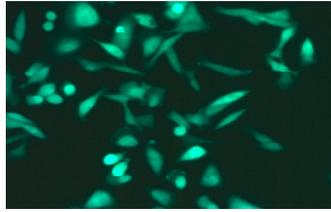
#### **Results and Discussion:**

##### **Identification of miR-145 targets in breast cancer cell lines ectopically expressing miR-145 by differential proteomic analysis.**

The focus of these analyses was towards the identification of regulatory gene targets of the breast cancer-associated microRNA, miR-145, in MDA-MB-231 and SK-BR-3 human breast cancer cells. The workflow utilized to achieve this goal (Fig 13) entailed restoration of miR-145 expression in MDA-MB-231 (MDA-145) and SK-BR-3 (SKB-145) cells followed by differential proteomic analyses of miR-145 expressing and scrambled, negative control miRNA conditions. 3'UTR gene sequences corresponding to proteins decreased in abundance were obtained from ENSEMBL and mined for miR-145 specific target “seed” motifs. Target candidates which overlapped between MDA-145 and SKB-145 analyses were investigated for their roles in breast cancer pathogenesis via assessment of published literature as well as Ingenuity Pathway Analysis

**Restore *miR-145* expression in  
MDA-MB-231 and SK-BR-3 cells**

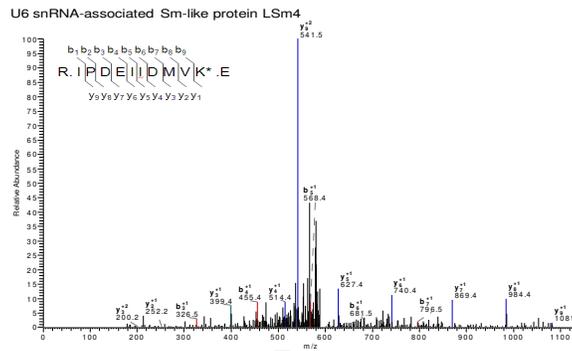
Stable: pri-*miR-145* (IDT)/ pcDNA 6.2-GW EmGFP miR (Invitrogen)



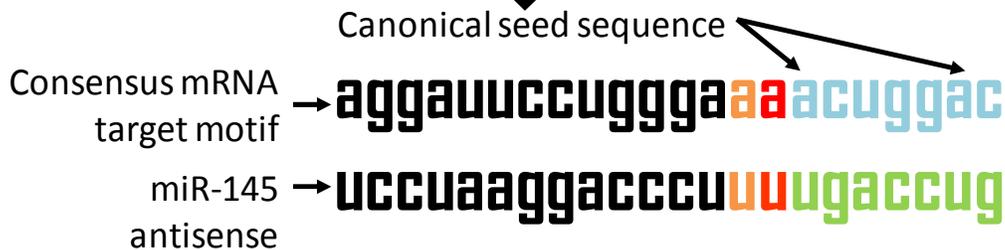
Stable expression  
of pri-*miR-145*



**Liquid Chromatography-Mass Spectrometry (LC-MS)  
Spectral Count Analysis– LTQ-XL & LTQ-Orbitrap XL**

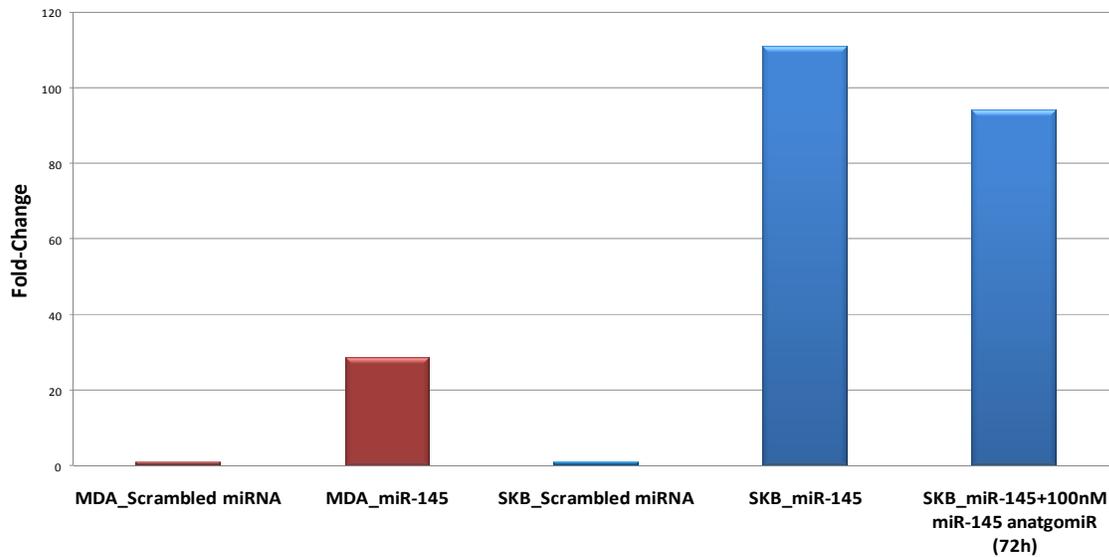


- Determine significantly, differentially expressed proteins:  $\log_2$  0.25-fold
- Obtain 3'UTR sequences corresponding to identified proteins: biomart.org.
- Search 3'UTR sequences for miR-145 motifs: (6mer, 7mers).
- Functional analysis of target candidates: Ingenuity Pathway Analysis.



**Figure 13: miRNA Target Discovery workflow.**

### Restoration of miR-145 in MDA-MB-231 and SK-BR-3 (qPCR, normalized to RNU43)



**Figure 14:** Quantitative PCR analysis of miR-145 expression in MDA-MB-231 and SK-BR-3 cells stably-expressing a vector encoding pri-miR-145 or a scrambled, negative control miRNA. Further, SKB cells stably expressing miR-145 were transfected with 100nM of a miR-145 antagomiR (72h). The small RNA RNU43 functioned as a loading control which miR-145 conditions were normalized against.

software. Further, analysis of the impact of miR-145 restoration in MDA-MB-231 on cell survival, cell cycle regulation and migratory characteristics is also discussed.

#### **Differential proteomic analysis of MDA-MB-231 and SK-BR-3 stably expressing miR-145.**

Confirmation of miR-145 restoration in breast cancer cell lines via qPCR analysis revealed MDA-145 cells exhibited a 28.5-fold increase in miR-145 levels relative to MDA-NEG cells and in SKB-145 cells, a 111.0-fold increase relative to SKB-NEG cells after selection for stable expression (Fig 11). Further, qPCR analysis of SKB-145 following transfection with 100nM of a

<i>Condition</i>	<i>Total Proteins by 2 unique peptides</i>	<i>Total Proteins -0.25 log<sub>2</sub>-fold</i>	<i>% Proteins Decreased</i>
MDA-145 vs. MDA-NEG	1294	313	24%
SKB-145 vs. SKB-NEG	1588	627	39%
SKB-145 vs. SKB-145KD	952	387 (increased)	41% (increased)

<i>Condition</i>	<i>Decreased (Increased for SKB-145KD) Proteins with mapped 3'UTR sequences</i>	<i>Decreased (Increased in SKB-145KD) Proteins with at least ONE 3'UTR, miR-145 "seed"</i>	<i>% Decreased Proteins with miR-145 "seed"</i>
MDA-145 vs. MDA-NEG	272	33	12%
SKB-145 vs. SKB-NEG	571	280	49%
SKB-145 vs. SKB-145KD	304	47	15%

**Table 3: Summary of proteomic analyses and miR-145 specific effects.** Quantitative proteomic analysis of MDA-MB-231 and SK-BR-3 cells stably expressing *miR-145* and the prevalence of proteins bearing a miR-145 “seed” motif in corresponding 3’UTR sequences.

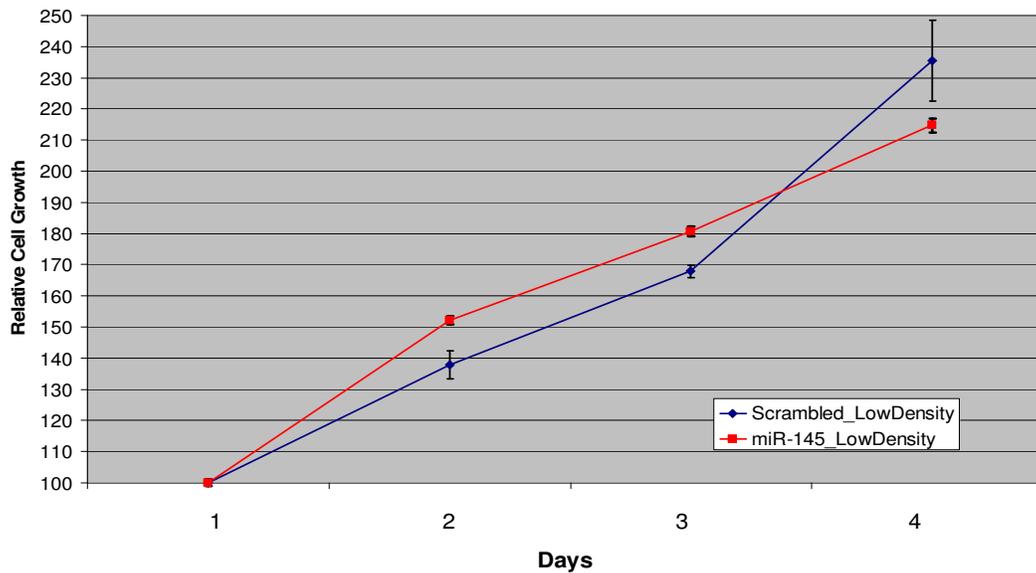
miR-145 specific antagomiR and incubation for 72 h resulted in a 15% decrease in miR-145 expression levels (Figure 14). The results of global proteomic analyses of the following cell line comparisons, MDA-145 vs. MDA-NEG, SKB-145 vs. SKB-NEG and SKB-145 vs. SKB-145KD are summarized in (Table 3). Assessment of total proteins identified by at least two peptides in MDA-145 and SKB-145 proteomic analyses revealed a RSD of 14.4%, whereas total proteins decreased below log<sub>2</sub> -0.25-fold between these two analyses yielded an RSD of 47.2%. This disparity in proteins decreased in abundance between MDA-145 and SKB-145 is commensurate with the markedly increased expression levels of miR-145 achieved in SKB-145, i.e. 3.9-fold

HGNC ID	Protein Name	Fold-Change (STDEV)	Cellular Localization	Functional Type	Predicted?	6mer SEED	7mer SEED
CCT7	T-complex protein 1 subunit eta	-0.57 ( 0.18)	Cytoplasm	other	yes	1	
ILK	Integrin-linked protein kinase	-2.81 ( 0.87)	Plasma Membrane	kinase	no	1	
MRPL54	39S ribosomal protein L54, mitochondrial	-2.32 ( 0.18)	unknown	other	no	1	
NDUFS2	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	-0.89 ( 0.3)	Cytoplasm	enzyme	yes	1	
PCYT2	Ethanolamine-phosphate cytidyltransferase	-2.32 ( 0.52)	Cytoplasm	enzyme	no	1	
PPP6C	Serine/threonine-protein phosphatase 6	-0.29 ( 0.09)	Nucleus	phosphatase	no	1	
RAB5C	Ras-related protein Rab-5C	-0.34 ( 0.02)	Cytoplasm	enzyme	yes	1	
<b>SLC16A3</b>	Monocarboxylate transporter 4	-2.81 ( 1.69)	Plasma Membrane	transporter	yes		1
SLC33A1	Acetyl-coenzyme A transporter 1	-2.32 ( 0.52)	Cytoplasm	transporter	no		1
TXNDC12	Thioredoxin domain-containing protein 12	-0.93 ( 0.05)	Cytoplasm	enzyme	yes	1	
<b>VPS26A</b>	Vacuolar protein sorting-associated protein 26A	-1.08 ( 0.35)	Cytoplasm	transporter	yes	1	

**Table 4: Mutually-observed miR-145 target candidates.** 11 miR-145 target candidates mutually observed between MDA-145 and SKB-145 analyses. Proteins in bold exhibited reverse expression trends in SKB-145 cells in response to transient expression with a miR-145-specific antagomiR.

higher levels in SKB-145 vs. MDA-145 cells. Assessment of miR-145 specific activity by identification of miR-145 “seed” motifs in 3’ UTR sequences corresponding to proteins exhibiting abundances below  $\log_2$  -0.25-fold change revealed 12% of proteins decreased in MDA-145 versus 49% of proteins decreased in SKB-145 analyses contained a miR-145 “seed”. This ~4-fold increase in miR-145 “seed” containing proteins in SKB-145 analyses can be further explained by the high levels of miR-145 expression achieved in SKB-145 cells, further suggesting that the impact of miR-145 on gene regulatory functions may be concentration dependent. As the populations of target candidates discerned from MDA-145 and SKB-145 analyses greatly differed in size, data sets were correlated to discern mutually observed targets, revealing 11 proteins which contained a 3’ UTR miR-145 “seed” motif as being decreased in

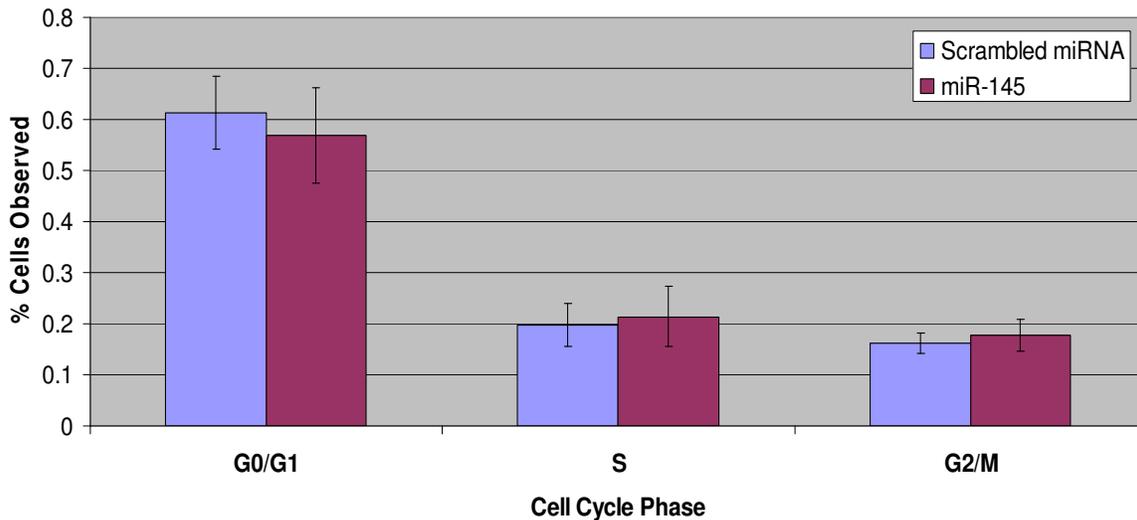
both cell lines analyzed (Table 3). Assessment of the status of these 11 candidates as being previously predicted as putative miR-145 targets was ascertained by correlation with predicted miR-145 targets derived from TargetScan, MIRANDA and PITA prediction algorithms, revealing that 6 of these proteins are predicted miR-145 targets, with the remaining 5 therefore representing potentially novel target candidates (Table 4). Assessment of the functional roles of these 11 target candidates revealed modulation of proteins associated with metabolic and mitochondrial functions, such as 39S ribosomal protein L54, mitochondrial (MRPL54), NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial (NDUFS2) and ethanolamine-phosphate cytidyltransferase (PCYT2) as well as proteins associated with regulation of molecular transport, such as monocarboxylate transporter 4 (SLC16A3), acetyl-coenzyme A transporter 1 (SLC33A1) and vacuolar protein sorting-associated protein 26A (VPS26A) (IPA software). Further analysis revealed several of these factors have been implicated in several cancers, such as chaperonin containing TCP1, subunit 7 (eta) (CCT7), which has been observed as being increased in non-small cell lung cancer tissues derived from smoking vs. non-smoking patients via cDNA analysis<sup>265</sup>, as well as the RAS related protein Rab-5C (RAB5C), which has been observed in lipid rafts of cervical cancer cells<sup>266</sup>, glycolipoprotein-rich regions of the cellular plasma membrane which promote the assembly of signaling molecules, and has been further implicated in the regulation of the pro-oncogenic mitogen-activated protein kinase signaling cascade in these cell types.<sup>267</sup> Additional investigations of these 11 miR-145 target candidates revealed two factors which have previously been implicated in breast cancer pathogenesis, i.e. monocarboxylate transporter 4 (SLC16A3) as well as integrin-linked kinase (ILK).<sup>268-270</sup> SLC16A3 mRNA has been previously observed in breast cancer tissues, being at increased levels in HER2-negative breast cancers specifically, and functional studies have



**Figure 15: Growth curve of MDA-145 vs. MDA-NEG cells.** Triplicate wells were analyzed via MTT analysis over a four-day time-course. Data displayed is normalized to day 1 OD values and is representative of 3 independent replicates.

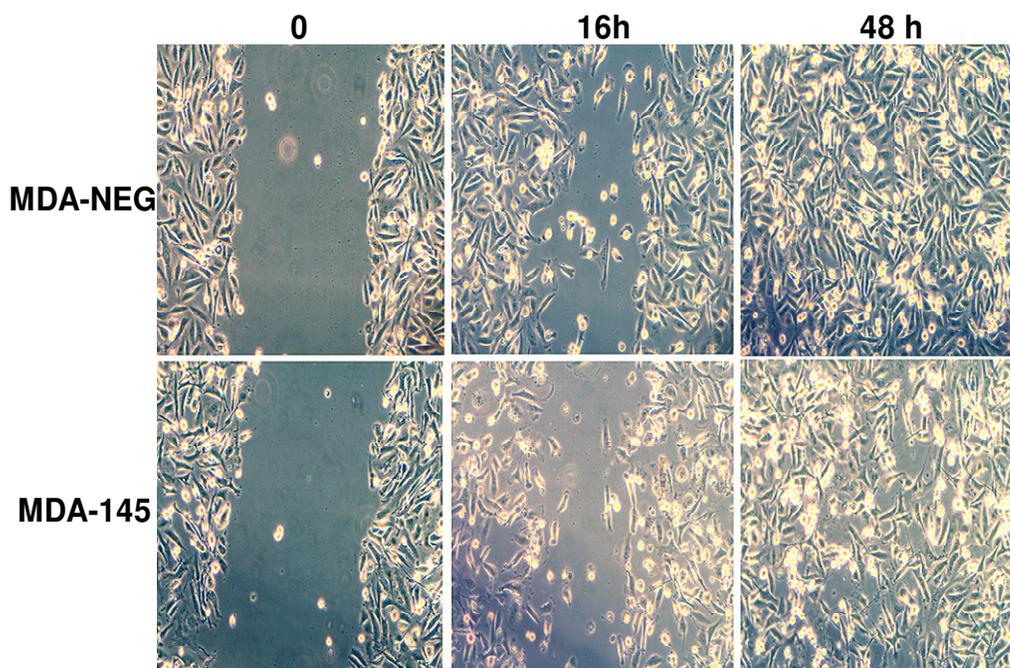
revealed this transporter mediates the cellular efflux of monocarboxylic acids, such as lactate; a by-product of pyruvate metabolism during glycolysis which is the predominant mechanism by which ATP is produced by metastatic tumor cells.<sup>20-22, 269, 271</sup> Analysis of SLC16A3 in MDA-MB-231 cells, a model of metastatic breast cancer, revealed expression levels of SLC16A3 were significantly increased in this cell line relative to normal mammary tissues and further, that silencing of SLC16A3 in MDA-MB-231 resulted in a decrease in cell migration.<sup>269</sup> These findings are interesting when considered in the context of recent evidence indicating the decreases in invasive and migratory activity observed in MDA-MB-231 cells ectopically expressing miR-145.<sup>254, 255</sup> Though other groups have shown that alternate miR-145 targets underlie modulation of cell migration in MDA-MB-231<sup>254, 255</sup>, miR-145-mediated targeting of

SLC16A3 may also be a factor at play in this event. An additional miR-145 target candidate also implicated in breast cancer pathogenesis, integrin-linked kinase (ILK), has been shown to function as an oncogene and further, to be overexpressed in various cancers including breast.<sup>268,</sup>  
<sup>270</sup> ILK is a member of a plasma membrane-localized, multi-protein complex which transduces signaling from cellular integrins and growth factor receptors, such as receptor tyrosine kinases, to the actin cytoskeleton.<sup>270</sup> Activation of ILK occurs in response to direct interaction with integrin or growth factor receptor proteins in a phosphatidylinositol 3-kinase (PI3K)-dependent manner, resulting in phosphorylation of downstream effectors; such as activation of protein kinase-B/AKT and inhibition of glycogen synthase kinase 3-beta (GSK3 $\beta$ ), yielding a cellular phenotype that is resistant to anoikis and which exhibits increased cell survival and proliferation characteristics.<sup>268, 270</sup> Furthermore, ILK has been shown to promote cell migration and invasion of tumor cells, as studies of small molecule inhibitors of ILK have resulted in inhibition of these activities.<sup>270, 272</sup> Thus, further in the context of recently published data revealing decreased migratory activity in miR-145 expressing metastatic breast cancer cells<sup>254, 255</sup>, miR-145-mediated targeting of ILK may also be a factor underlying this event. Additionally, recent evidence revealed miR-145 expressing MDA-MB-231 cells exhibited significant remodeling of the actin cytoskeleton, resulting in reduced actin stress fiber and filopodia formation, supporting the observed decreased in migratory phenotype.<sup>254</sup> ILK-mediated promotion of cell migration has also been shown to be accompanied by actin remodeling and increased formation of actin stress fibers, adding further support to the hypothesis that miR-145-mediated targeting of ILK may contribute to the decreases in cell migration. Interestingly however, the use of ILK-specific small molecule inhibitors or expression of a kinase deficient variant of ILK in MDA-MB-231 has been shown to induce significant anoikis in this cell line.<sup>268</sup> These findings are inconsistent with recent



**Figure 16: Cell cycle analysis of MDA-145 vs. MDA-NEG cells.** Data displayed is representative of sub-confluent cells harvested over a four day time-course. % cells observed for each cell cycle phases were averaged. Data is representative of 3 independent replicates.

evidence indicating that miR-145 expressing MDA-MB-231 cells do not exhibit increases in apoptotic or growth inhibitory responses.<sup>253</sup> This inconsistency may be due to concentration-dependent effects of miR-145 on ILK versus alternate inhibitory strategies.<sup>268</sup> However, it is possible that the panel of genes targeted by miR-145 produces a molecular context which results in inhibition of the migratory activity functions induced by ILK, but which does not impact the influence of ILK signaling on cell growth and survival. Furthermore, previous evidence has shown that the growth inhibitory activity of miR-145 is *wt* p53 or ER $\alpha$ -dependent, two characteristics which MDA-MB-231 and SKBR-3 cells lack.



**Figure 17: Wound healing analysis of MDA-145 vs. MDA-NEG cells.**

Wounded cells were imaged in 12-24 hour intervals. Data is representative of 2 independent replicates.

Confirmatory studies involving global proteomic analyses of SKB-145 cells transfected with a miR-145-specific anatgomiR vs. SKB-145 are summarized in (Table 3). In this context, proteins which were increased above  $\log_2 +0.25$ -fold change in SKB-145KD conditions were characterized for the presence of a miR-145 3'UTR "seed sequence". Assessment of the 11 mutually, differentially abundant miR-145 target candidates observed between MDA-145 and SKB-145 analyses revealed 2 proteins which exhibited a reversed expression profile in SKB-145KD studies, SLC16A3 and VPS26A (Table 4, bolded). These results support initial observations of SLC16A3 and VPS26A as being targets of miR-145, and further that they may

represent candidates which are acutely sensitive to miR-145 levels as SKB-145KD cells exhibited only a 15% decrease in miR-145 levels versus SKB-145.

**Phenotypic analysis of growth, cell cycle and migration characteristics in MDA-MB-231 cells stably expressing miR-145.**

Analysis of the impact of miR-145 on basal growth characteristics of MDA-MB-231 was achieved by comparative analysis of MDA-145 and MDA-NEG cell growth over a four day time-course utilizing an MTT cell proliferation assay (Fig 15). Results revealed that miR-145 had no impact on the basal growth rate of MDA-MB-231 cells. These findings are further corroborated by cell cycle analysis of MDA-145 and MDA-NEG over a four day time-course which also revealed no significant variation in cell cycle profiles between these conditions (Fig 16). These results have been corroborated by another group<sup>253</sup>, with further evidence, as aforementioned, indicating miR-145-mediated growth inhibition as occurring only in breast cancer cell lines expressing *wt* p53 or ER $\alpha$ .<sup>253</sup> Assessment of the impact of miR-145 on migratory activity of MDA-MB-231 was achieved via comparison of MDA-NEG and MDA-145 cells utilizing a wound healing assay over a 48 hour period (Figure 17). Results revealed a modest phenotype indicating impaired migratory ability in MDA-145 cells, with disorganization of cells at the leading edge of the wound at 16 hours and incomplete monolayer formation at 48 hours post-wound initiation. Impaired migratory, as well as invasive and metastatic potential, have also recently been described in MDA-MB-231 cells stably expressing miR-145.<sup>254, 255</sup> Further, the impaired migratory phenotype in MDA-145 cells reported by Gotte et al. was more profound than we had observed and is likely due to the miR-145 expression levels in which they report in MDA-MB-231 versus our model system; which were over 10,000-fold higher than

control cells versus our model in which we achieved only a ~30-fold increase in miR-145 levels. These findings again indicate the regulatory impacts of miR-145 likely occur in a concentration-dependent fashion.

## **Conclusion**

These analyses have yielded a panel of experimentally observed miR-145 target candidates in two cell line models of breast cancer, MDA-MB-231 and SK-BR-3, and further revealed 11 target candidates which were mutually observed between both analyses, suggesting these factors may be conserved regulatory targets. Mutually observed targets reveal modulation of proteins associated with cellular metabolism, mitochondrial function and molecular transport, further revealing two proteins which have previously been associated with breast cancer, SLC16A3 and ILK. The experimental design of this miR-145 target discovery platform was directed towards the identification of intact 3'UTR "seed" sequences for miR-145 to increase the confidence of target candidates observed. However, significant evidence has been shown that these seed regions tolerate mismatches, such as G:U base-pairing<sup>256</sup>, which has been observed in seed regions of validated miR-145 targets, such as c-myc<sup>251</sup> or insulin-receptor substrate-1 (IRS-1)<sup>273</sup>. Further discovery of validated miR-145 targets will provide insights into the possible "seed" and downstream binding characteristics which support effective miR-145 activity, offering evidence which will guide future target discovery efforts. When pairing target discovery and phenotypic data with previously published evidence of miR-145 activity in breast cancer cells, these analyses provide evidence to support that miR-145 activity is concentration-dependent. This observation is vital when considering the possible therapeutic utility of restoring miR-145 expression in breast cancer cells, which has been proposed.<sup>254</sup> In conclusion, these analyses

provide an ensemble of miR-145 regulatory targets in two cell line models of human breast cancer which warrant further investigation for their role in breast cancer pathogenesis. Furthermore, evidence is provided to support previous observations associated with the phenotypic modulation induced by miR-145 in metastatic breast cancer cells as well as the application of global, quantitative proteomic analyses as a strategy to elucidate the impact of microRNA function on the cellular transcriptome and proteome.

#### **4.0 CHAPTER 3: DIFFERENTIAL PROTEOMIC ANALYSIS OF LATE-STAGE AND RECURRENT BREAST CANCER FROM FORMALIN-FIXED PARAFFIN EMBEDDED TISSUES.**

Breast cancers exhibit a high degree of molecular heterogeneity, a characteristic that can confound the accurate determination of disease prognosis, thus more incisive tools are required to define the molecular characteristics underlying neoplastic progression as well as which provide greater insights into risk of disease recurrence.<sup>25, 26</sup> In the context of protein-based prognostic biomarkers for breast cancer, determination of the ER, PR and HER2 status by immunohistochemical (IHC) analysis of patient tumor biopsies remains the gold standard for assessment of disease prognosis and provides the foundation for which patient treatment options are selected.<sup>46, 76</sup> Further, profiling of the urokinase plasminogen activator (uPA)/plasminogen activator inhibitor (PAI-1) proteins by enzyme-linked immunosorbent assay (ELISA) in early-stage breast cancer has been shown to highly correlate with disease prognosis with increased expression of either protein being associated with poor disease outcome and an elevated risk of disease recurrence.<sup>46, 76</sup> Still, the performance of many protein-based biomarkers as prognostic

tools in breast cancer, such as biomarkers of cell proliferation, like Ki-67 (proliferating cell nuclear antigen, PCNA) or cyclin D, have been criticized due to significant inter-patient variability and inconsistent pathological scoring methodologies.<sup>76</sup> Discovery of protein candidates that function to better discern stages of disease progression and provide insights into disease prognosis and recurrence risk would greatly aid in defining the facile molecular tools needed to achieve more confident diagnostic and prognostic assessments.

One resource for the discovery of breast cancer biomarkers are formalin-fixed, paraffin embedded (FFPE) tissues.<sup>80, 123</sup> Gene expression profiling in FFPE tissues has proven fruitful, but are often complicated by the impact of formalin fixation on nucleic acids as this can result in fragmentation and degradation of mRNA products due to the addition of monoethylol moieties to nucleotides and the formation of methylene-bridged, adenine dimers relative to mRNA products recovered from fresh tissue sources.<sup>125, 126</sup> However, multiple investigations have demonstrated high concordance in proteins identified from fresh and FFPE tissues sources indicating that the impact of fixation on MS-based analysis of proteins derived from FFPE tissue as being minimal.<sup>123, 127</sup>

Several comparative, global proteomic analyses of tumor samples derived from FFPE tissues have been performed<sup>129, 131, 274</sup>, though few that have focused specifically on breast cancer.<sup>132</sup> We have undertaken a global proteomic analysis of pathologically-defined tumor regions obtained by LM from archival, FFPE tissues from twenty-five breast cancer patients presenting with stage 0, II and III disease at the time of diagnosis, with a subset of stage II patients exhibiting recurrent (R) disease after initial disease diagnosis. The goal of these analyses was to

identify protein abundance characteristics indicative of early to late-stage disease progression as well as of recurrent disease in breast cancer. Data analysis revealed 113 proteins that significantly differentiated patients diagnosed at early (stage 0, n=7) versus late-stage (stage III, n=9) disease and 42 proteins that significantly differentiated stage II patients that did (stage II R, n=5) or did not (stage II NR, n=4) exhibit recurrent disease. Verification of differentially abundant proteins was achieved via immunohistochemical (IHC) analyses of primary FFPE tissue samples for protein candidates from each stage comparison, i.e. thrombospondin-1 (TSP-1) for stage 0 versus stage III disease and Protein DJ-1 (PARK7) for stage II NR versus stage II R. These data reveal differentially abundant proteins indicative of neoplastic progression and disease recurrence that provide insights into the molecular characteristics underlying breast cancer pathogenesis.

## **Materials and Methods**

### **Breast cancer tissue sample preparation**

Twenty-five FFPE breast cancer patient tissue samples (Table 5) were sectioned onto LM slides (Director Slides, Expression Pathology Inc., Gaithersburg, MD) for tissue microdissection (LMD6000, Leica Microsystems, Bannockburn, IL). Disease recurrence in patients presenting with stage II disease was determined two years following initial diagnosis. Pathologically-defined regions of cancerous epithelium were procured from all twenty-five patient samples and tryptic digests were prepared for global proteomic profiling utilizing the Liquid Tissue FFPE proteomic sample preparation kit (Expression Pathology, Inc., Gaithersburg, MD) as per the

manufacturer's recommendations. Samples were re-suspended at a final concentration of 0.2 ng/ $\mu$ L in 0.1% trifluoroacetic acid (TFA).

### **Liquid chromatography-tandem mass spectrometry**

Total peptide digests were resolved by nanoflow reverse-phase liquid chromatography (Ultimate 3000, Dionex Inc., Sunnyvale, CA) coupled online via electrospray ionization to a linear ion trap mass spectrometer (LTQ, ThermoFisher Scientific, Inc., San Jose, CA). Triplicate injections of 1.2  $\mu$ g of peptides derived from each patient tissue sample were resolved on a 100  $\mu$ m i.d. by 360  $\mu$ m o.d. by 200 mm long fused silica capillary columns (Polymicro Technologies, Phoenix, AZ) slurry-packed in-house with 5  $\mu$ m, 300 Å pore size C-18 silica-bonded stationary phase (Jupiter, Phenomenex, Torrance, CA). After sample injection, peptides were eluted from the column using a linear gradient of 2% mobile phase B (0.1% formic acid (FA) in acetonitrile to 40% mobile phase B over 125 min at a constant flow rate of 250 nL/min followed by a column wash consisting of 95% B for an additional 20 min at a constant flow rate of 400 nL/min. The LTQ MS was configured to collect broadband mass spectra ( $m/z$  375-1800) from which the seven-most abundant peptide molecular ions dynamically determined from the MS scan were selected for tandem MS using a relative CID energy of 35%. Dynamic exclusion was utilized to minimize redundant selection of peptides for CID.

### **MS data analysis**

Peptide identifications were obtained by searching the LC-MS/MS data utilizing SEQUEST (BioWorks, v3.2, ThermoScientific) on a 72-node Beowulf cluster against a UniProt-derived human proteome database (version 10/08, 56,301 protein entries) obtained from the European

Bioinformatics Institute (EBI) using the following parameters: trypsin (KR); full enzymatic-cleavage; two missed cleavages sites; 1.5 Da peptide mass tolerance peptide tolerance, 0.5 Da fragment ion tolerance and variable modifications for methionine oxidation ( $m/z$  15.99492). Resulting peptide identifications were filtered according to specific SEQUEST scoring criteria: delta correlation ( $\Delta C_n$ )  $\geq 0.08$  and charge state dependent cross correlation (Xcorr)  $\geq 1.9$  for  $[M+H]^+$ ,  $\geq 2.2$  for  $[M+2H]^{2+}$ , and  $\geq 3.5$  for  $[M+3H]^{3+}$ . These criteria resulted in a false discovery rate (FDR) of 5.84% for all peptides identified as determined by searching the entire data set against a decoy human database where the protein sequences were reversed.<sup>178</sup> Protein abundance was derived by spectral counting (SC) and peptides whose sequences mapped to multiple protein isoforms were grouped as per the principle of parsimony.<sup>179</sup> To determine statistically significant changes in protein abundance across patient samples by disease stage subgroups, a hierarchical supervised cluster analysis of peptides identified from stage 0 versus stage III and stage II NR disease recurrence (Stage II NR) versus stage II with disease recurrence (Stage II R) patient samples was performed in which the variance in total spectral count peptides identified was determined utilizing the Mann–Whitney rank-sum test (significance level  $p \leq 0.05$ , Fisher’s exact test) paired with the filter criteria requiring that 60% of the samples in a supervised group had a minimum peptide count of 2 or greater for a given protein.

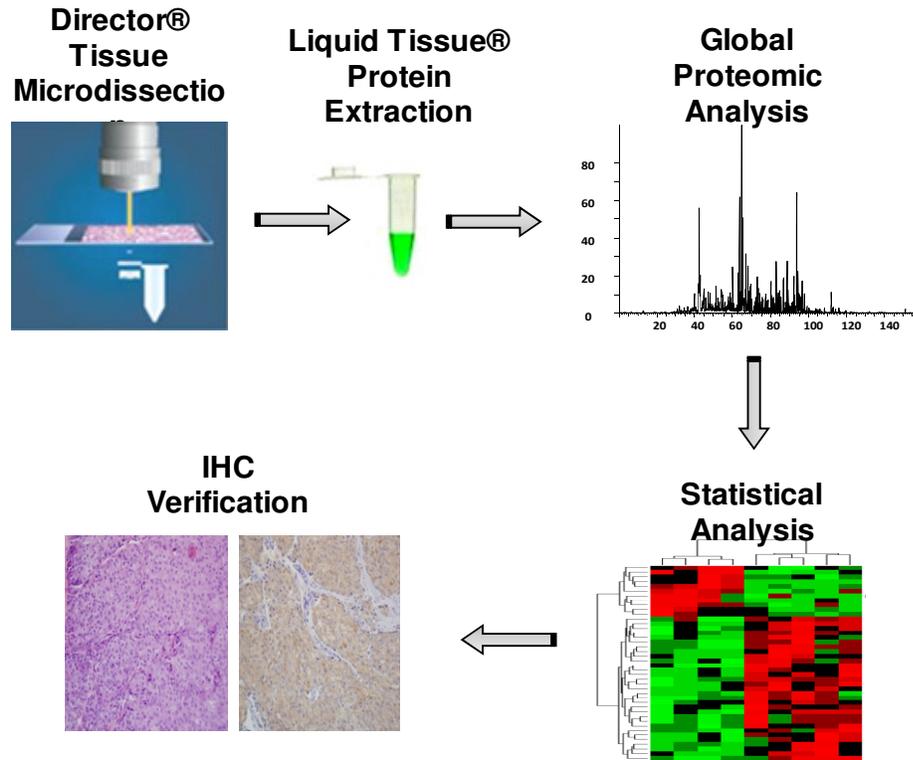
### **Bioinformatic analyses**

Uniprot accessions corresponding to significantly, differentially abundant proteins were mapped to HUGO (HGNC) gene symbols utilizing Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems, [www.ingenuity.com](http://www.ingenuity.com)). Accessions that failed to map were converted to HGNC identifiers by manual inspection at [www.uniprot.org](http://www.uniprot.org) and re-mapped to IPA to maximize protein

identifications available for downstream bioinformatic analyses; a unique HGNC designation for accession # P84243 from comparative stage II analysis could not be determined and was thus not utilized in downstream functional analysis. Functional analysis of significant protein lists were performed utilizing the “Core Analysis” function in IPA and inferred biofunctions exhibiting a  $p < 0.05$  and a minimum of two associated proteins were considered significant.

### **Immunohistochemical Verification**

Immunohistochemical verification (MDR Global Systems LLC, Windber, PA) was performed on primary FFPE tissue sections and antigen recovery was achieved utilizing the Dako PT Module (Dako, Denmark). Immunostaining was performed on a Dako autostainer and all slides were counterstained with hematoxylin and blued with 1% ammonia water. Antibody conditions were as follow: 1: estrogen receptor (ER alpha, ERA), (clone 1D5/mouse, 1:200, Dako), 2: progesterone receptor (PGR), (clone PGR636/mouse, 1:400, Dako) 3: Her2/Herceptest, (polyclonal/rabbit, FDA guidelines, Dako) 4: Thrombospondin-1 (TSP-1), (1:50, clone A6.1/mouse, Abcam) and 5: Park7/DJ1, (1:800, polyclonal/goat, Abcam).



**Figure 18: Significant, differentially abundant proteins identified in comparisons of stage 0 and stage III breast cancer patient tissue samples.** Hierarchical supervised cluster analysis of stage 0 and stage III patient tissue data sets yielded 113 proteins which significantly differentiated these groups.

**Table 5: Vital statistics of 25 patients presenting with discrete stages of disease.**

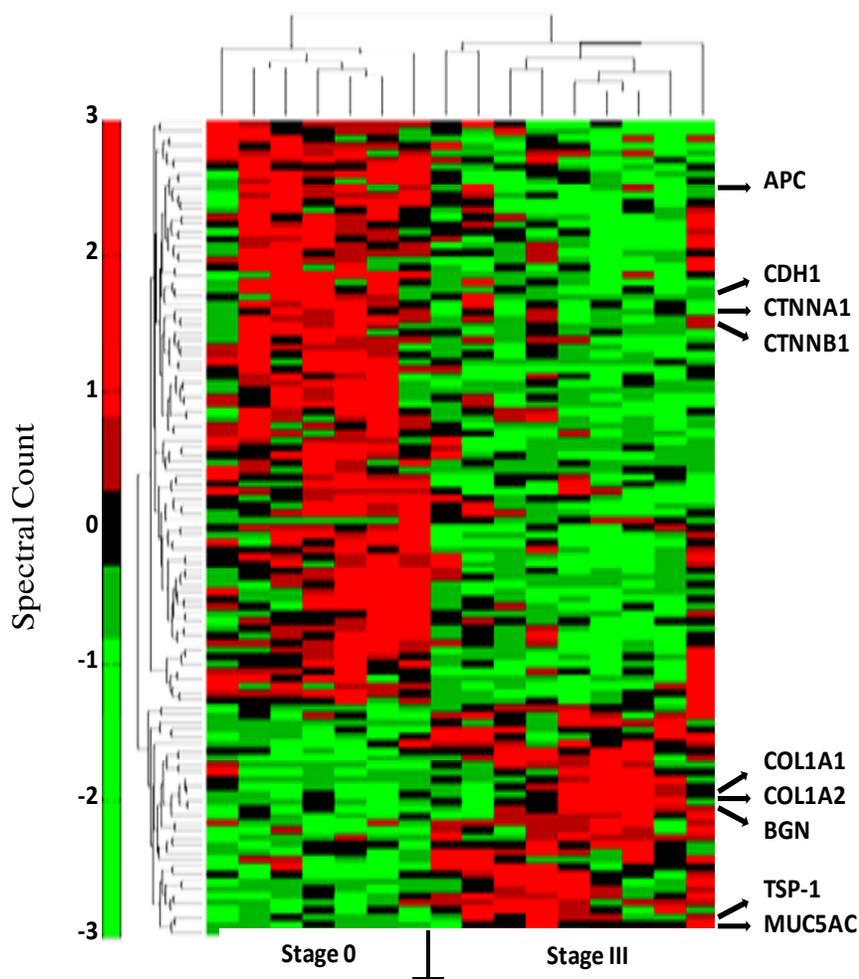
<i>Diagnosis</i>	<i>Patients (n)</i>	<i>Age</i>	<i>DCIS</i>	<i>LCS</i>	<i>ILC</i>	<i>IDC</i>	<i>Comb/Other</i>	<i>ER+</i>	<i>HER2+</i>	<i>Basal-like</i>
Stage 0	7	58(±11)	4	1	0	1	1	6	1	0
Stage II NR	4	58(±12)	0	0	0	1	3	0	1	3
Stage II R	5	56(±21)	1	0	0	2	2	0	2	3
Stage III	9	62(±11)	0	0	4	4	1	6	0	3

## **Results and Discussion**

Global proteomic analyses of pathologically-defined regions of cancerous epithelium, microdissected from FFPE-tissue samples derived from twenty-five breast cancer patients presenting with distinct stages of disease (Table 5) yielded 115,549 total peptide identifications (FDR of 5.84%) and 9,437 proteins identified by at least two peptides (Figure 18). Merits of analytical performance were monitored by comparing peptides identified between replicate injections for all 25 patient samples, revealing an average relative standard deviation (RSD) of 9.5% ( $\pm$  3.9%). Equivalency of protein digest input for LC-MS/MS analyses was determined by comparison of spectral count values for total peptides identified across all twenty-five patient samples where an RSD of 13.4 % was found, as well as for total peptides identified that corresponded to glyceraldehyde 3-phosphate dehydrogenase (GAPDH), a protein commonly utilized as a loading control for western blot analysis, revealing an RSD of 23.0% for this protein.<sup>180</sup> These measures critically underpin the determination of significant, differentially abundant proteins by spectral counting without the need for normalization.

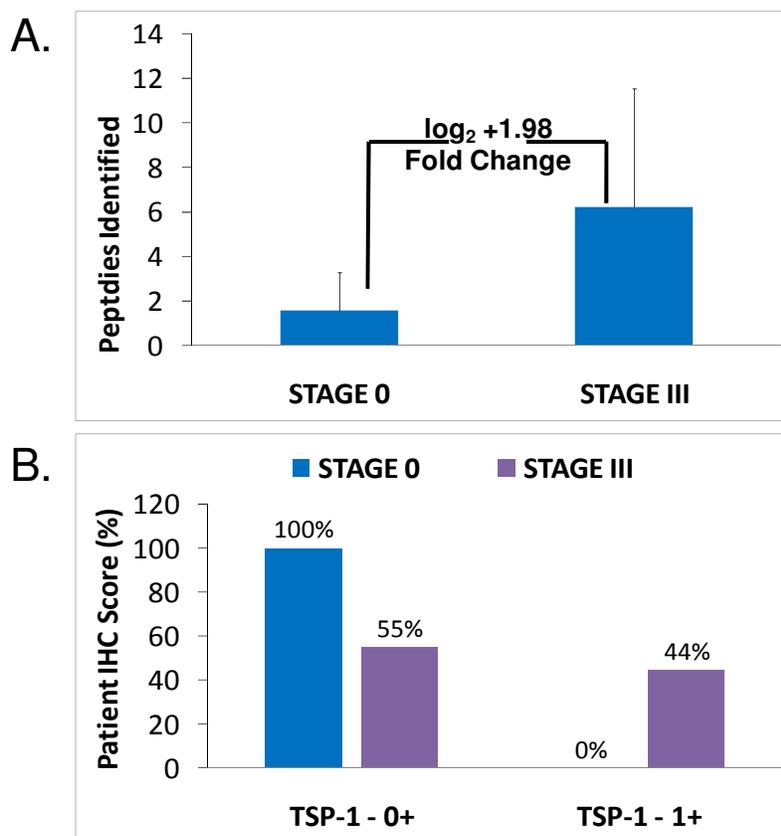
### **Characterization of differentially abundant proteins between stage 0 and III breast cancer.**

Significant, differentially abundant proteins were established by hierarchical supervised cluster analysis in which at least 60% of the samples within a supervised group exhibited spectral count values of 2 or greater for a given protein of interest (Figures 19 and 23). Characterization of differential protein abundance indicative of early versus late-stage breast cancer was achieved by comparative analysis of proteomic data sets derived from stage 0 (n=7) and stage 3 (n=9) breast



**Figure 19: Significant, differentially abundant proteins identified in comparisons of stage 0 and stage III breast cancer patient tissue samples.** Hierarchical supervised cluster analysis of stage 0 and stage III patient tissue data sets yielded 113 proteins which significantly differentiated these groups.

cancer patient tissue samples (Figure 19). Results revealed a total of 113 proteins, 32 that were increased and 81 decreased in abundance in stage III vs. 0 disease, which significantly differentiated patients presenting with these disease stages. Stage III breast cancer is defined by

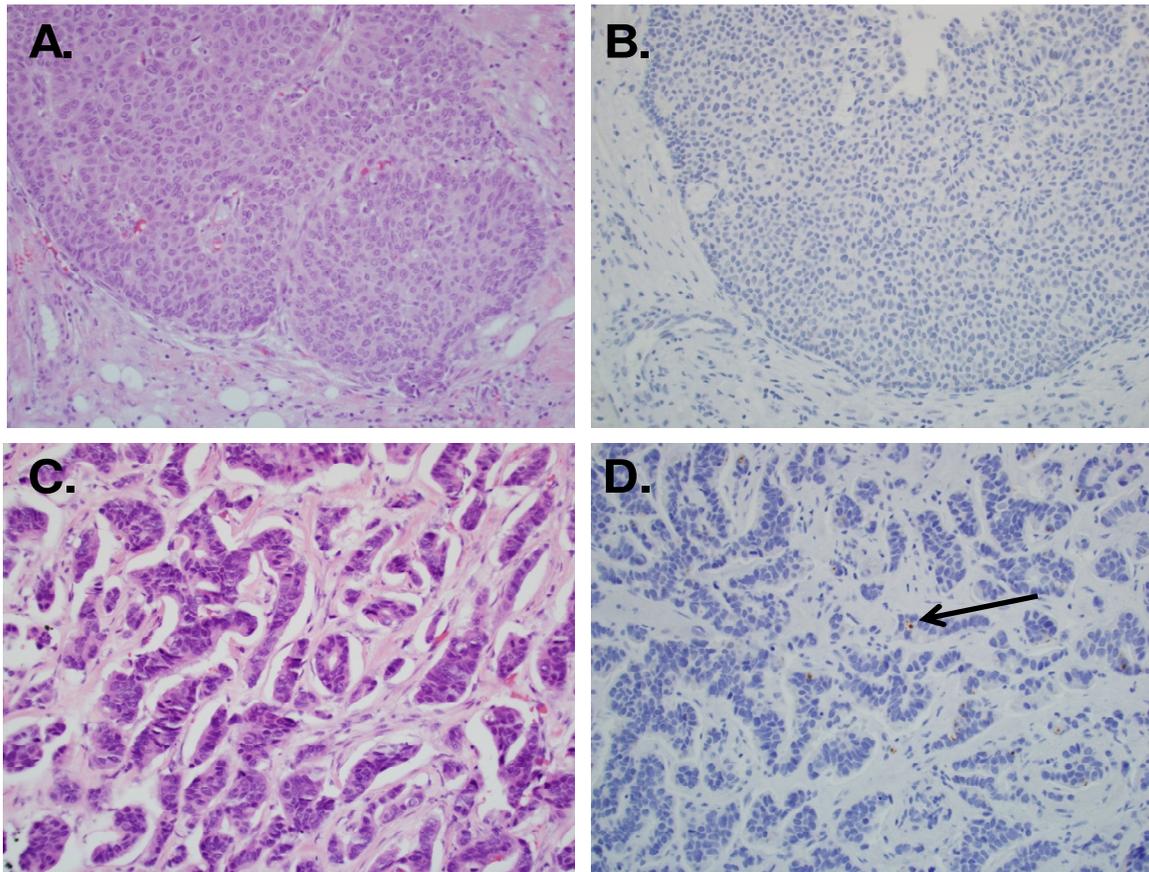


**Figure 20: Differential abundance profile for Thrombospondin-1 (TSP-1) by spectral count peptides (A) and immunohistochemical verification (B) from Stage 0 and Stage III breast cancer patient tissue samples. A.** Data reported indicates average peptides identified for TSP-1 by spectral count between Stage 0 (n=7) and Stage III (n=9) samples. Results revealed a  $\log_2 +1.98$ -fold increased in TSP-1 in stage III versus stage 0 patients. **B.** Data reported indicates the percentage of patient tissues exhibiting an IHC score of 0+ or 1+ for TSP-1 in Stage 0 (n=6) and Stage III (n=9) samples. Results revealed 44% of stage III versus 0% of stage 0 patients exhibited an IHC score of 1+ for TSP-1.

evidence of tumor cell populations that have invaded from the primary tumor site and spread into regional breast lymph nodes, or by a primary tumor that has extended to the chest wall or skin.<sup>44</sup>

<sup>275, 276</sup> Locoregional invasion of tumor cells occurs as a product of phenotypic alterations, such

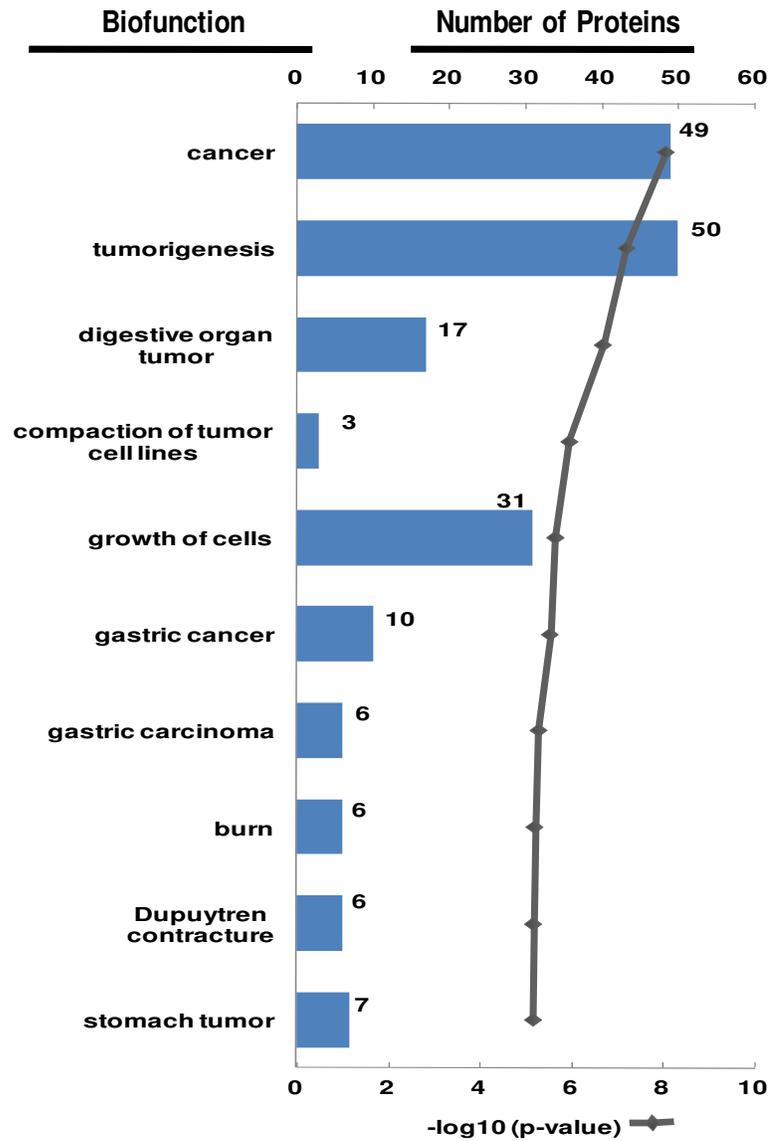
as a loss of contact-inhibited growth and modulation of extracellular matrix (ECM) protein composition, which can be precipitated by selective pressures induced by the tumor microenvironment that accompany accelerated tumor growth, such as increased hypoxic and acidic conditions.<sup>9-11, 16, 18</sup> Manual inspection of the 113 differentially abundant proteins observed between stage 0 and III patients revealed modulation of several extracellular matrix proteins in stage III disease, such as multiple collagen isoforms as well as biglycan (BGN, log<sub>2</sub> +2.61-fold change in stage III versus stage 0 disease), mucin 5AC (MUC5AC, log<sub>2</sub> +1.32) and thrombospondin-1 (TSP1, log<sub>2</sub> +1.98) (Figure 19 and Figure 20A and 20B – TSP-1 data). Among the differentially abundant collagen isoforms observed, the type I collagens, COL1A1 (log<sub>2</sub> +0.94) and COL1A2 (log<sub>2</sub> +1.19), were found to be increased in abundance in stage III patients and corresponding mRNA for these factors has previously been shown to be overexpressed in invasive breast carcinoma.<sup>277</sup> Increased deposition and aberrant cross-linking of collagen is associated with the development of invasive breast cancer, the result of which contributes to “stiffening” of the ECM and is a factor that has been shown to drive progression of *in situ* disease.<sup>278, 279</sup> The ECM-localized proteoglycan BGN was observed as being increased in stage III patients and previous evidence has revealed BGN mRNA levels are increased with breast cancer disease progression, as evidenced by comparisons of patient-matched IDC versus DCIS breast cancer tissues.<sup>280</sup> The ECM glycoprotein MUC5AC was also observed as being increased in stage III disease, and though this mucin isoform has been previously shown to be increased in cancerous versus normal breast tissues<sup>281</sup> a lack of correlation between MUC5AC expression and other clinicopathological parameters utilized in breast cancer diagnostics, such as disease grade, revealed this factor to have weak prognostic value.<sup>282</sup> Further, the ECM glycoprotein TSP-1 was observed as being increased in stage III patients and previous evidence



**Figure 21: Representative immunohistochemical (IHC) analysis of Thrombospondin-1 (TSP-1) in stage 0 (A & B) versus stage III (C & D) patient tissues.** A: Hematoxylin and eosin staining of stage 0 patient tissue. B: IHC staining for TSP-1 in corresponding stage 0 patient tissue. C: Hematoxylin and eosin staining of stage III patient tissue. D: IHC staining for TSP-1 in corresponding stage III patient tissue. Observed vacuolar staining pattern for TSP-1 denoted by arrow.

has indicated that TSP-1 levels are increased in breast cancer tissues as well as found at high levels in malignant breast secretions<sup>283-285</sup> However, clinical assessment of TSP-1 expression in specific histological subtypes and stages of breast cancer disease progression have been ambiguous.<sup>286, 287</sup> Increased TSP-1 expression is predominantly observed in breast tumor stroma

and early studies of this phenomenon comparing invasive ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS) breast cancer subtypes revealed TSP-1 levels were significantly higher in stromal tissues adjacent to invasive disease as well as in both tumor and stromal cell types in malignant versus normal breast tissues.<sup>287</sup> Recent evidence characterizing TSP-1 expression in DCIS has shown that decreased TSP-1 levels in adjacent tumor stroma is associated with more aggressive disease characteristics, such as evidence of high histological grade and increased tissue necrosis, correlating with a poorer disease prognosis in these individuals.<sup>286</sup> TSP-1 is an inhibitor of angiogenesis and exogenous expression of TSP-1 has been shown to elicit both anti-tumorigenic and anti-angiogenic effects in breast cancer cell lines.<sup>283, 285</sup> Assessment of TSP-1 expression associated with microvessel density, a measure of angiogenesis, in breast cancer tissues revealed both positive and negative associations of TSP-1 with these features, findings that complicate defining a central role for TSP-1 in angiogenesis.<sup>283</sup> We confirmed the increased abundance of TSP-1 in stage III patients observed by differential proteomic analysis (Figure 20A) by IHC of TSP-1 in stage 0 (n=6) and stage III (n=9) patient tissue samples (Figure 20B and Figure 21) which revealed a complete lack of TSP-1 staining in stage 0 tissues relative to stage III tissues of which 44% exhibited an IHC score of 1+ (Figure 20B). Further, TSP-1 staining in stage III tissues was observed as being predominantly localized to cellular vacuoles or intra-cytoplasmic lumina of tumor cell populations (Figure 21). Interestingly, IHC analysis of TSP-1 abundance in stage II patient tissues also revealed increases of TSP-1 in stage II R patients versus stage II NR, though this profile was not significantly observed in proteomic analyses. Our results therefore substantiate previous reports indicating that TSP-1 expression is increased with invasive breast cancer and provide evidence to support this abundance profile that indicate a higher risk for developing recurrent disease.



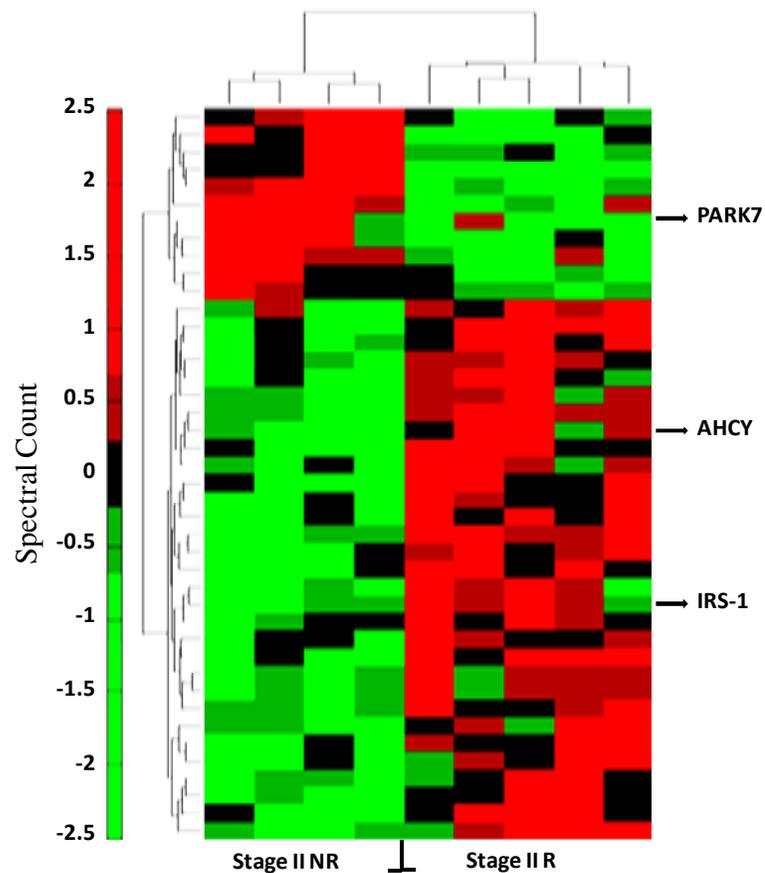
**Figure 22: Significant, differentially abundant proteins identified in comparisons of stage 0 and stage III breast cancer patient tissue samples. Top 10 significant biofunctions observed in proteins differentiating stage 0 and stage III patients. Data reported indicates significant p-values and mean number of proteins observed which clustered with a specific biofunction.**

## **Analysis of functional biological characteristics amongst differentially abundant proteins between stage 0 and III breast cancer.**

Functional analysis of the 113 proteins differentiating stage 0 and stage III patients revealed the top ten significant biofunctions to be predominantly related to the overarching theme of cancer, specifically being associated with the regulation of tumorigenesis as well as being previously implicated in gastrointestinal cancer pathogenesis (Figure 2B). One significant biofunction category - namely "compaction of tumor cell lines" - was of interest as this cellular phenomenon is associated with the maintenance of intercellular interactions between tumor cells, an event that is disrupted in invasive and metastatic breast tumor cells.<sup>11, 288</sup> This biofunction category contained three proteins, E-cadherin (CDH1,  $\log_2$  -1.57 in stage III versus stage 0), alpha-catenin (CTNNA1,  $\log_2$  -1.24) and beta-catenin (CTNNB1,  $\log_2$  -2.56), all of which are core constituents of cellular adherens junctions that comprise plasma membrane-localized, multi-protein complexes which mediate intercellular adhesion and communication via direct interaction with the cellular actin cytoskeleton.<sup>11, 284</sup> The formation of intercellular adherens junctions, which is facilitated through interaction of adjoining cadherin proteins, produces a contact inhibition signal that halts cell proliferation, the response to which is lost during cellular transformation.<sup>11</sup> Modulation of expression of CDH1, CTNNA1 and CTNNB1 are canonical characteristics of transformed mammary epithelial cells, with loss of CDH1 specifically being a hallmark of invasive and metastatic breast tumor cells, such as those that have undergone epithelial to mesenchymal transition, with this decrease further being found to correlate with a poor disease prognosis.<sup>11, 284, 289</sup> Loss of both CTNNA1 and CTNNB1 have also been reported in breast carcinomas, particularly in invasive disease variants or as precursors to these subtypes.<sup>284, 290, 291</sup> Interestingly, loss of expression of CTNNB1 with disease progression in particular is

inconsistent with the oncogenic role CTNNB1 plays in the pathogenesis of a variety of cancers including breast.<sup>288</sup> The canonical role of CTNNB1 is to function as a signaling intermediate between CDH1 and cellular actin at cellular adherens junctions.<sup>288</sup> However, CTNNB1 can further function as a transcriptional co-factor, mediating translation of upstream signaling by the Wnt/Frizzled G-protein coupled receptor family via translocation to the nucleus and interaction with lymphoid enhancer factor/t-cell factor (LEF/TCF) transcription factors resulting in transcriptional activation of LEF/TCF gene targets, such as cyclin D1.<sup>288</sup> Insights into the role of CTNNB1 dysregulation in cancer have emerged from studies of familial adenomatous polyposis, a heritable form of colorectal cancer that often occurs due to mutations in the tumor suppressor gene adenomatous polyposis coli (APC), a protein directly implicated in regulating CTNNB1 stability.<sup>292, 293</sup> APC functions as a cytoplasmic scaffolding protein that facilitates the targeting of CTNNB1 for ubiquitin-mediated proteolysis.<sup>293</sup> This is achieved through APC recruiting glycogen synthase kinase 3 beta (GSK3 $\beta$ ) that directly phosphorylates cytoplasmic CTNNB1, targeting it for degradation.<sup>293</sup> Loss of APC results in increased stability of CTNNB1 and increased CTNNB1-dependent transcriptional activity that drives cellular transformation.<sup>293</sup> Notably, we observed APC (log<sub>2</sub> -1.1) abundance as being decreased in stage III versus stage 0 disease. When pairing the observed decrease in CTNNB1 and APC abundance in stage III disease, a context is produced in which the diminished population of CTNNB1 may possess increased stability and thus function to further drive disease progression. Indeed, modulation of APC via mutation or via epigenetic mechanisms has been observed in invasive breast cancers.<sup>292-</sup>

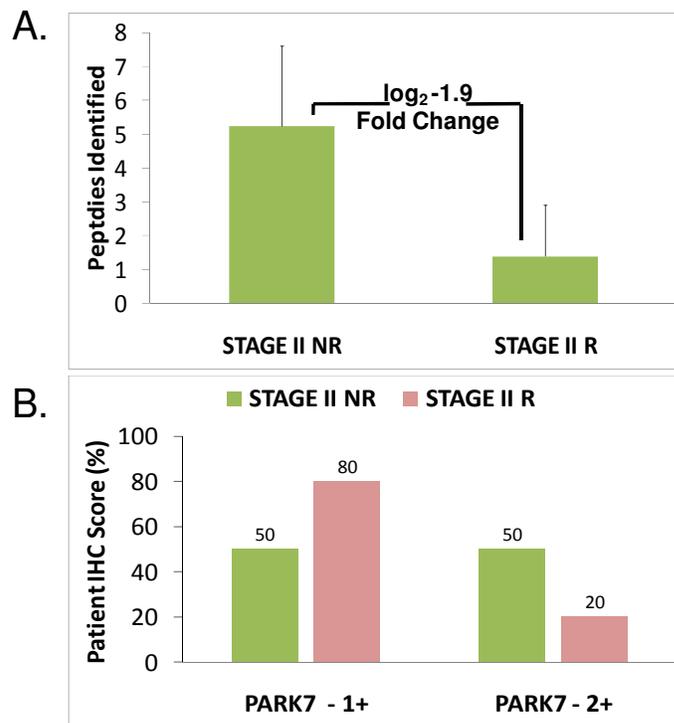
294



**Figure 23: Significant, differentially abundant proteins and subsequent functional characteristics identified in comparisons of stage II NR and stage II R breast cancer patient tissue samples.** Hierarchical Supervised cluster analysis of stage II NR and stage II R patient tissue data sets yielded 42 proteins which significantly differentiated these groups.

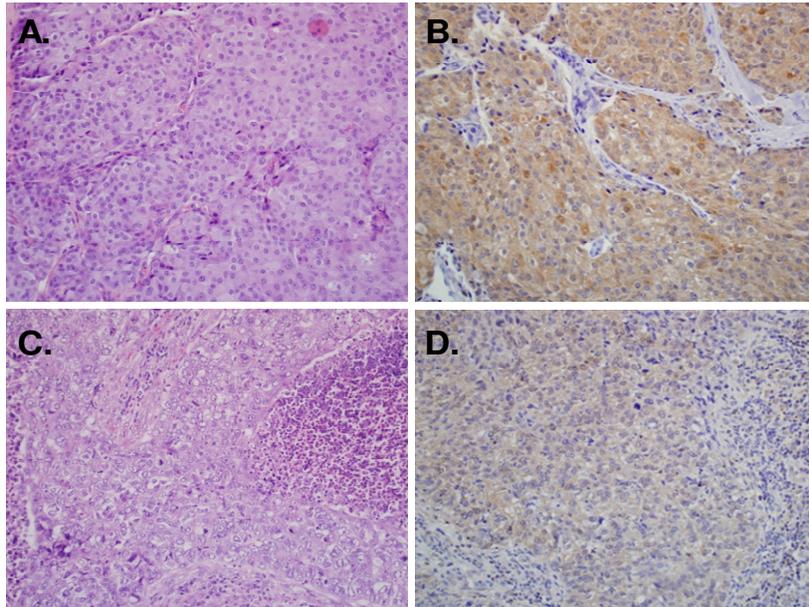
**Characterization of differentially abundant proteins between stage II non-recurrent and recurrent breast cancer.**

Characterization of differential protein abundance in patients diagnosed with stage II disease that did (stage II R) or did not exhibit recurrent disease (stage II NR) two years following initial diagnosis was achieved by comparative analysis of proteomic data sets derived from stage II NR



**Figure 24: Differential abundance profile for Protein DJ-1 (PARK7) by spectral count peptides (A) and immunohistochemical verification (B) from Stage II NR and Stage II R breast cancer patient tissue samples.** A. Data reported indicates average peptides identified for PARK7 by spectral count between Stage II NR (n=4) and Stage II R (n=5) samples. Results revealed a  $\log_2$  - 1.9-fold decrease in PARK7 in stage II R versus stage II NR patients. B. Data reported indicates the percentage of patient tissues exhibiting an IHC score of 1+ or 2+ for PARK7 in Stage II NR (n=4) and Stage II R (n=5) samples. Results revealed 20% of stage II R versus 50% of stage II NR patients exhibited an IHC score of 2+ for PARK7.

(n=4) and stage II R (n=5) patient tissue samples (Figure 23). Results revealed a total of 42 proteins, 31 that were increased and 11 decreased in abundance, which significantly differentiated stage II NR versus II R patients (Figure 23). Stage II breast cancer is defined as



**Figure 25: Representative immunohistochemical (IHC) analysis of Protein DJ-1 (PARK7) in stage II NR (A & B) versus stage II R (C & D) patient tissues.** A: Hematoxylin and eosin staining of stage II NR patient tissue. B: IHC staining for PARK7 in corresponding stage II NR patient tissue. C: Hematoxylin and eosin staining of stage II R patient tissue. D: IHC staining for PARK7 in corresponding stage II R patient tissue.

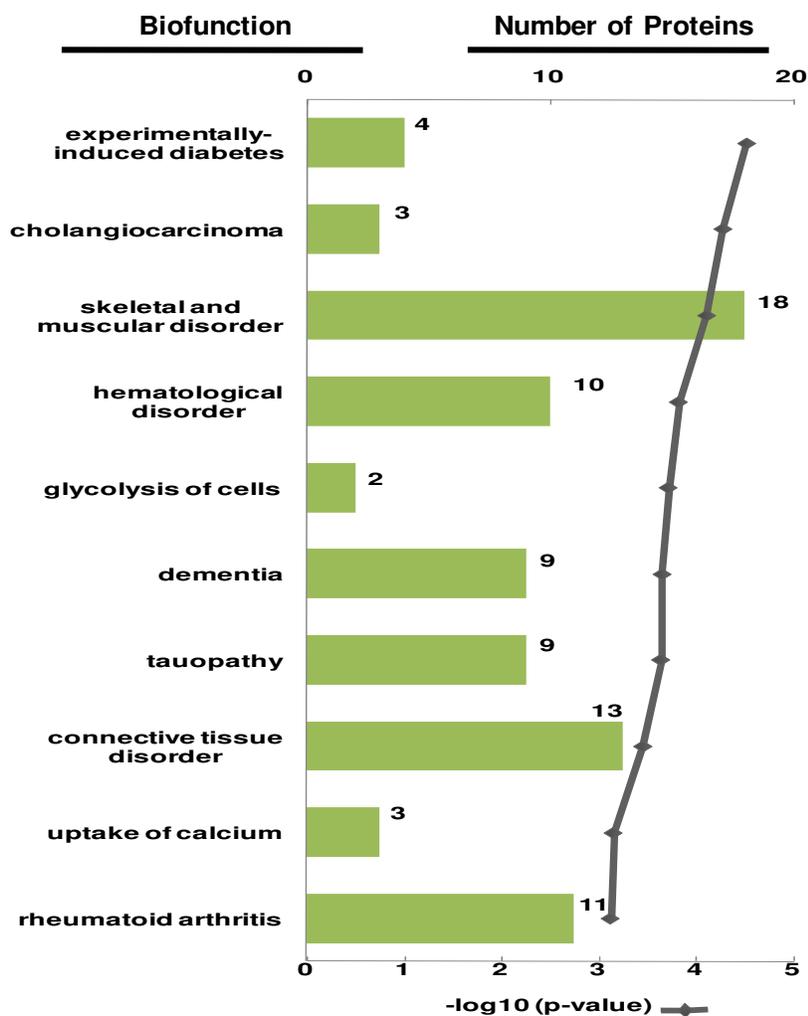
evidence of a primary tumor ranging from 2 to 5.0 cm in size with or without locoregional invasion of tumor cells to as many as three axillary breast lymph nodes.<sup>44, 276</sup> The mechanisms underlying disease recurrence in breast cancer are diverse and include the development of acquired resistance in tumor cells to adjuvant therapeutics, such as trastuzumab (herceptin), a monoclonal antibody-based therapeutic utilized in the treatment of HER2+ breast cancer<sup>295</sup>, as well as due to the impact of increased immune cell signaling following surgical treatment of breast cancer, which has been shown to suppress immune cell function and produce a signaling environment that can promote the growth of residual disease or of nascent tumor cell

populations.<sup>296, 297</sup> Of the 42 proteins that differentiated stage II NR and II R patients, adenosylhomocysteinase (AHCY,  $\log_2$  +2.55 fold in stage II R vs. II NR) was found to be significantly increased in patients exhibiting recurrent disease, a protein that has previously been shown to be increased in cancerous versus normal breast tissues.<sup>298</sup> AHCY catalyzes the conversion of S-adenosylhomocysteine to homocysteine and elevated homocysteine levels have been associated with increased DNA hypomethylation and subsequent aberrant gene expression that has been related to the acquisition of a drug resistant phenotype.<sup>299, 300</sup> Indeed, incubation of a human cell line model of ER+ breast cancer, MCF-7, with homocysteine resulted in resistance to the chemotherapeutic agents doxorubicin and cisplatin as well as upregulation of drug resistance markers.<sup>300</sup> The increased abundance of AHCY in stage II R patient tissues may thus support increased homocysteine production resulting in tumor cell phenotypes that are more refractory to chemotherapeutic treatment. Further assessment of proteins differentiating stage II NR and II R patients revealed protein DJ-1 (PARK7,  $\log_2$  -1.91) as being decreased in stage II R tissues, a factor that is commonly mutated in early-onset, recessive Parkinson's disease and shown to function as a neuroprotectant against oxidative stress in this context and has further been shown to function as an oncogene in various cancer subtypes including breast.<sup>301, 302</sup> Specifically, PARK7 has been observed as being increased in breast cancer tissues and further found at increased levels in serum derived from breast cancer patients, a finding that has led to the proposal of PARK7 as a prognostic marker for breast cancer.<sup>302-304</sup> Investigations into the oncogenic role of PARK7 in lymph node-negative breast cancer tissues has revealed expression of this factor to correlate with increased levels of phosphorylated PKB/Akt protein, a downstream signaling intermediate of the PI3K pathway that is aberrantly activated in tumor cells resulting in an increased cell survival phenotype, as well as decreased levels of the tumor

suppressor PTEN, an antagonist of PI3K signaling and a factor that PARK7 has been shown to directly suppress.<sup>302</sup> However, in instances of breast cancer tissues that exhibited non-phosphorylated PKB/Akt proteins, evidence of PARK7 expression was less frequent, suggesting alternative signaling pathways may be emphasized in these contexts.<sup>302</sup> The observed decrease of PARK7 abundance in stage II R vs. stage II NR breast cancer tissues, a profile that was also observed as significant in comparisons of stage 0 and stage III breast cancer ( $\log_2$  -1.44), supports this hypothesis. We confirmed the decreased abundance of PARK7 in stage II R patients observed by differential proteomics (Figure 24A) via IHC analysis of PARK7 in stage II NR (n=4) and stage II R (n=5) patient tissue samples (Figure 24B and Figure 25) that revealed an equal distribution of stage II NR patients exhibiting a score of 1+ and 2+ for PARK7 and 80% of stage II R patients exhibiting a 1+ score versus 20% that were 2+ for PARK7, verifying our discovery LC-MS/MS data. However, IHC confirmatory studies for decreased PARK7 abundance in stage III versus stage 0 patients only modestly recapitulated the LC-MS/MS observations (44.4% of stage III samples indicating a +2 score versus 55.5% of stage 0 samples), indicating this abundance profile may be more specific for stage II breast tumor cells with the capacity to produce recurrent disease.

**Analysis of functional biological characteristics amongst differentially abundant proteins between non-recurrent and recurrent stage II breast cancer.**

Functional analysis of proteins differentiating recurrence in stage II patients revealed the top ten significant biofunctions to be related to various functional themes, including neurodegenerative (dementia and tauopathy) and inflammatory diseases (rheumatoid arthritis and connective tissue



**Figure 26: Significant, differentially abundant proteins and subsequent functional characteristics identified in comparisons of stage II NR and stage II R breast cancer patient tissue samples.** Top 10 significant biofunctions observed in proteins differentiating stage II NR and stage II R patients. Data reported indicates significant p-values and mean number of proteins observed which clustered with a specific biofunction.

disorders) as well modulation of metabolic functions, such as uptake of calcium, experimentally-induced diabetes and glycolysis of cells (Figure 26). One candidate of interest, being increased in stage II R patients versus II NR, was insulin receptor substrate-1 (IRS-1,  $\log_2 +2.14$ ) that was

further associated with five of the top ten significant biofunctions, i.e. skeletal and muscular disorder, hematological disorder, connective tissue disorder, uptake of calcium and rheumatoid arthritis. IRS-1 functions as a scaffolding protein and signaling intermediate for the insulin-like growth factor-1 receptor signaling cascade and has been implicated in the pathogenesis of several cancers including breast.<sup>305, 306</sup> IRS-1 has been shown to function as an oncogene in breast cancer, being associated with regulation of cell proliferation from gain and loss-of-function analyses in cell line models of breast cancer as well as transgenic murine models.<sup>305, 306</sup> Assessment of IRS-1 expression in clinical breast cancer tissues has revealed ambiguous abundance profiles for this factor, with evidence of both increased and decreased levels being associated with breast cancer disease progression.<sup>305, 306</sup> Interestingly, a study addressing the prognostic utility of assessing IRS-1 abundance in clinical breast cancer tissues revealed that higher levels correlated with an increased risk of disease recurrence following surgical intervention.<sup>306, 307</sup> Indeed, the increased abundance profile we observe for IRS-1 in stage II R versus stage II NR patients supports these observations and provides provocative evidence that assessment of IRS-1 abundance may provide prognostic value in assessing disease recurrence risk.

## **Conclusion**

A comparative, global proteomic analysis of pathologically-defined regions of cancerous epithelium, laser microdissected from FFPE-tissue samples derived from twenty-five breast cancer patients at distinct stages of disease progression revealed differentially abundant proteins indicative of disease stage as well provided functional insights into disease biology underlying *in situ* breast cancer, such as the coordinate modulation of CTNNB1 abundance and related

regulatory machinery, i.e. decreased APC abundance, with progression towards stage III disease and the increase in IRS-1 abundance in stage II R patient tissues that experienced disease recurrence. These data provide robustly, differentially abundant protein candidates that warrant further validation for potential prognostic and predictive power as diagnostic tools in breast cancer. Verification by IHC analysis of primary breast cancer tissues for two candidates derived from these studies support this suggestion, confirming of the observed increase of TSP-1 abundance and loss of PARK7 abundance in later stage disease (stage III) and in tissues derived from patients exhibiting recurrent disease (stage II R) initially observed by proteomic analysis, respectively. These data provide evidence to support the utility of FFPE tissues as a robust resource for retrospective biomarker studies and functional analyses utilizing MS-based proteomics workflows. Further, these data reveal several provocative protein candidates indicative of disease stage and recurrence that warrant continued investigation for their diagnostic utility and biological relevance in breast cancer pathogenesis.

## **5.0 CHAPTER 4: DIFFERENTIAL PROTEOMICS ANALYSIS OF LYMPH-NODE METASTASIS IN ESTROGEN-RECEPTOR POSITIVE BREAST CANCER FROM FORMALIN-FIXED, PARAFFIN EMBEDDED TISSUES**

Estrogen receptor positive (ER+) breast cancer, also known as luminal breast cancer, is characterized by expression of ER as well as variable levels of progesterone receptor (PR) and the HER2 growth factor receptor.<sup>26, 33, 37, 39, 308</sup> The expression characteristics of these latter proteins comprise the basis for further sub-classification, of which there are currently three recognized categories; luminal-A, luminal-B and luminal-C (Table 1).<sup>33, 37, 39, 308</sup> A luminal

breast cancer diagnosis is associated with better overall survival and prognosis relative to HER2+ or basal-like breast cancer subtypes, with luminal-A in particular as exhibiting the highest patient outcome relative to the luminal-B subtype due in part to luminal-A cancers expressing more ER-associated genes and fewer genes associated with cellular proliferation and, further, due to luminal-A cancers tending to be of lower grade than luminal-B.<sup>39, 308, 309</sup> Irrespective of the specific molecular breast cancer subtype identified at the time of diagnosis, evidence of locoregional invasion of primary tumor cells, such as into axillary or sentinel lymph nodes (LN) of the breast, correlates with a poorer disease prognosis overall.<sup>310</sup> Lymph node involvement is commonly established by clinicopathological analyses which include the utilization of axillary ultrasound and ultrasound guided fine-needle aspiration cytology, as well as by characterization of disseminated tumor cell populations in biopsied LN tissues by pathological analysis.<sup>44, 311, 312</sup> However, even with these diagnostic strategies in place, evidence of metastasis has been reported in up to 50% of individuals which initially present with apparent localized disease that has further been observed in 30% of individuals diagnosed with disease lacking apparent LN involvement, indicating that current methodologies lack the sufficient sensitivity needed to identify invasive disease characteristics at the outset of diagnosis.<sup>311</sup> Though the discovery and investigation of biomarkers directed towards the identification of LN disseminated tumor cells as well as which identify characteristics of breast tumor cells capable of LN invasion have been described, none of these candidates have been adopted into regular clinical use.<sup>311, 313, 314</sup>

Thus, as luminal breast cancers represent the most commonly diagnosed clinical subtype of breast cancer, the identification of patients presenting with more aggressive disease

characteristics amongst this population, such as those exhibiting locoregional / LN invasion, may provide a basis in which to consider more aggressive therapeutic options for these individuals, such as treatment with ER-specific endocrine therapy, such as the ER antagonist tamoxifen, or aromatase inhibitors, an enzyme responsible for estrogen synthesis.<sup>39, 315</sup> We therefore undertook a global proteomic analysis of pathologically defined regions of tumor epithelium derived from archival, formalin-fixed paraffin-embedded (FFPE) tissues obtained from a large cohort of ER+ breast cancer patients (n=35) which did (n=16) or did not (n=19) exhibit LN involvement by pathological analysis with the goal of identifying protein abundance characteristics which significantly differentiated luminal-type breast cancer tissues which have or have not undergone LN invasion. Results revealed 7 proteins which significantly differentiated these subgroups, two of which (ARL6IP5,  $\log_2$  -1.07 in LN+ versus LN- disease) and RNA-binding protein 39 (RBM39,  $\log_2$  + 0.66) have been previously implicated in estrogen signaling in breast cancer. Validation efforts of these candidates by immunohistochemical analysis utilizing a tissue microarray of ER+, LN- and LN+ breast cancer tissues are currently underway.

## **Materials and Methods**

### **Breast cancer tissue sample preparation**

Pathologically-defined tumor regions from seventy-five formalin-fixed paraffin-embedded (FFPE) breast cancer patient tissue samples derived from patients with ER+ or ER- breast cancers with (LN+) or without (LN-) lymph node metastasis were processed for global proteomic analysis (Dr. Rohit Bhargava, Department of Pathology, Magee Womens Hospital, Pittsburgh, PA). Processing entailed de-paraffinization of tissue sections via successive 5.0 min

incubations in Xylenes followed by a dehydration and wash regimen consisting of 30 s in 100% ethanol (EtOH), 30 sec. in 95% EtOH, 30 sec. in 70% EtOH and 30 sec. in H<sub>2</sub>O. Sections were lightly stained in Harris hematoxylin for 30 s and washed for 1.0 min in tap water, dehydrated for 30 s in 95% EtOH and air dried for 10 min. Defined regions were then manually scraped from an average of three slides per patient and pooled into 200  $\mu$ L of 80 mM ammonium bicarbonate (AMB) and 20 % acetonitrile (ACN). Samples underwent two freeze-thaw cycles before being incubated at 95 °C for 1 hr followed by 65 °C for 2 h. 1.0  $\mu$ g of porcine sequencing grade modified trypsin (Promega, Madison, WI) was then added to each sample followed by incubation at 37 °C overnight. Samples were then briefly lyophilized to remove ambient ACN, centrifuged and total peptide concentrations were determined from resulting supernatants utilizing the bicinchoninic assay (BCA) and a protein standard curve derived from a diluted stock of 2.0  $\mu$ g/mL  $\alpha$ -casein (Sigma Aldrich). Samples which yielded at least 3.0  $\mu$ g of peptide product (total: n = 48, i.e. ER+ (n = 35) ER- (n = 13) were utilized for downstream LC-MS analyses and were lyophilized and re-suspended to a final concentration of 0.1  $\mu$ g/ $\mu$ L in 0.1% TFA.

### **Liquid chromatography-tandem mass spectrometry analyses**

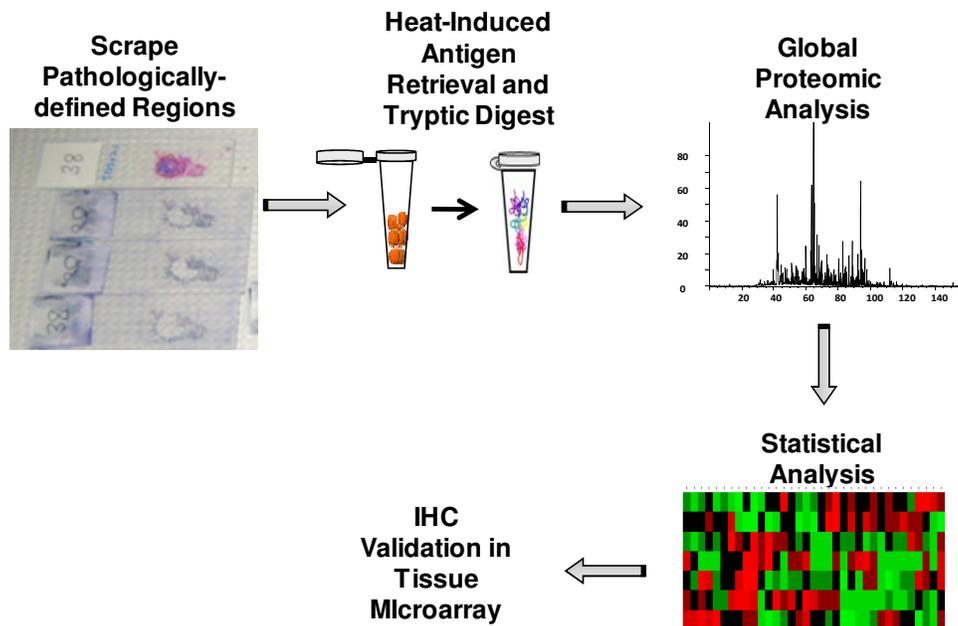
Peptide digests were resolved by nanoflow reverse-phase liquid chromatography (Ultimate 3000, Dionex Inc., Sunnyvale, CA) coupled online via electrospray ionization to a hybrid linear ion trap-Orbitrap Velos mass spectrometer (LTQ-Orbitrap Velos, ThermoFisher Scientific, Inc., San Jose, CA). Quadruplicate injections of 5  $\mu$ L of peptide extracts were resolved on 100  $\mu$ m i.d. by 360  $\mu$ m o.d. by 200 mm long fused silica capillary columns (Polymicro Technologies, Phoenix, AZ) slurry-packed in-house with 5  $\mu$ m, 300 Å pore size C-18 silica-bonded stationary phase (Jupiter, Phenomenex, Torrance, CA). After sample injection, peptides were eluted from the

column using a linear gradient of 2% mobile phase B (0.1% FA in ACN) to 40% mobile phase B over 125 min at a constant flow rate of 200 nL/min followed by a column wash consisting of 95% B for an additional 30 min at a constant flow rate of 400 nL/min. The LTQ-Orbitrap Velos MS was configured to collect high resolution ( $R=60,000$  at  $m/z$  400) broadband mass spectra ( $m/z$  375-1800) from which the twenty-most abundant peptide molecular ions dynamically determined from the MS scan were selected for tandem MS using a relative CID energy of 35%. Dynamic exclusion was utilized to minimize redundant selection of peptides for CID.

### **Peptide Identification and Spectral Count Analysis**

Peptide identifications were obtained by searching the LC-MS/MS data utilizing SEQUEST (BioWorks, v3.2, ThermoScientific) on a 72-node Beowulf cluster against a UniProt-derived human proteome database (version 6/10) obtained from the European Bioinformatics Institute (EBI) using the following parameters: trypsin (KR); full enzymatic-cleavage; two missed cleavages sites; 20 ppm peptide mass tolerance peptide tolerance, 0.5 amu fragment ion tolerance; and variable modifications for methionine oxidation ( $m/z$  15.99492). Resulting peptide identifications were filtered according to specific SEQUEST scoring criteria: delta correlation ( $\Delta C_n$ )  $\geq 0.08$  and charge state dependent cross correlation ( $X_{corr}$ )  $\geq 1.9$  for  $[M+H]^+$ ,  $\geq 2.2$  for  $[M+2H]^{2+}$ , and  $\geq 3.5$  for  $[M+3H]^{3+}$ . To determine statistically significant changes in protein abundance across ER+ patient samples (n=35), a hierarchical supervised cluster analysis of peptides identified from ER+/ LN- (n=19) versus ER+/ LN+ (n=16) samples was performed in which the variance in total spectral count peptides identified was determined utilizing the Mann-Whitney rank-sum test (significance level  $p \leq 0.05$ , Fisher's exact test) paired with the

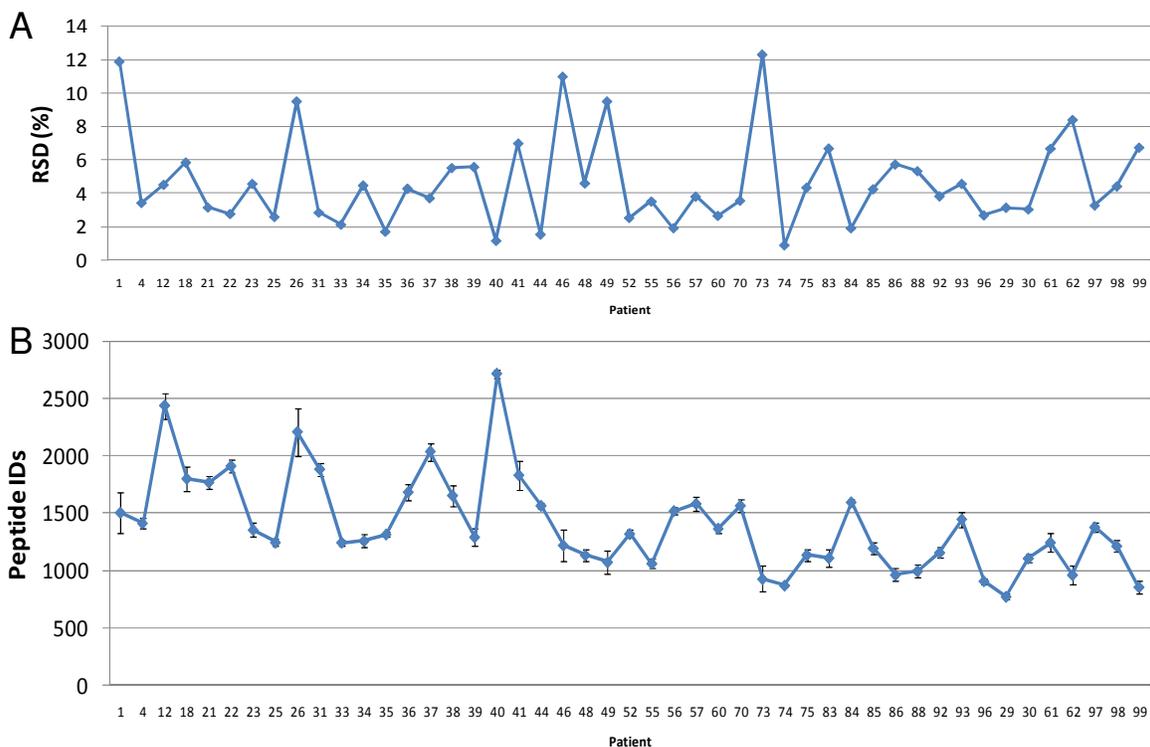
filter criteria requiring that 50% of the samples in a supervised group had a minimum peptide count of 2 or greater for a given protein.



**Figure 27: Analytical workflow utilized for differential proteomic analysis of ER+ formalin-fixed paraffin embedded (FFPE) primary breast cancer tissues.**

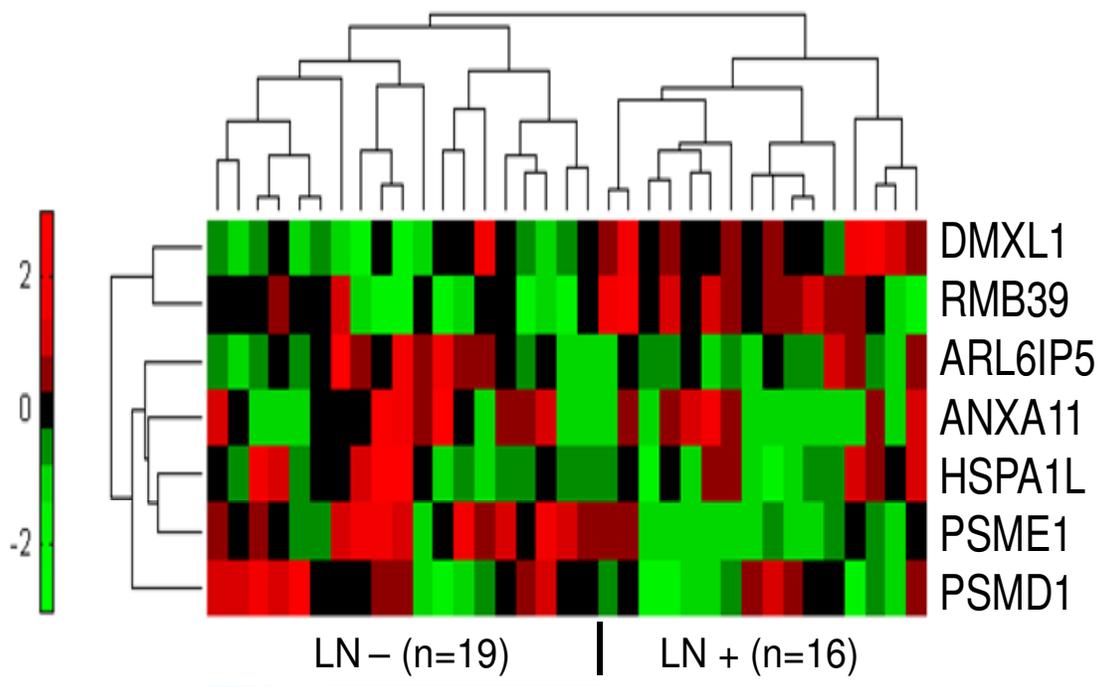
### **Results and Discussion:**

Global proteomic analyses of pathologically-defined regions of cancerous epithelium scraped from FFPE-tissue samples derived from forty-eight ER+ breast cancer patients (n=35) and ER- (n=13) disease yielded 19,947 total peptide identifications (Fig 27). Merits of analytical performance were monitored by comparing peptides identified between replicate injections for all forty-eight patient samples, revealing an average relative standard deviation (RSD) of 5.3% ( $\pm 5.0\%$ ) (Fig 28A). Equivalency of protein digest input for LC-MS/MS analyses was



**Figure 28: Metric of LC-MS performance during analysis of 48 patient samples.** (A): Assessment of peptides identified between LC-MS/MS technical replicates revealed a mean RSD of 5.3% ( $\pm 5.0$ ) across data set. (B): Average peptides identified for 48 patient tissue samples analyzed in quadruplicate over a ~6-week week period. Assessment of peptides identified across data set revealed a 28.7 % RSD in total peptides identified.

determined by comparison of spectral count values for total peptides identified, indicating that an average of 1456 ( $\pm 417$ ) peptide identifications was achieved across all forty- eight patient samples (RSD of 28.7%) (Figure 28B). Compilation of these results revealed an average of 528 ( $\pm 128$ ) proteins was identified by at least two peptides across all samples analyzed (RSD of 24.3%). These measures critically underpin the determination of significant, differentially abundant proteins by spectral counting without the need for normalization.



**Figure 29: Hierarchical supervised cluster analysis of proteomic analysis derived from 35 ER+ patient tissue samples  $\pm$  LN metastasis:** Cluster analysis was performed on total spectral count peptides utilizing the Mann–Whitney rank-sum test (significance level  $p \leq 0.05$ , Fisher’s exact test) paired with the filter criteria requiring that 50% of the samples in a supervised group had a minimum peptide count of 2 or greater for a given protein. Results revealed a total of 7 proteins which significantly differentiated these groups by this criteria.

**Characterization of differentially abundant proteins in ER+ patients with (LN+) or without (LN-) Lymph Node Metastasis.**

Significant, differentially abundant proteins were established by hierarchical supervised cluster analysis of ER+/ LN+ (n=16) or ER+/ LN- (n=19) proteomic data sets in which at least 50% of the samples within a supervised group exhibited spectral count values of 2 or greater for a given

protein of interest (Fig 29). Results revealed 7 proteins which significantly differentiated these two sample groups (Table 6). Among these seven candidates, two proteins have previously been implicated in estrogen signaling in breast cancer, ADP-ribosylation-like factor 6 interacting protein 5 (ARL6IP5,  $\log_2$  -1.07 in LN+ versus LN- disease)<sup>316</sup> and RNA-binding motif protein 39 (RBM39,  $\log_2$  + 0.66).<sup>317, 318</sup>

ARL6IP5, also known as JWA, is a microtubule-localized cytoskeletal protein that has been implicated in estrogen signaling as treatment of an ER+ human breast cancer cell line, MCF7, with fenvalerate, a pyrethroid insecticide shown to possess estrogenic activity, resulted in a dose-dependent decrease in ARL6IP5 expression and a coordinate increase in cell proliferation.<sup>316</sup> Additional analyses of ARL6IP5 function have revealed this protein plays a regulatory role in cellular differentiation and apoptosis signaling and has been shown to be induced in response to cellular stressors, such as heat shock and oxidative stress, in a variety of cellular systems.<sup>319-321</sup> In the context of oxidative stress, ARL6IP5 has been shown to function as a protectant against benzo[a]pyrene, an environmental carcinogen, and H<sub>2</sub>O<sub>2</sub>-induced DNA damage in NIH-3T3 fibroblast cells and further has been shown to be induced in MCF7 cells in response to H<sub>2</sub>O<sub>2</sub> treatment<sup>320, 322, 323</sup> Furthermore, recent evidence investigating this phenomenon has revealed that oxidative stress-induced ARL6IP5 promotes expression of XRCC1, a key DNA repair mediator associated with the base excision repair pathway, being further shown to complex with this protein at single-strand break sites in DNA as well as promote stabilization of XRCC1 through the prevention of ubiquitin-mediated proteolysis.<sup>22</sup> Interestingly, an additional function for ARL6IP5 has revealed this factor to be involved in regulation of cell migration, as loss of ARL6IP5 expression has been shown to greatly increase cell migratory characteristics of several

cancer cell lines including HeLa cells, B16 melanoma cells as well as HCCLM3 hepatocellular carcinoma cells.<sup>319</sup> Assessment of these findings in the context of the observed decrease of ARL6IP5 abundance in LN+ versus LN- disease in our analysis, breast cancer cell types comprising ER+/LN+ disease may support a more robust response to estrogen leading to loss of ARL6IP5 versus ER+/LN- cells, thus resulting in an increase in cell migratory activity as well as impaired DNA repair capabilities which may promote the development of more aberrant and aggressive tumor cell types.

HGNC ID	Protein Name	LN+ vs. LN- Log <sub>2</sub> Fold-Change	Cellular Localization	Functional Type
ANXA11	annexin A11	-1.261	Nucleus	other
<b>ARL6IP5</b>	ADP-ribosylation-like factor 6 interacting protein 5	-1.074	Cytoplasm	other
DMXL1	Dmx-like 1	0.499	unknown	other
HSPA1L	heat shock 70kDa protein 1-like	-0.613	Cytoplasm	other
PSMD1	proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	-0.624	Cytoplasm	other
PSME1	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	-1.004	Cytoplasm	other
<b>RBM39</b>	RNA binding motif protein 39	0.663	Nucleus	transcription regulator

**Table 6: Significant proteins differentiating ER+/LN+ and ER+/LN- breast cancer tissues:** Seven proteins identified which significantly differentiated ER+/LN+ and ER+/LN- groups. HGNC ID's for proteins previously implication in estrogen signaling in breast cancer are in bold.

RBM39, also known as CAPER, is a nuclear-localized factor that has been observed as being associated with spliceosome complexes and further to function as a specific transcriptional co-activator for AP-1, ER- $\alpha$  and ER- $\beta$ .<sup>317</sup> RBM39 has been shown to directly interact with ER isoforms and to promote transcription of genes containing estrogen promoter response elements in response to exogenous treatment with E2 estradiol.<sup>317</sup> Further analysis of mammary epithelia

derived from transgenic mice homozygous null for caveolin-1, a plasma membrane-bound scaffolding protein that has been shown to be exclusively mutated in ER- $\alpha$  human breast cancers, revealed increased expression of RBM39 protein in these tissues which were further hyper-responsive to exogenous estrogen relative to wild-type controls.<sup>318</sup> Our observed increase in RBM39 abundance in LN+ versus LN- patients therefore supports the aforementioned hypothesis that ER+/LN+ breast tumor cell types may be more responsive to estrogen versus ER+/LN- cells. These observed differential abundance profiles for ARL6IP5 and RBM39 in LN+ versus LN- disease may thus produce a molecular context which supports more aggressive tendencies which translate to greater invasive potential.

## **Conclusion**

These analyses have yielded a panel of seven differentially abundant proteins which significantly differentiated a large cohort of ER+ breast cancer patients presenting with or without lymph node metastasis. Further inspection of these results revealed two proteins, ARL6IP5 and RBM39, which have been previously implicated in estrogen signaling in breast cancer that exhibit an abundance profile which supports that ER+/LN+ breast tumor cells may have an increased capacity to respond to estrogen and thus may be a factor underlying the invasive characteristics exhibited by these cell types. Validation efforts of these candidates by immunohistochemical analysis utilizing a tissue microarray of ER+, LN- and LN+ breast cancer tissues are currently underway.

## 6.0: SUMMARY AND CONCLUSIONS

The molecular heterogeneity of breast cancer underscores the benefits to be gained from analysis of this disease utilizing high-throughput proteomic analyses. This dissertation details the investigations of key subjects in breast cancer biology focused on the characterization of endogenous and experimentally-induced disease biology characteristics utilizing the application of LC-MS based proteomic analyses of both *in vitro* models of breast cancer as well as primary clinical samples.

Data is presented which describes a combined global and functional proteomic strategy to discern governing functional roles for mutually, differentially abundant proteins which are observed across three tumorigenic cell lines models of human breast cancer corresponding to the most common molecular subtypes of disease clinically encountered, ER+, HER+ and basal-like disease, relative to a non-transformed model cell line of normal mammary epithelium. Though microarray evidence has revealed that breast tumor cells bearing these discrete hormone and growth factors receptor profiles exhibit subtype-specific gene expression signatures, observations of a mutual population of proteins exhibiting virtually identical abundance trends across these three subtypes of breast cancer indicates that there are central molecular characteristics which are conserved across breast cancer cells which may represent core characteristics of disease biology underlying these systems. Further functional analysis revealed a predominance of differentially abundant proteins associated with the regulation of cell spreading, with further analysis indicating many of these pro-cell spreading factors being lost across all tumorigenic breast cancer cells. Protein network analysis of these mutual candidates revealed a central factor associated with regulation of cell spreading , further known to be constitutively active in breast

cancer cells, focal adhesion kinase (FAK), as being associated with several of these dysregulated cell spreading-associated factors as well as other mutually dysregulated candidates which have been shown to inhibit expression of activity of FAK in other cell types and cancer disease pathologies. As FAK has been shown to play an integral role in maintaining characteristics of breast tumor cell tumorigenicity, these data revealed several targets which warrant further investigation for their role in regulating FAK in breast cancer. Further data is provided detailing a facile workflow towards the elucidation of functional biological characteristics in complex proteome data sets, an area of growing interest in proteomic research. Further analyses are described which focus on the fairly novel topic of applying proteomic analysis to the discovery of microRNA targets, a research area which has exploded in the last decade with the discovery of this previously unknown post-transcriptional gene regulatory mechanism. Investigation into this subject has revealed the parameters governing microRNA-mediated gene targeting to be complex, such as the tolerance of variable base-pairing within microRNA-target duplex “seed” regions as well as the impact of RNA secondary structure on microRNA target site accessibility. Further, the clear delineation of the characteristics which govern regulation of microRNA-mediated target degradation or translational repression have yet to be established. The efforts described herein were two-fold in that they were focused on 1) devising a microRNA target discovery platform utilizing proteomic analysis and further 2) towards the elucidation of experimentally observed targets of a microRNA which has been implicated in breast cancer pathogenesis, miR-145. The results of these efforts yielded a proteomics-based microRNA target discovery workflow focused on the identification of high confidence microRNA targets bearing intact 3’UTR “seed” regions as well as the identification of a panel of microRNA target candidates which were mutually regulated across two cell line models of metastatic breast

cancer, MDA and SKB cells. Comparative analysis of the robustly different levels of miR-145 achieved upon restoration of expression in MDA and SKB cells with the prevalence of miR-145 “seed” containing protein targets observed as decreased in abundance in response to miR-145 provided evidence to support the hypothesis that microRNA activity is concentration dependent. This is a vital issue to consider in the context of therapeutic microRNA restoration or silencing strategies which have been proposed for various disease associated microRNAs. Further investigation of microRNA activity which attempts to mimic physiologically observed levels of a microRNA in endogenous tissue would provide a better experimental context in which to investigate microRNA function as well as which would provide expression benchmarks for which to achieve in gene therapy strategies aimed at modulating microRNA levels.

Data is further described which focuses on the global proteomic analysis of clinical breast cancer tissues towards the identification of protein characteristics representative of disease stage as well as indicative of recurrent disease. As high-throughput sample analysis is suited to proteomic workflows, the opportunity to perform large-scale analysis of clinical samples is an attractive capability towards the discovery of conserved protein abundance characteristics which are indicative of *in situ* disease. Further, the ability to draw upon basic research findings describing the roles of different protein candidates in various cellular systems and disease processes enables the elucidation of functional biological characteristics from large-scale proteome analyses which may better clarify disease and tissue-specific molecular processes. These data further add to the growing arsenal of proteomic analyses utilizing formalin-fixed, paraffin embedded (FFPE) tissues samples for retrospective biomarker investigations. The large archives of FFPE tissue banks world-wide represent an excellent resource in which to obtain large numbers of clinical

samples for analysis which will add sufficient power to observations made that may more readily expedite findings to clinical applications. The results of the FFPE proteomic analyses described herein adds further evidence to support the utility of these sample sources for retrospective proteomic investigations. This was evidenced in comparisons of early and late-stage breast cancer revealed differential abundance profiles for several proteins are well established in breast cancer pathogenesis, such as a loss of CDH1 (E-Cadherin) as well as modulation of CTNNB1 (beta-catenin) and CTNNB1 regulatory machinery, i.e. APC (adenomatous polyposis coli). Further, the verification of the abundance profiles for two novel factors associated with different disease stage and recurrent disease characteristics by immunohistochemical analysis, TSP-1 (thrombospondin) and PARK7 (Protein DJ-1) adds further credence to the utility of these sample sources for proteomic study. Lastly, the identification of factors differentially abundant in ER+ breast cancer tissues with and without lymph node (LN) metastasis, which have been previously implicated in estrogen signaling and further which exhibited abundance profiles supporting the hypothesis that ER+ breast tumor cells capable of achieving LN metastasis may be more responsive to estrogen signaling, further adds support to the observations made utilizing these sample sources. The differential abundance profiles for proteins associated with disease stage as well as indicative of recurrent disease identified in these analyses warrant further investigation for their utility in the clinical diagnosis of breast cancer.

The results reported herein overall detail the development of novel proteomic workflows as well as the identification of both established and novel protein abundance characteristics associated with neoplastic progression in breast cancer which will benefit the further development of investigating functional biology in proteomic analyses and further will result in expansion of the

basic biological understanding of breast cancer as well as which will provide prospective clinical tools towards the more facile diagnosis of this disease.

## **7.0 BIBLIOGRAPHY:**

1. Jemal, A.; Siegel, R.; Xu, J.; Ward, E., Cancer Statistics, 2010. *CA Cancer J Clin.*
2. Salih, A. K.; Fentiman, I. S., Breast cancer prevention: present and future. *Cancer Treat Rev* **2001**, *27* (5), 261-73.
3. Kenemans, P.; Verstraeten, R. A.; Verheijen, R. H., Oncogenic pathways in hereditary and sporadic breast cancer. *Maturitas* **2004**, *49* (1), 34-43.
4. Hopper, J. L., Genetic epidemiology of female breast cancer. *Semin Cancer Biol* **2001**, *11* (5), 367-74.
5. Polyak, K., Breast cancer: origins and evolution. *J Clin Invest* **2007**, *117* (11), 3155-63.
6. McSherry, E. A.; Donatello, S.; Hopkins, A. M.; McDonnell, S., Molecular basis of invasion in breast cancer. *Cell Mol Life Sci* **2007**, *64* (24), 3201-18.
7. Keen, J. C.; Davidson, N. E., The biology of breast carcinoma. *Cancer* **2003**, *97* (3 Suppl), 825-33.
8. Anderson, E.; Clarke, R. B.; Howell, A., Estrogen responsiveness and control of normal human breast proliferation. *J Mammary Gland Biol Neoplasia* **1998**, *3* (1), 23-35.
9. Hahn, W. C.; Weinberg, R. A., Modelling the molecular circuitry of cancer. *Nat Rev Cancer* **2002**, *2* (5), 331-41.
10. Hanahan, D.; Weinberg, R. A., The hallmarks of cancer. *Cell* **2000**, *100* (1), 57-70.
11. Bissell, M. J.; Radisky, D., Putting tumours in context. *Nat Rev Cancer* **2001**, *1* (1), 46-54.
12. Ghajar, C. M.; Bissell, M. J., Extracellular matrix control of mammary gland morphogenesis and tumorigenesis: insights from imaging. *Histochem Cell Biol* **2008**, *130* (6), 1105-18.

13. Al-Hajj, M.; Wicha, M. S.; Benito-Hernandez, A.; Morrison, S. J.; Clarke, M. F., Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **2003**, *100* (7), 3983-8.
14. Smalley, M.; Ashworth, A., Stem cells and breast cancer: A field in transit. *Nat Rev Cancer* **2003**, *3* (11), 832-44.
15. Moulis, S.; Sgroi, D. C., Re-evaluating early breast neoplasia. *Breast Cancer Res* **2008**, *10* (1), 302.
16. Kass, L.; Erler, J. T.; Dembo, M.; Weaver, V. M., Mammary epithelial cell: influence of extracellular matrix composition and organization during development and tumorigenesis. *Int J Biochem Cell Biol* **2007**, *39* (11), 1987-94.
17. Streuli, C. H.; Gilmore, A. P., Adhesion-mediated signaling in the regulation of mammary epithelial cell survival. *J Mammary Gland Biol Neoplasia* **1999**, *4* (2), 183-91.
18. Bissell, M. J.; Weaver, V. M.; Lelievre, S. A.; Wang, F.; Petersen, O. W.; Schmeichel, K. L., Tissue structure, nuclear organization, and gene expression in normal and malignant breast. *Cancer Res* **1999**, *59* (7 Suppl), 1757-1763s; discussion 1763s-1764s.
19. Graslund, S.; Nordlund, P.; Weigelt, J.; Hallberg, B. M.; Bray, J.; Gileadi, O.; Knapp, S.; Oppermann, U.; Arrowsmith, C.; Hui, R.; Ming, J.; dhe-Paganon, S.; Park, H. W.; Savchenko, A.; Yee, A.; Edwards, A.; Vincentelli, R.; Cambillau, C.; Kim, R.; Kim, S. H.; Rao, Z.; Shi, Y.; Terwilliger, T. C.; Kim, C. Y.; Hung, L. W.; Waldo, G. S.; Peleg, Y.; Albeck, S.; Unger, T.; Dym, O.; Prilusky, J.; Sussman, J. L.; Stevens, R. C.; Lesley, S. A.; Wilson, I. A.; Joachimiak, A.; Collart, F.; Dementieva, I.; Donnelly, M. I.; Eschenfeldt, W. H.; Kim, Y.; Stols, L.; Wu, R.; Zhou, M.; Burley, S. K.; Emtage, J. S.; Sauder, J. M.; Thompson, D.; Bain, K.; Luz, J.; Gheyi, T.; Zhang, F.; Atwell, S.; Almo, S. C.; Bonanno, J. B.; Fiser, A.; Swaminathan, S.; Studier, F. W.; Chance, M. R.; Sali, A.; Acton, T. B.; Xiao, R.; Zhao, L.; Ma, L. C.; Hunt, J. F.; Tong, L.; Cunningham, K.; Inouye, M.; Anderson, S.; Janjua, H.; Shastry, R.; Ho, C. K.; Wang, D.; Wang, H.; Jiang, M.; Montelione, G. T.; Stuart, D. I.; Owens, R. J.; Daenke, S.; Schutz, A.; Heinemann, U.; Yokoyama, S.; Bussow, K.; Gunsalus, K. C., Protein production and purification. *Nat Methods* **2008**, *5* (2), 135-46.
20. Tredan, O.; Galmarini, C. M.; Patel, K.; Tannock, I. F., Drug resistance and the solid tumor microenvironment. *J Natl Cancer Inst* **2007**, *99* (19), 1441-54.
21. Berg, J. M.; Tymoczko, J. L.; Stryer, L., *Biochemistry*. 6th ed.; W.H. Freeman: New York, 2007; p 1 v. (various pagings).
22. Warburg, O., On respiratory impairment in cancer cells. *Science* **1956**, *124* (3215), 269-70.

23. Leong, S. P.; Cady, B.; Jablons, D. M.; Garcia-Aguilar, J.; Reintgen, D.; Jakub, J.; Pendas, S.; Duhaime, L.; Cassell, R.; Gardner, M.; Giuliano, R.; Archie, V.; Calvin, D.; Mensha, L.; Shivers, S.; Cox, C.; Werner, J. A.; Kitagawa, Y.; Kitajima, M., Clinical patterns of metastasis. *Cancer Metastasis Rev* **2006**, *25* (2), 221-32.
24. Fu, J.; Jeffrey, S. S., Transcriptomic signatures in breast cancer. *Mol Biosyst* **2007**, *3* (7), 466-72.
25. Perou, C. M.; Sorlie, T.; Eisen, M. B.; van de Rijn, M.; Jeffrey, S. S.; Rees, C. A.; Pollack, J. R.; Ross, D. T.; Johnsen, H.; Akslen, L. A.; Fluge, O.; Pergamenschikov, A.; Williams, C.; Zhu, S. X.; Lonning, P. E.; Borresen-Dale, A. L.; Brown, P. O.; Botstein, D., Molecular portraits of human breast tumours. *Nature* **2000**, *406* (6797), 747-52.
26. Jeffrey, S. S.; Lonning, P. E.; Hillner, B. E., Genomics-based prognosis and therapeutic prediction in breast cancer. *J Natl Compr Canc Netw* **2005**, *3* (3), 291-300.
27. Sims, A. H.; Howell, A.; Howell, S. J.; Clarke, R. B., Origins of breast cancer subtypes and therapeutic implications. *Nat Clin Pract Oncol* **2007**, *4* (9), 516-25.
28. Sorlie, T.; Perou, C. M.; Tibshirani, R.; Aas, T.; Geisler, S.; Johnsen, H.; Hastie, T.; Eisen, M. B.; van de Rijn, M.; Jeffrey, S. S.; Thorsen, T.; Quist, H.; Matese, J. C.; Brown, P. O.; Botstein, D.; Eystein Lonning, P.; Borresen-Dale, A. L., Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **2001**, *98* (19), 10869-74.
29. Sgroi, D. C.; Teng, S.; Robinson, G.; LeVangie, R.; Hudson, J. R., Jr.; Elkahoul, A. G., In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res* **1999**, *59* (22), 5656-61.
30. Bhargava, R.; Beriwal, S.; Dabbs, D. J.; Ozbek, U.; Soran, A.; Johnson, R. R.; Brufsky, A. M.; Lembersky, B. C.; Ahrendt, G. M., Immunohistochemical surrogate markers of breast cancer molecular classes predicts response to neoadjuvant chemotherapy: a single institutional experience with 359 cases. *Cancer* **116** (6), 1431-9.
31. Tang, P.; Skinner, K. A.; Hicks, D. G., Molecular classification of breast carcinomas by immunohistochemical analysis: are we ready? *Diagn Mol Pathol* **2009**, *18* (3), 125-32.
32. Cheang, M. C.; Voduc, D.; Bajdik, C.; Leung, S.; McKinney, S.; Chia, S. K.; Perou, C. M.; Nielsen, T. O., Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res* **2008**, *14* (5), 1368-76.
33. Tamimi, R. M.; Baer, H. J.; Marotti, J.; Galan, M.; Galaburda, L.; Fu, Y.; Deitz, A. C.; Connolly, J. L.; Schnitt, S. J.; Colditz, G. A.; Collins, L. C., Comparison of molecular

phenotypes of ductal carcinoma in situ and invasive breast cancer. *Breast Cancer Res* **2008**, *10* (4), R67.

34. Manni, A.; Arafah, B.; Pearson, O. H., Estrogen and progesterone receptors in the prediction of response of breast cancer to endocrine therapy. *Cancer* **1980**, *46* (12 Suppl), 2838-41.

35. Elledge, R.; Hayes, D. F.; Lerner, R., Hormone receptors in breast cancer: Measurement and clinical implications. *Uptodate. Waltham, MA.* **2010**.

36. Slamon, D. J.; Clark, G. M.; Wong, S. G.; Levin, W. J.; Ullrich, A.; McGuire, W. L., Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **1987**, *235* (4785), 177-82.

37. Bhargava, R.; Dabbs, D. J., Luminal B breast tumors are not HER2 positive. *Breast Cancer Res* **2008**, *10* (5), 404; author reply 405.

38. Bhargava, R.; Beriwal, S.; Dabbs, D. J.; Ozbek, U.; Soran, A.; Johnson, R. R.; Brufsky, A. M.; Lembersky, B. C.; Ahrendt, G. M., Immunohistochemical surrogate markers of breast cancer molecular classes predicts response to neoadjuvant chemotherapy: a single institutional experience with 359 cases. *Cancer* **2010**, *116* (6), 1431-9.

39. Brenton, J. D.; Carey, L. A.; Ahmed, A. A.; Caldas, C., Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol* **2005**, *23* (29), 7350-60.

40. Esserman, L. J.; Joe, B. N., Diagnostic evaluation of women with suspected breast cancer. *Uptodate. Waltham, MA.* **2010**.

41. Taghian, A.; El-Ghamry, M. N.; Merajver, S. D., Clinical features and management of locally advanced breast cancer. *Uptodate. Waltham, MA.* **2010**.

42. Bleiweiss, I. J., Pathology of breast cancer: The invasive carcinomas. *Uptodate. Waltham, MA.* **2010**.

43. Bleiweiss, I. J., Pathology of breast cancer: The in situ carcinomas. *Uptodate. Waltham, MA.* **2010**.

44. Hayes, D. F., Tumor node metastasis (TNM) staging classification for breast cancer. *Uptodate. Waltham, MA.* **2010**.

45. Hayes, D. F., Measurement of prognostic factors in breast cancer. *Uptodate. Waltham, MA.* **2010**.

46. Martin, M.; Gonzalez Palacios, F.; Cortes, J.; de la Haba, J.; Schneider, J., Prognostic and predictive factors and genetic analysis of early breast cancer. *Clin Transl Oncol* **2009**, *11* (10), 634-42.
47. Wilkins, M., Proteomics data mining. *Expert Rev Proteomics* **2009**, *6* (6), 599-603.
48. Wu, C. C.; MacCoss, M. J., Shotgun proteomics: tools for the analysis of complex biological systems. *Curr Opin Mol Ther* **2002**, *4* (3), 242-50.
49. Gygi, S. P.; Aebersold, R., Mass spectrometry and proteomics. *Curr Opin Chem Biol* **2000**, *4* (5), 489-94.
50. Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P., The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **2005**, *35* (3), 265-73.
51. Siuzdak, G., *The Expanding Role of Mass Spectrometry in Biotechnology*. 2 ed.; MCC Press: 2006.
52. Garrels, J. I.; Franza, B. R.; Patterson, S. D.; Latham, K.; Soltzer, S.; Chang, C.; Latter, G., Protein Databases Constructed by Quantitative 2D Gel Analysis and Protein Identification from 2D gels. *Journal of Protein Chemistry* **1992**, *11* (2), 394-395.
53. McDonald, W. H.; Yates, J. R., 3rd, Proteomic tools for cell biology. *Traffic* **2000**, *1* (10), 747-54.
54. McDonald, W. H.; Yates, J. R., 3rd, Shotgun proteomics and biomarker discovery. *Dis Markers* **2002**, *18* (2), 99-105.
55. Stokes, M. P.; Rush, J.; Macneill, J.; Ren, J. M.; Sprott, K.; Nardone, J.; Yang, V.; Beausoleil, S. A.; Gygi, S. P.; Livingstone, M.; Zhang, H.; Polakiewicz, R. D.; Comb, M. J., Profiling of UV-induced ATM/ATR signaling pathways. *Proc Natl Acad Sci U S A* **2007**, *104* (50), 19855-60.
56. Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P., Multidimensional separations-based shotgun proteomics. *Chem Rev* **2007**, *107* (8), 3654-86.
57. Eng, J. K.; McCormack, A. L.; Yates II, J. R., An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom* **1994**, *5*, 976-989.

58. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551-67.
59. Sadygov, R. G.; Cociorva, D.; Yates, J. R., 3rd, Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods* **2004**, *1* (3), 195-202.
60. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005**, *40* (4), 430-43.
61. Picotti, P.; Aebersold, R.; Domon, B., The implications of proteolytic background for shotgun proteomics. *Mol Cell Proteomics* **2007**, *6* (9), 1589-98.
62. Olsen, J. V.; Ong, S. E.; Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **2004**, *3* (6), 608-14.
63. Patel, V. J.; Thalassinou, K.; Slade, S. E.; Connolly, J. B.; Crombie, A.; Murrell, J. C.; Scrivens, J. H., A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J Proteome Res* **2009**, *8* (7), 3752-9.
64. Bogdanov, B.; Smith, R. D., Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom Rev* **2005**, *24* (2), 168-200.
65. Zhu, W.; Smith, J. W.; Huang, C. M., Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**, *2010*, 840518.
66. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002**, *1* (5), 376-86.
67. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004**, *3* (12), 1154-69.
68. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008**, *26* (12), 1367-72.

69. Hondermarck, H.; Dolle, L.; El Yazidi-Belkoura, I.; Vercoutter-Edouart, A. S.; Adriaenssens, E.; Lemoine, J., Functional proteomics of breast cancer for signal pathway profiling and target discovery. *J Mammary Gland Biol Neoplasia* **2002**, *7* (4), 395-405.
70. Souchelnytskyi, S., Proteomics in studies of signal transduction in epithelial cells. *J Mammary Gland Biol Neoplasia* **2002**, *7* (4), 359-71.
71. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C., The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **2000**, *21* (6), 1104-15.
72. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001**, *19* (3), 242-7.
73. Second, T. P.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V., Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal Chem* **2009**, *81* (18), 7757-65.
74. Hondermarck, H.; Tastet, C.; El Yazidi-Belkoura, I.; Toillon, R. A.; Le Bourhis, X., Proteomics of breast cancer: the quest for markers and therapeutic targets. *J Proteome Res* **2008**, *7* (4), 1403-11.
75. Dua, R. S.; Isacke, C. M.; Gui, G. P., The intraductal approach to breast cancer biomarker discovery. *J Clin Oncol* **2006**, *24* (7), 1209-16.
76. Harris, L.; Fritsche, H.; Mennel, R.; Norton, L.; Ravdin, P.; Taube, S.; Somerfield, M. R.; Hayes, D. F.; Bast, R. C., Jr., American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* **2007**, *25* (33), 5287-312.
77. Baek, D.; Villen, J.; Shin, C.; Camargo, F. D.; Gygi, S. P.; Bartel, D. P., The impact of microRNAs on protein output. *Nature* **2008**, *455* (7209), 64-71.
78. Selbach, M.; Schwanhaussner, B.; Thierfelder, N.; Fang, Z.; Khanin, R.; Rajewsky, N., Widespread changes in protein synthesis induced by microRNAs. *Nature* **2008**, *455* (7209), 58-63.
79. Yang, Y.; Chaerkady, R.; Beer, M. A.; Mendell, J. T.; Pandey, A., Identification of miR-21 targets in breast cancer cells using a quantitative proteomic approach. *Proteomics* **2009**, *9* (5), 1374-84.

80. Abramovitz, M.; Leyland-Jones, B., A systems approach to clinical oncology: focus on breast cancer. *Proteome Sci* **2006**, *4*, 5.
81. Wang, Y.; Hanley, R.; Klemke, R. L., Computational methods for comparison of large genomic and proteomic datasets reveal protein markers of metastatic cancer. *J Proteome Res* **2006**, *5* (4), 907-15.
82. Adam, P. J.; Boyd, R.; Tyson, K. L.; Fletcher, G. C.; Stamps, A.; Hudson, L.; Poyser, H. R.; Redpath, N.; Griffiths, M.; Steers, G.; Harris, A. L.; Patel, S.; Berry, J.; Loader, J. A.; Townsend, R. R.; Daviet, L.; Legrain, P.; Parekh, R.; Terrett, J. A., Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. *J Biol Chem* **2003**, *278* (8), 6482-9.
83. Imai, K.; Ichibangase, T.; Saitoh, R.; Hoshikawa, Y., A proteomics study on human breast cancer cell lines by fluorogenic derivatization-liquid chromatography/tandem mass spectrometry. *Biomed Chromatogr* **2008**, *22* (11), 1304-14.
84. Kulasingam, V.; Diamandis, E. P., Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. *Mol Cell Proteomics* **2007**, *6* (11), 1997-2011.
85. Lai, T. C.; Chou, H. C.; Chen, Y. W.; Lee, T. R.; Chan, H. T.; Shen, H. H.; Lee, W. T.; Lin, S. T.; Lu, Y. C.; Wu, C. L.; Chan, H. L., Secretomic and proteomic analysis of potential breast cancer markers by two-dimensional differential gel electrophoresis. *J Proteome Res* *9* (3), 1302-22.
86. Patwardhan, A. J.; Strittmatter, E. F.; Camp, D. G., 2nd; Smith, R. D.; Pallavicini, M. G., Quantitative proteome analysis of breast cancer cell lines using <sup>18</sup>O-labeling and an accurate mass and time tag strategy. *Proteomics* **2006**, *6* (9), 2903-15.
87. Smith, L.; Welham, K. J.; Watson, M. B.; Drew, P. J.; Lind, M. J.; Cawkwell, L., The proteomic analysis of cisplatin resistance in breast cancer cells. *Oncol Res* **2007**, *16* (11), 497-506.
88. Wu, S. L.; Hancock, W. S.; Goodrich, G. G.; Kunitake, S. T., An approach to the proteomic analysis of a breast cancer cell line (SKBR-3). *Proteomics* **2003**, *3* (6), 1037-46.
89. Hathout, Y.; Gehrman, M. L.; Chertov, A.; Fenselau, C., Proteomic phenotyping: metastatic and invasive breast cancer. *Cancer Lett* **2004**, *210* (2), 245-53.
90. Patwardhan, A. J.; Strittmatter, E. F.; Camp, D. G., 2nd; Smith, R. D.; Pallavicini, M. G., Comparison of normal and breast cancer cell lines using proteome, genome, and interactome data. *J Proteome Res* **2005**, *4* (6), 1952-60.

91. Chen, S. T.; Pan, T. L.; Juan, H. F.; Chen, T. Y.; Lin, Y. S.; Huang, C. M., Breast tumor microenvironment: proteomics highlights the treatments targeting secretome. *J Proteome Res* **2008**, *7* (4), 1379-87.
92. Menon, R.; Omenn, G. S., Proteomic characterization of novel alternative splice variant proteins in human epidermal growth factor receptor 2/neu-induced breast cancers. *Cancer Res* **70** (9), 3440-9.
93. Choong, L. Y.; Lim, S.; Chong, P. K.; Wong, C. Y.; Shah, N.; Lim, Y. P., Proteome-wide profiling of the MCF10AT breast cancer progression model. *PLoS One* **5** (6), e11030.
94. Kim, S. H.; Miller, F. R.; Tait, L.; Zheng, J.; Novak, R. F., Proteomic and phosphoproteomic alterations in benign, premalignant and tumor human breast epithelial cells and xenograft lesions: biomarkers of progression. *Int J Cancer* **2009**, *124* (12), 2813-28.
95. Miller, F. R., Xenograft models of premalignant breast disease. *J Mammary Gland Biol Neoplasia* **2000**, *5* (4), 379-91.
96. Orazine, C. I.; Hincapie, M.; Hancock, W. S.; Hattersley, M.; Hanke, J. H., A proteomic analysis of the plasma glycoproteins of a MCF-7 mouse xenograft: a model system for the detection of tumor markers. *J Proteome Res* **2008**, *7* (4), 1542-54.
97. Hu, X.; Zhang, Y.; Zhang, A.; Li, Y.; Zhu, Z.; Shao, Z.; Zeng, R.; Xu, L. X., Comparative serum proteome analysis of human lymph node negative/positive invasive ductal carcinoma of the breast and benign breast disease controls via label-free semiquantitative shotgun technology. *OMICS* **2009**, *13* (4), 291-300.
98. Mathelin, C.; Cromer, A.; Wendling, C.; Tomasetto, C.; Rio, M. C., Serum biomarkers for detection of breast cancers: A prospective study. *Breast Cancer Res Treat* **2006**, *96* (1), 83-90.
99. Pietrowska, M.; Marczak, L.; Polanska, J.; Behrendt, K.; Nowicka, E.; Walaszczyk, A.; Chmura, A.; Deja, R.; Stobiecki, M.; Polanski, A.; Tarnawski, R.; Widlak, P., Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early stage breast cancer. *J Transl Med* **2009**, *7*, 60.
100. Schaub, N. P.; Jones, K. J.; Nyalwidhe, J. O.; Cazares, L. H.; Karbassi, I. D.; Semmes, O. J.; Feliberti, E. C.; Perry, R. R.; Drake, R. R., Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients. *J Am Coll Surg* **2009**, *208* (5), 970-8; discussion 978-80.
101. Tamesa, M. S.; Kuramitsu, Y.; Fujimoto, M.; Maeda, N.; Nagashima, Y.; Tanaka, T.; Yamamoto, S.; Oka, M.; Nakamura, K., Detection of autoantibodies against cyclophilin A and

triosephosphate isomerase in sera from breast cancer patients by proteomic analysis. *Electrophoresis* **2009**, *30* (12), 2168-81.

102. Zeidan, B. A.; Cutress, R. I.; Murray, N.; Coulton, G. R.; Hastie, C.; Packham, G.; Townsend, P. A., Proteomic analysis of archival breast cancer serum. *Cancer Genomics Proteomics* **2009**, *6* (3), 141-7.

103. Rompp, A.; Dekker, L.; Taban, I.; Jenster, G.; Boogerd, W.; Bonfrer, H.; Spengler, B.; Heeren, R.; Smitt, P. S.; Luider, T. M., Identification of leptomeningeal metastasis-related proteins in cerebrospinal fluid of patients with breast cancer by a combination of MALDI-TOF, MALDI-FTICR and nanoLC-FTICR MS. *Proteomics* **2007**, *7* (3), 474-81.

104. Coombes, K. R.; Fritsche, H. A., Jr.; Clarke, C.; Chen, J. N.; Baggerly, K. A.; Morris, J. S.; Xiao, L. C.; Hung, M. C.; Kuerer, H. M., Quality control and peak finding for proteomics data collected from nipple aspirate fluid by surface-enhanced laser desorption and ionization. *Clin Chem* **2003**, *49* (10), 1615-23.

105. Coombes, K. R.; Tsavachidis, S.; Morris, J. S.; Baggerly, K. A.; Hung, M. C.; Kuerer, H. M., Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics* **2005**, *5* (16), 4107-17.

106. Varnum, S. M.; Covington, C. C.; Woodbury, R. L.; Petritis, K.; Kangas, L. J.; Abdullah, M. S.; Pounds, J. G.; Smith, R. D.; Zangar, R. C., Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. *Breast Cancer Res Treat* **2003**, *80* (1), 87-97.

107. Alexander, H.; Stegner, A. L.; Wagner-Mann, C.; Du Bois, G. C.; Alexander, S.; Sauter, E. R., Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. *Clin Cancer Res* **2004**, *10* (22), 7500-10.

108. Pavlou, M. P.; Kulasingam, V.; Sauter, E. R.; Kliethermes, B.; Diamandis, E. P., Nipple aspirate fluid proteome of healthy females and patients with breast cancer. *Clin Chem* **2001**, *56* (5), 848-55.

109. Paweletz, C. P.; Trock, B.; Pennanen, M.; Tsangaris, T.; Magnant, C.; Liotta, L. A.; Petricoin, E. F., 3rd, Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. *Dis Markers* **2001**, *17* (4), 301-7.

110. Pawlik, T. M.; Fritsche, H.; Coombes, K. R.; Xiao, L.; Krishnamurthy, S.; Hunt, K. K.; Pusztai, L.; Chen, J. N.; Clarke, C. H.; Arun, B.; Hung, M. C.; Kuerer, H. M., Significant differences in nipple aspirate fluid protein expression between healthy women and those with

breast cancer demonstrated by time-of-flight mass spectrometry. *Breast Cancer Res Treat* **2005**, *89* (2), 149-57.

111. Pawlik, T. M.; Hawke, D. H.; Liu, Y.; Krishnamurthy, S.; Fritsche, H.; Hunt, K. K.; Kuerer, H. M., Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. *BMC Cancer* **2006**, *6*, 68.

112. Sauter, E. R.; Davis, W.; Qin, W.; Scanlon, S.; Mooney, B.; Bromert, K.; Folk, W. R., Identification of a beta-casein-like peptide in breast nipple aspirate fluid that is associated with breast cancer. *Biomark Med* **2009**, *3* (5), 577-88.

113. Sauter, E. R.; Shan, S.; Hewett, J. E.; Speckman, P.; Du Bois, G. C., Proteomic analysis of nipple aspirate fluid using SELDI-TOF-MS. *Int J Cancer* **2005**, *114* (5), 791-6.

114. Sauter, E. R.; Zhu, W.; Fan, X. J.; Wassell, R. P.; Chervoneva, I.; Du Bois, G. C., Proteomic analysis of nipple aspirate fluid to detect biologic markers of breast cancer. *Br J Cancer* **2002**, *86* (9), 1440-3.

115. Rower, C.; Vissers, J. P.; Koy, C.; Kipping, M.; Hecker, M.; Reimer, T.; Gerber, B.; Thiesen, H. J.; Glocker, M. O., Towards a proteome signature for invasive ductal breast carcinoma derived from label-free nanoscale LC-MS protein expression profiling of tumorous and glandular tissue. *Anal Bioanal Chem* **2009**, *395* (8), 2443-56.

116. Alldridge, L.; Metodieva, G.; Greenwood, C.; Al-Janabi, K.; Thwaites, L.; Sauven, P.; Metodiev, M., Proteome profiling of breast tumors by gel electrophoresis and nanoscale electrospray ionization mass spectrometry. *J Proteome Res* **2008**, *7* (4), 1458-69.

117. Cowherd, S. M.; Espina, V. A.; Petricoin, E. F., 3rd; Liotta, L. A., Proteomic analysis of human breast cancer tissue with laser-capture microdissection and reverse-phase protein microarrays. *Clin Breast Cancer* **2004**, *5* (5), 385-92.

118. Sanders, M. E.; Dias, E. C.; Xu, B. J.; Mobley, J. A.; Billheimer, D.; Roder, H.; Grigorieva, J.; Dowsett, M.; Arteaga, C. L.; Caprioli, R. M., Differentiating proteomic biomarkers in breast cancer by laser capture microdissection and MALDI MS. *J Proteome Res* **2008**, *7* (4), 1500-7.

119. Gromov, P.; Gromova, I.; Bunkenborg, J.; Cabezon, T.; Moreira, J. M.; Timmermans-Wielenga, V.; Roepstorff, P.; Rank, F.; Celis, J. E., Up-regulated proteins in the fluid bathing the tumour cell microenvironment as potential serological markers for early detection of cancer of the breast. *Mol Oncol* *4* (1), 65-89.

120. Cabezon, T.; Celis, J. E.; Skibshoj, I.; Klingelhofer, J.; Grigorian, M.; Gromov, P.; Rank, F.; Myklebust, J. H.; Maeldansmo, G. M.; Lukanidin, E.; Ambartsumian, N., Expression of S100A4 by a variety of cell types present in the tumor microenvironment of human breast cancer. *Int J Cancer* **2007**, *121* (7), 1433-44.
121. Celis, J. E.; Gromov, P.; Cabezon, T.; Moreira, J. M.; Ambartsumian, N.; Sandelin, K.; Rank, F.; Gromova, I., Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. *Mol Cell Proteomics* **2004**, *3* (4), 327-44.
122. Johann, D. J.; Rodriguez-Canales, J.; Mukherjee, S.; Prieto, D. A.; Hanson, J. C.; Emmert-Buck, M.; Blonder, J., Approaching solid tumor heterogeneity on a cellular basis by tissue proteomics using laser capture microdissection and biological mass spectrometry. *J Proteome Res* **2009**, *8* (5), 2310-8.
123. Hood, B. L.; Conrads, T. P.; Veenstra, T. D., Mass spectrometric analysis of formalin-fixed paraffin-embedded tissue: unlocking the proteome within. *Proteomics* **2006**, *6* (14), 4106-14.
124. Kiernan, J. A., Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: What they are and what they do. *Microscopy Today*. **2000**, *00-1*, 8-12
125. Kim, C.; Paik, S., Gene-expression-based prognostic assays for breast cancer. *Nat Rev Clin Oncol* **2010**, *7* (6), 340-7.
126. Gnanapragasam, V. J., Unlocking the molecular archive: the emerging use of formalin-fixed paraffin-embedded tissue for biomarker research in urological cancer. *BJU Int* **2010**, *105* (2), 274-8.
127. Shi, S. R.; Liu, C.; Balgley, B. M.; Lee, C.; Taylor, C. R., Protein extraction from formalin-fixed, paraffin-embedded tissue sections: quality evaluation by mass spectrometry. *J Histochem Cytochem* **2006**, *54* (6), 739-43.
128. Berg, D.; Hipp, S.; Malinowsky, K.; Bollner, C.; Becker, K. F., Molecular profiling of signalling pathways in formalin-fixed and paraffin-embedded cancer tissues. *Eur J Cancer* **2010**, *46* (1), 47-55.
129. Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Furusato, B.; Lucas, D. A.; Ringeisen, B. R.; Sesterhenn, I. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B., Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol Cell Proteomics* **2005**, *4* (11), 1741-53.
130. Hood, B. L.; Conrads, T. P.; Veenstra, T. D., Unravelling the proteome of formalin-fixed paraffin-embedded tissue. *Brief Funct Genomic Proteomic* **2006**, *5* (2), 169-75.

131. Becker, K. F.; Schott, C.; Hipp, S.; Metzger, V.; Porschewski, P.; Beck, R.; Nahrig, J.; Becker, I.; Hofler, H., Quantitative protein analysis from formalin-fixed tissues: implications for translational clinical research and nanoscale molecular diagnosis. *J Pathol* **2007**, *211* (3), 370-8.
132. Ronci, M.; Bonanno, E.; Colantoni, A.; Pieroni, L.; Di Ilio, C.; Spagnoli, L. G.; Federici, G.; Urbani, A., Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations. *Proteomics* **2008**, *8* (18), 3702-14.
133. Gast, M. C.; Schellens, J. H.; Beijnen, J. H., Clinical proteomics in breast cancer: a review. *Breast Cancer Res Treat* **2009**, *116* (1), 17-29.
134. Kulasingam, V.; Diamandis, E. P., Tissue culture-based breast cancer biomarker discovery platform. *Int J Cancer* **2008**, *123* (9), 2007-12.
135. Roche, S.; Gabelle, A.; Lehmann, S., Clinical proteomics of the cerebrospinal fluid: Towards the discovery of new biomarkers. *Proteomics Clinical Applications* **2008**, *2* (3), 428-436.
136. Shores, K. S.; Knapp, D. R., Assessment approach for evaluating high abundance protein depletion methods for cerebrospinal fluid (CSF) proteomic analysis. *J Proteome Res* **2007**, *6* (9), 3739-51.
137. Jacobs, J. M.; Waters, K. M.; Kathmann, L. E.; Camp, D. G., 2nd; Wiley, H. S.; Smith, R. D.; Thrall, B. D., The mammary epithelial cell secretome and its regulation by signal transduction pathways. *J Proteome Res* **2008**, *7* (2), 558-69.
138. Mbeunkui, F.; Metge, B. J.; Shevde, L. A.; Pannell, L. K., Identification of differentially secreted biomarkers using LC-MS/MS in isogenic cell lines representing a progression of breast cancer. *J Proteome Res* **2007**, *6* (8), 2993-3002.
139. Klein, P. M.; Lawrence, J. A., Lavage and nipple aspiration of breast ductal fluids: a source of biomarkers for environmental mutagenesis. *Environ Mol Mutagen* **2002**, *39* (2-3), 127-33.
140. Laronga, C.; Drake, R. R., Proteomic approach to breast cancer. *Cancer Control* **2007**, *14* (4), 360-8.
141. Petrakis, N. L., Physiologic, biochemical, and cytologic aspects of nipple aspirate fluid. *Breast Cancer Res Treat* **1986**, *8* (1), 7-19.
142. Das, D. K.; Al-Ayadhy, B.; Ajrawi, M. T.; Shaheen, A. A.; Sheikh, Z. A.; Malik, M.; Pathan, S. K.; Ebrahim, B.; Francis, I. M.; Satar, S. A.; Abdulla, M. A.; Luthra, U. K.; Junaid, T.

A., Cytodiagnosis of nipple discharge: a study of 602 samples from 484 cases. *Diagn Cytopathol* **2001**, 25 (1), 25-37.

143. Kurebayashi, J., [Biomarkers in breast cancer]. *Gan To Kagaku Ryoho* **2004**, 31 (7), 1021-6.

144. Noble, J.; Dua, R. S.; Locke, I.; Eeles, R.; Gui, G. P.; Isacke, C. M., Proteomic analysis of nipple aspirate fluid throughout the menstrual cycle in healthy pre-menopausal women. *Breast Cancer Res Treat* **2007**, 104 (2), 191-6.

145. Duffy, M. J., Urokinase plasminogen activator and its inhibitor, PAI-1, as prognostic markers in breast cancer: from pilot to level 1 evidence studies. *Clin Chem* **2002**, 48 (8), 1194-7.

146. Foekens, J. A.; Look, M. P.; Bolt-de Vries, J.; Meijer-van Gelder, M. E.; van Putten, W. L.; Klijn, J. G., Cathepsin-D in primary breast cancer: prognostic evaluation involving 2810 patients. *Br J Cancer* **1999**, 79 (2), 300-7.

147. Duffy, M. J., Serum tumor markers in breast cancer: are they of clinical value? *Clin Chem* **2006**, 52 (3), 345-51.

148. Nicolini, A.; Anselmi, L.; Michelassi, C.; Carpi, A., Prolonged survival by 'early' salvage treatment of breast cancer patients: a retrospective 6-year study. *Br J Cancer* **1997**, 76 (8), 1106-11.

149. Teng, P. N.; Rungruang, B. J.; Hood, B. L.; Sun, M.; Flint, M. S.; Bateman, N. W.; Dhir, R.; Bhargava, R.; Richard, S. D.; Edwards, R. P.; Conrads, T. P., Assessment of buffer systems for harvesting proteins from tissue interstitial fluid for proteomic analysis. *J Proteome Res* **2010**, 9 (8), 4161-9.

150. Gromov, P.; Gromova, I.; Bunkenborg, J.; Cabezon, T.; Moreira, J. M.; Timmermans-Wielenga, V.; Roepstorff, P.; Rank, F.; Celis, J. E., Up-regulated proteins in the fluid bathing the tumour cell microenvironment as potential serological markers for early detection of cancer of the breast. *Mol Oncol* **2010**, 4 (1), 65-89.

151. Saah, A. J.; Hoover, D. R., "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. *Ann Intern Med* **1997**, 126 (1), 91-4.

152. van 't Veer, L. J.; Dai, H.; van de Vijver, M. J.; He, Y. D.; Hart, A. A.; Mao, M.; Peterse, H. L.; van der Kooy, K.; Marton, M. J.; Witteveen, A. T.; Schreiber, G. J.; Kerkhoven, R. M.; Roberts, C.; Linsley, P. S.; Bernards, R.; Friend, S. H., Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **2002**, 415 (6871), 530-6.

153. Carey, L. A., Prognostic molecular profiles of breast cancer. *Uptodate*. Waltham, MA. **2010**.
154. Tsuda, H., Individualization of breast cancer based on histopathological features and molecular alterations. *Breast Cancer* **2008**, *15* (2), 121-32.
155. Drukier, A. K.; Grigoriev, I.; Brown, L. R.; Tomaszewski, J. E.; Sainsbury, R.; Godovac-Zimmermann, J., Looking for Thom's biomarkers with proteomics. *J Proteome Res* **2006**, *5* (8), 2046-8.
156. Chuang, H. Y.; Lee, E.; Liu, Y. T.; Lee, D.; Ideker, T., Network-based classification of breast cancer metastasis. *Mol Syst Biol* **2007**, *3*, 140.
157. Malik, R.; Dulla, K.; Nigg, E. A.; Korner, R., From proteome lists to biological impact--tools and strategies for the analysis of large MS data sets. *Proteomics* *10* (6), 1270-83.
158. Cline, M. S.; Smoot, M.; Cerami, E.; Kuchinsky, A.; Landys, N.; Workman, C.; Christmas, R.; Avila-Campilo, I.; Creech, M.; Gross, B.; Hanspers, K.; Isserlin, R.; Kelley, R.; Killcoyne, S.; Lotia, S.; Maere, S.; Morris, J.; Ono, K.; Pavlovic, V.; Pico, A. R.; Vailaya, A.; Wang, P. L.; Adler, A.; Conklin, B. R.; Hood, L.; Kuiper, M.; Sander, C.; Schmulevich, I.; Schwikowski, B.; Warner, G. J.; Ideker, T.; Bader, G. D., Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2007**, *2* (10), 2366-82.
159. Ideker, T.; Sharan, R., Protein networks in disease. *Genome Res* **2008**, *18* (4), 644-52.
160. Poisson, L. M.; Ghosh, D., Statistical Issues and Analyses of in vivo and in vitro Genomic Data in order to Identify Clinically Relevant Profiles. *Cancer Inform* **2007**, *3*, 231-43.
161. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W. H.; Pages, F.; Trajanoski, Z.; Galon, J., ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25* (8), 1091-3.
162. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **2000**, *25* (1), 25-9.
163. Dennis, G., Jr.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A., DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **2003**, *4* (5), P3.

164. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003**, *13* (11), 2498-504.
165. Gatza, M. L.; Lucas, J. E.; Barry, W. T.; Kim, J. W.; Wang, Q.; Crawford, M. D.; Datto, M. B.; Kelley, M.; Mathey-Prevot, B.; Potti, A.; Nevins, J. R., A pathway-based classification of human breast cancer. *Proc Natl Acad Sci U S A* *107* (15), 6994-9.
166. Bateman, N. W.; Sun, M.; Hood, B. L.; Flint, M. S.; Conrads, T. P., Defining central themes in breast cancer biology by differential proteomics: conserved regulation of cell spreading and focal adhesion kinase. *J Proteome Res* *9* (10), 5311-24.
167. Powell, S., Common threads in breast cancer proteomes. *J Proteome Res* *9* (10), 4875.
168. Bateman, N. W.; Sun, M.; Bhargava, R.; Hood, B. L.; Darfler, M. M.; Krizman, D. B.; Conrads, T. P., Differential Proteomic Analysis of Late-Stage and Recurrent Breast Cancer from Formalin-Fixed Paraffin-Embedded Tissues. *J Proteome Res* 2010. Accepted 12/2010. Manuscript ID # pr-2010-01073s.R1.
169. Charafe-Jauffret, E.; Ginestier, C.; Monville, F.; Finetti, P.; Adelaide, J.; Cervera, N.; Fekairi, S.; Xerri, L.; Jacquemier, J.; Birnbaum, D.; Bertucci, F., Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* **2006**, *25* (15), 2273-84.
170. Engel, L. W.; Young, N. A., Human breast carcinoma cells in continuous culture: a review. *Cancer Res* **1978**, *38* (11 Pt 2), 4327-39.
171. Lacroix, M.; Leclercq, G., Relevance of breast cancer cell lines as models for breast tumours: an update. *Breast Cancer Res Treat* **2004**, *83* (3), 249-89.
172. Neve, R. M.; Chin, K.; Fridlyand, J.; Yeh, J.; Baehner, F. L.; Fevr, T.; Clark, L.; Bayani, N.; Coppe, J. P.; Tong, F.; Speed, T.; Spellman, P. T.; DeVries, S.; Lapuk, A.; Wang, N. J.; Kuo, W. L.; Stilwell, J. L.; Pinkel, D.; Albertson, D. G.; Waldman, F. M.; McCormick, F.; Dickson, R. B.; Johnson, M. D.; Lippman, M.; Ethier, S.; Gazdar, A.; Gray, J. W., A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **2006**, *10* (6), 515-27.
173. Seibert, K.; Shafie, S. M.; Triche, T. J.; Whang-Peng, J. J.; O'Brien, S. J.; Toney, J. H.; Huff, K. K.; Lippman, M. E., Clonal variation of MCF-7 breast cancer cells in vitro and in athymic nude mice. *Cancer Res* **1983**, *43* (5), 2223-39.
174. Fogh, J.; Fogh, J. M.; Orfeo, T., One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. *J Natl Cancer Inst* **1977**, *59* (1), 221-6.

175. Soule, H. D.; Maloney, T. M.; Wolman, S. R.; Peterson, W. D., Jr.; Brenz, R.; McGrath, C. M.; Russo, J.; Pauley, R. J.; Jones, R. F.; Brooks, S. C., Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res* **1990**, *50* (18), 6075-86.
176. Konety, B. R.; Getzenberg, R. H., Nuclear structural proteins as biomarkers of cancer. *J Cell Biochem* **1999**, *Suppl 32-33*, 183-91.
177. Luo, M.; Guan, J. L., Focal adhesion kinase: a prominent determinant in breast cancer initiation, progression and metastasis. *Cancer Lett* **289** (2), 127-39.
178. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007**, *4* (3), 207-14.
179. Marengo, E.; Robotti, E.; Bobba, M.; Gosetti, F., The principle of exhaustiveness versus the principle of parsimony: a new approach for the identification of biomarkers from proteomic spot volume datasets based on principal component analysis. *Anal Bioanal Chem* **397** (1), 25-41.
180. Ferguson, R. E.; Carroll, H. P.; Harris, A.; Maher, E. R.; Selby, P. J.; Banks, R. E., Housekeeping proteins: a preliminary study illustrating some limitations as useful references in protein expression studies. *Proteomics* **2005**, *5* (2), 566-71.
181. McDonald, J. H., Handbook of Biological Statistics. *Sparky House Publishing, Baltimore, Maryland*. **2009**, (Data Transformations), 160-164.
182. Sommers, C. L.; Byers, S. W.; Thompson, E. W.; Torri, J. A.; Gelmann, E. P., Differentiation state and invasiveness of human breast cancer cell lines. *Breast Cancer Res Treat* **1994**, *31* (2-3), 325-35.
183. Cuvelier, D.; Thery, M.; Chu, Y. S.; Dufour, S.; Thiery, J. P.; Bornens, M.; Nassoy, P.; Mahadevan, L., The universal dynamics of cell spreading. *Curr Biol* **2007**, *17* (8), 694-9.
184. Gilcrease, M. Z., Integrin signaling in epithelial cells. *Cancer Lett* **2007**, *247* (1), 1-25.
185. Goldfinger, L. E.; Stack, M. S.; Jones, J. C., Processing of laminin-5 and its functional consequences: role of plasmin and tissue-type plasminogen activator. *J Cell Biol* **1998**, *141* (1), 255-65.
186. Stahl, S.; Weitzman, S.; Jones, J. C., The role of laminin-5 and its receptors in mammary epithelial cell branching morphogenesis. *J Cell Sci* **1997**, *110* ( Pt 1), 55-63.

187. Martin, S. S.; Leder, P., Human MCF10A mammary epithelial cells undergo apoptosis following actin depolymerization that is independent of attachment and rescued by Bcl-2. *Mol Cell Biol* **2001**, *21* (19), 6529-36.
188. Tait, L.; Soule, H. D.; Russo, J., Ultrastructural and immunocytochemical characterization of an immortalized human breast epithelial cell line, MCF-10. *Cancer Res* **1990**, *50* (18), 6087-94.
189. Wang, H. B.; Dembo, M.; Wang, Y. L., Substrate flexibility regulates growth and apoptosis of normal but not transformed cells. *Am J Physiol Cell Physiol* **2000**, *279* (5), C1345-50.
190. Bambang, I. F.; Lu, D.; Li, H.; Chiu, L. L.; Lau, Q. C.; Koay, E.; Zhang, D., Cytokeratin 19 regulates endoplasmic reticulum stress and inhibits ERp29 expression via p38 MAPK/XBP-1 signaling in breast cancer cells. *Exp Cell Res* **2009**, *315* (11), 1964-74.
191. Nerlich, A. G.; Lebeau, A.; Hagedorn, H. G.; Sauer, U.; Schleicher, E. D., Morphological aspects of altered basement membrane metabolism in invasive carcinomas of the breast and the larynx. *Anticancer Res* **1998**, *18* (5A), 3515-20.
192. Prince, J. M.; Klinowska, T. C.; Marshman, E.; Lowe, E. T.; Mayer, U.; Miner, J.; Aberdam, D.; Vestweber, D.; Gusterson, B.; Streuli, C. H., Cell-matrix interactions during development and apoptosis of the mouse mammary gland in vivo. *Dev Dyn* **2002**, *223* (4), 497-516.
193. Valladares, A.; Hernandez, N. G.; Gomez, F. S.; Curiel-Quezada, E.; Madrigal-Bujaidar, E.; Vergara, M. D.; Martinez, M. S.; Arenas Aranda, D. J., Genetic expression profiles and chromosomal alterations in sporadic breast cancer in Mexican women. *Cancer Genet Cytogenet* **2006**, *170* (2), 147-51.
194. Colorado, P. C.; Torre, A.; Kamphaus, G.; Maeshima, Y.; Hopfer, H.; Takahashi, K.; Volk, R.; Zamborsky, E. D.; Herman, S.; Sarkar, P. K.; Ericksen, M. B.; Dhanabal, M.; Simons, M.; Post, M.; Kufe, D. W.; Weichselbaum, R. R.; Sukhatme, V. P.; Kalluri, R., Anti-angiogenic cues from vascular basement membrane collagen. *Cancer Res* **2000**, *60* (9), 2520-6.
195. Long, M. Y.; Li, H. H.; Xu, J. Y.; Lai, D. M.; Weng, Z. H., [Inhibitory effects of transfection of arresten gene on liver metastasis from colorectal cancer in nude mice]. *Ai Zheng* **2008**, *27* (10), 1039-43.
196. Tandle, A.; Blazer, D. G., 3rd; Libutti, S. K., Antiangiogenic gene therapy of cancer: recent developments. *J Transl Med* **2004**, *2* (1), 22.

197. Kamphaus, G. D.; Colorado, P. C.; Panka, D. J.; Hopfer, H.; Ramchandran, R.; Torre, A.; Maeshima, Y.; Mier, J. W.; Sukhatme, V. P.; Kalluri, R., Canstatin, a novel matrix-derived inhibitor of angiogenesis and tumor growth. *J Biol Chem* **2000**, *275* (2), 1209-15.
198. Bax, D. V.; Mahalingam, Y.; Cain, S.; Mellody, K.; Freeman, L.; Younger, K.; Shuttleworth, C. A.; Humphries, M. J.; Couchman, J. R.; Kielty, C. M., Cell adhesion to fibrillin-1: identification of an Arg-Gly-Asp-dependent synergy region and a heparin-binding site that regulates focal adhesion formation. *J Cell Sci* **2007**, *120* (Pt 8), 1383-92.
199. Golden, H. B.; Watson, L. E.; Lal, H.; Verma, S. K.; Foster, D. M.; Kuo, S. R.; Sharma, A.; Frankel, A.; Dostal, D. E., Anthrax toxin: pathologic effects on the cardiovascular system. *Front Biosci* **2009**, *14*, 2335-57.
200. Hauzenberger, D.; Olivier, P.; Gundersen, D.; Ruegg, C., Tenascin-C inhibits beta1 integrin-dependent T lymphocyte adhesion to fibronectin through the binding of its fnIII 1-5 repeats to fibronectin. *Eur J Immunol* **1999**, *29* (5), 1435-47.
201. Horiuchi, K.; Amizuka, N.; Takeshita, S.; Takamatsu, H.; Katsuura, M.; Ozawa, H.; Toyama, Y.; Bonewald, L. F.; Kudo, A., Identification and characterization of a novel protein, periostin, with restricted expression to periosteum and periodontal ligament and increased expression by transforming growth factor beta. *J Bone Miner Res* **1999**, *14* (7), 1239-49.
202. Jin, L.; Kern, M. J.; Otey, C. A.; Wamhoff, B. R.; Somlyo, A. V., Angiotensin II, focal adhesion kinase, and PRX1 enhance smooth muscle expression of lipoma preferred partner and its newly identified binding partner palladin to promote cell migration. *Circ Res* **2007**, *100* (6), 817-25.
203. Kim, J. E.; Kim, S. J.; Lee, B. H.; Park, R. W.; Kim, K. S.; Kim, I. S., Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-beta-induced gene, betaig-h3. *J Biol Chem* **2000**, *275* (40), 30907-15.
204. Lange, K.; Kammerer, M.; Saupe, F.; Hegi, M. E.; Grotegut, S.; Fluri, E.; Orend, G., Combined lysophosphatidic acid/platelet-derived growth factor signaling triggers glioma cell migration in a tenascin-C microenvironment. *Cancer Res* **2008**, *68* (17), 6942-52.
205. Liu, X. S.; Luo, H. J.; Yang, H.; Wang, L.; Kong, H.; Jin, Y. E.; Wang, F.; Gu, M. M.; Chen, Z.; Lu, Z. Y.; Wang, Z. G., Palladin regulates cell and extracellular matrix interaction through maintaining normal actin cytoskeleton architecture and stabilizing beta1-integrin. *J Cell Biochem* **2007**, *100* (5), 1288-300.
206. Nishiya, N.; Tachibana, K.; Shibamura, M.; Mashimo, J. I.; Nose, K., Hic-5-reduced cell spreading on fibronectin: competitive effects between paxillin and Hic-5 through interaction with focal adhesion kinase. *Mol Cell Biol* **2001**, *21* (16), 5332-45.

207. Popova, S. N.; Barczyk, M.; Tiger, C. F.; Beertsen, W.; Zigrino, P.; Aszodi, A.; Miosge, N.; Forsberg, E.; Gullberg, D., Alpha11 beta1 integrin-dependent regulation of periodontal ligament function in the erupting mouse incisor. *Mol Cell Biol* **2007**, *27* (12), 4306-16.
208. Prieto, A. L.; Edelman, G. M.; Crossin, K. L., Multiple integrins mediate cell attachment to cytotactin/tenascin. *Proc Natl Acad Sci U S A* **1993**, *90* (21), 10154-8.
209. Summers, L.; Kielty, C.; Pinteaux, E., Adhesion to fibronectin regulates interleukin-1 beta expression in microglial cells. *Mol Cell Neurosci* **2009**, *41* (2), 148-55.
210. Werner, E.; Kowalczyk, A. P.; Faundez, V., Anthrax toxin receptor 1/tumor endothelium marker 8 mediates cell spreading by coupling extracellular ligands to the actin cytoskeleton. *J Biol Chem* **2006**, *281* (32), 23227-36.
211. Yayon, A.; Klagsbrun, M., Autocrine transformation by chimeric signal peptide-basic fibroblast growth factor: reversal by suramin. *Proc Natl Acad Sci U S A* **1990**, *87* (14), 5346-50.
212. Zagzag, D.; Shiff, B.; Jallo, G. I.; Greco, M. A.; Blanco, C.; Cohen, H.; Hukin, J.; Allen, J. C.; Friedlander, D. R., Tenascin-C promotes microvascular cell migration and phosphorylation of focal adhesion kinase. *Cancer Res* **2002**, *62* (9), 2660-8.
213. Zhang, M.; Liu, J.; Cheng, A.; Deyoung, S. M.; Saltiel, A. R., Identification of CAP as a costameric protein that interacts with filamin C. *Mol Biol Cell* **2007**, *18* (12), 4731-40.
214. Calaf, G. M.; Echiburru-Chau, C.; Zhao, Y. L.; Hei, T. K., BigH3 protein expression as a marker for breast cancer. *Int J Mol Med* **2008**, *21* (5), 561-8.
215. Korah, R.; Das, K.; Lindy, M. E.; Hameed, M.; Wieder, R., Coordinate loss of fibroblast growth factor 2 and laminin 5 expression during neoplastic progression of mammary duct epithelium. *Hum Pathol* **2007**, *38* (1), 154-60.
216. Murabito, J. M.; Rosenberg, C. L.; Finger, D.; Kreger, B. E.; Levy, D.; Splansky, G. L.; Antman, K.; Hwang, S. J., A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med Genet* **2007**, *8 Suppl 1*, S6.
217. Davies, G.; Rmali, K. A.; Watkins, G.; Mansel, R. E.; Mason, M. D.; Jiang, W. G., Elevated levels of tumour endothelial marker-8 in human breast cancer and its clinical significance. *Int J Oncol* **2006**, *29* (5), 1311-7.
218. Ruan, K.; Bao, S.; Ouyang, G., The multifaceted role of periostin in tumorigenesis. *Cell Mol Life Sci* **2009**, *66* (14), 2219-30.

219. Zhang, Y.; Zhang, G.; Li, J.; Tao, Q.; Tang, W., The expression analysis of periostin in human breast cancer. *J Surg Res* **160** (1), 102-6.
220. Heitzer, M. D.; DeFranco, D. B., Mechanism of action of Hic-5/androgen receptor activator 55, a LIM domain-containing nuclear receptor coactivator. *Mol Endocrinol* **2006**, *20* (1), 56-64.
221. Tumbarello, D. A.; Turner, C. E., Hic-5 contributes to epithelial-mesenchymal transformation through a RhoA/ROCK-dependent pathway. *J Cell Physiol* **2007**, *211* (3), 736-47.
222. Hancox, R. A.; Allen, M. D.; Holliday, D. L.; Edwards, D. R.; Pennington, C. J.; Guttery, D. S.; Shaw, J. A.; Walker, R. A.; Pringle, J. H.; Jones, J. L., Tumour-associated tenascin-C isoforms promote breast cancer cell invasion and growth by matrix metalloproteinase-dependent and independent mechanisms. *Breast Cancer Res* **2009**, *11* (2), R24.
223. Midwood, K. S.; Orend, G., The role of tenascin-C in tissue injury and tumorigenesis. *J Cell Commun Signal* **2009**.
224. Flaxman, B. A.; Van Scott, E. J., Growth of normal human mammary gland epithelium in vitro. *Cancer Res* **1972**, *32* (11), 2407-12.
225. Kaneda, A.; Kaminishi, M.; Yanagihara, K.; Sugimura, T.; Ushijima, T., Identification of silencing of nine genes in human gastric cancers. *Cancer Res* **2002**, *62* (22), 6645-50.
226. Pogue-Geile, K. L.; Chen, R.; Bronner, M. P.; Crnogorac-Jurcevic, T.; Moyes, K. W.; Downen, S.; Otey, C. A.; Crispin, D. A.; George, R. D.; Whitcomb, D. C.; Brentnall, T. A., Palladin mutation causes familial pancreatic cancer and suggests a new cancer mechanism. *PLoS Med* **2006**, *3* (12), e516.
227. Pirone, D. M.; Liu, W. F.; Ruiz, S. A.; Gao, L.; Raghavan, S.; Lemmon, C. A.; Romer, L. H.; Chen, C. S., An inhibitory role for FAK in regulating proliferation: a link between limited adhesion and RhoA-ROCK signaling. *J Cell Biol* **2006**, *174* (2), 277-88.
228. Lahlou, H.; Sanguin-Gendreau, V.; Zuo, D.; Cardiff, R. D.; McLean, G. W.; Frame, M. C.; Muller, W. J., Mammary epithelial-specific disruption of the focal adhesion kinase blocks mammary tumor progression. *Proc Natl Acad Sci U S A* **2007**, *104* (51), 20302-7.
229. Wendt, M. K.; Schiemann, W. P., Therapeutic targeting of the focal adhesion complex prevents oncogenic TGF-beta signaling and metastasis. *Breast Cancer Res* **2009**, *11* (5), R68.
230. Streuli, C. H.; Edwards, G. M., Control of normal mammary epithelial phenotype by integrins. *J Mammary Gland Biol Neoplasia* **1998**, *3* (2), 151-63.

231. Boudreau, N. J.; Jones, P. L., Extracellular matrix and integrin signalling: the shape of things to come. *Biochem J* **1999**, *339* ( Pt 3), 481-8.
232. Pylayeva, Y.; Gillen, K. M.; Gerald, W.; Beggs, H. E.; Reichardt, L. F.; Giancotti, F. G., Ras- and PI3K-dependent breast tumorigenesis in mice and humans requires focal adhesion kinase signaling. *J Clin Invest* **2009**, *119* (2), 252-66.
233. Kobayashi, H.; Azumi, M.; Kimura, Y.; Sato, K.; Aoki, N.; Kimura, S.; Honma, M.; Iizuka, H.; Tateno, M.; Celis, E., Focal adhesion kinase as an immunotherapeutic target. *Cancer Immunol Immunother* **2009**, *58* (6), 931-40.
234. Sawhney, R. S.; Liu, W.; Brattain, M. G., A novel role of ERK5 in integrin-mediated cell adhesion and motility in cancer cells via Fak signaling. *J Cell Physiol* **2009**, *219* (1), 152-61.
235. Panka, D. J.; Mier, J. W., Canstatin inhibits Akt activation and induces Fas-dependent apoptosis in endothelial cells. *J Biol Chem* **2003**, *278* (39), 37632-6.
236. Sumitomo, M.; Shen, R.; Walburg, M.; Dai, J.; Geng, Y.; Navarro, D.; Boileau, G.; Papandreou, C. N.; Giancotti, F. G.; Knudsen, B.; Nanus, D. M., Neutral endopeptidase inhibits prostate cancer cell migration by blocking focal adhesion kinase signaling. *J Clin Invest* **2000**, *106* (11), 1399-407.
237. Wu, Y.; Chen, L.; Zheng, P. S.; Yang, B. B., beta 1-Integrin-mediated glioma cell adhesion and free radical-induced apoptosis are regulated by binding to a C-terminal domain of PG-M/versican. *J Biol Chem* **2002**, *277* (14), 12294-301.
238. Smollich, M.; Gotte, M.; Yip, G. W.; Yong, E. S.; Kersting, C.; Fischgrabe, J.; Radke, I.; Kiesel, L.; Wulfig, P., On the role of endothelin-converting enzyme-1 (ECE-1) and neprilysin in human breast cancer. *Breast Cancer Res Treat* **2007**, *106* (3), 361-9.
239. Iwaya, K.; Ogawa, H.; Izumi, M.; Kuroda, M.; Mukai, K., Stromal expression of CD10 in invasive breast carcinoma: a new predictor of clinical outcome. *Virchows Arch* **2002**, *440* (6), 589-93.
240. Alix-Panabieres, C.; Vendrell, J. P.; Slijper, M.; Pelle, O.; Barbotte, E.; Mercier, G.; Jacot, W.; Fabbro, M.; Pantel, K., Full-length cytokeratin-19 is released by human tumor cells: a potential role in metastatic progression of breast cancer. *Breast Cancer Res* **2009**, *11* (3), R39.
241. Suwihat, S.; Ricciardelli, C.; Tammi, R.; Tammi, M.; Auvinen, P.; Kosma, V. M.; LeBaron, R. G.; Raymond, W. A.; Tilley, W. D.; Horsfall, D. J., Expression of extracellular matrix components versican, chondroitin sulfate, tenascin, and hyaluronan, and their association with disease outcome in node-negative breast cancer. *Clin Cancer Res* **2004**, *10* (7), 2491-8.

242. Bartel, D. P., MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **2004**, *116* (2), 281-97.
243. Ding, L.; Han, M., GW182 family proteins are crucial for microRNA-mediated gene silencing. *Trends Cell Biol* **2007**, *17* (8), 411-6.
244. Filipowicz, W.; Bhattacharyya, S. N.; Sonenberg, N., Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **2008**, *9* (2), 102-14.
245. Rana, T. M., Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol* **2007**, *8* (1), 23-36.
246. Meltzer, P. S., Cancer genomics: small RNAs with big impacts. *Nature* **2005**, *435* (7043), 745-6.
247. Esquela-Kerscher, A.; Slack, F. J., Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **2006**, *6* (4), 259-69.
248. Sempere, L. F.; Christensen, M.; Silahdaroglu, A.; Bak, M.; Heath, C. V.; Schwartz, G.; Wells, W.; Kauppinen, S.; Cole, C. N., Altered MicroRNA expression confined to specific epithelial cell subpopulations in breast cancer. *Cancer Res* **2007**, *67* (24), 11612-20.
249. Blenkiron, C.; Goldstein, L. D.; Thorne, N. P.; Spiteri, I.; Chin, S. F.; Dunning, M. J.; Barbosa-Morais, N. L.; Teschendorff, A. E.; Green, A. R.; Ellis, I. O.; Tavare, S.; Caldas, C.; Miska, E. A., MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol* **2007**, *8* (10), R214.
250. Wang, V.; Wu, W., MicroRNA: A New Player in Breast Cancer Development. *Journal of Cancer Molecules* **2007**, *3* (5), 133-138.
251. Sachdeva, M.; Zhu, S.; Wu, F.; Wu, H.; Walia, V.; Kumar, S.; Elble, R.; Watabe, K.; Mo, Y. Y., p53 represses c-Myc through induction of the tumor suppressor miR-145. *Proc Natl Acad Sci U S A* **2009**, *106* (9), 3207-12.
252. Wang, S.; Bian, C.; Yang, Z.; Bo, Y.; Li, J.; Zeng, L.; Zhou, H.; Zhao, R. C., miR-145 inhibits breast cancer cell growth through RTKN. *Int J Oncol* **2009**, *34* (5), 1461-6.
253. Spizzo, R.; Nicoloso, M. S.; Lupini, L.; Lu, Y.; Fogarty, J.; Rossi, S.; Zagatti, B.; Fabbri, M.; Veronese, A.; Liu, X.; Davuluri, R.; Croce, C. M.; Mills, G.; Negrini, M.; Calin, G. A., miR-145 participates with TP53 in a death-promoting regulatory loop and targets estrogen receptor-alpha in human breast cancer cells. *Cell Death Differ* **17** (2), 246-54.

254. Gotte, M.; Mohr, C.; Koo, C. Y.; Stock, C.; Vaske, A. K.; Viola, M.; Ibrahim, S. A.; Peddibhotla, S.; Teng, Y. H.; Low, J. Y.; Ebnet, K.; Kiesel, L.; Yip, G. W., miR-145-dependent targeting of Junctional Adhesion Molecule A and modulation of fascin expression are associated with reduced breast cancer cell motility and invasiveness. *Oncogene*.
255. Sachdeva, M.; Mo, Y. Y., MicroRNA-145 suppresses cell invasion and metastasis by directly targeting mucin 1. *Cancer Res* **70** (1), 378-87.
256. Bentwich, I., Prediction and validation of microRNAs and their targets. *FEBS Lett* **2005**, *579* (26), 5904-10.
257. Grimson, A.; Farh, K. K.; Johnston, W. K.; Garrett-Engle, P.; Lim, L. P.; Bartel, D. P., MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **2007**, *27* (1), 91-105.
258. Enright, A. J.; John, B.; Gaul, U.; Tuschl, T.; Sander, C.; Marks, D. S., MicroRNA targets in *Drosophila*. *Genome Biol* **2003**, *5* (1), R1.
259. Kertesz, M.; Iovino, N.; Unnerstall, U.; Gaul, U.; Segal, E., The role of site accessibility in microRNA target recognition. *Nat Genet* **2007**, *39* (10), 1278-84.
260. Lewis, B. P.; Shih, I. H.; Jones-Rhoades, M. W.; Bartel, D. P.; Burge, C. B., Prediction of mammalian microRNA targets. *Cell* **2003**, *115* (7), 787-98.
261. Vinther, J.; Hedegaard, M. M.; Gardner, P. P.; Andersen, J. S.; Arctander, P., Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucleic Acids Res* **2006**, *34* (16), e107.
262. Chen, C. Z.; Li, L.; Lodish, H. F.; Bartel, D. P., MicroRNAs modulate hematopoietic lineage differentiation. *Science* **2004**, *303* (5654), 83-6.
263. Zeng, Y.; Cullen, B. R., Recognition and cleavage of primary microRNA transcripts. *Methods Mol Biol* **2006**, *342*, 49-56.
264. Bateman, N. W.; Tan, D.; Pestell, R. G.; Black, J. D.; Black, A. R., Intestinal tumor progression is associated with altered function of KLF5. *J Biol Chem* **2004**, *279* (13), 12093-101.
265. Lam, D. C.; Girard, L.; Ramirez, R.; Chau, W. S.; Suen, W. S.; Sheridan, S.; Tin, V. P.; Chung, L. P.; Wong, M. P.; Shay, J. W.; Gazdar, A. F.; Lam, W. K.; Minna, J. D., Expression of nicotinic acetylcholine receptor subunit genes in non-small-cell lung cancer reveals differences between smokers and nonsmokers. *Cancer Res* **2007**, *67* (10), 4638-47.

266. Foster, L. J.; De Hoog, C. L.; Mann, M., Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc Natl Acad Sci U S A* **2003**, *100* (10), 5813-8.
267. Vecchione, A.; Cooper, H. J.; Trim, K. J.; Akbarzadeh, S.; Heath, J. K.; Wheldon, L. M., Protein partners in the life history of activated fibroblast growth factor receptors. *Proteomics* **2007**, *7* (24), 4565-78.
268. Attwell, S.; Roskelley, C.; Dedhar, S., The integrin-linked kinase (ILK) suppresses anoikis. *Oncogene* **2000**, *19* (33), 3811-5.
269. Gallagher, S. M.; Castorino, J. J.; Wang, D.; Philp, N. J., Monocarboxylate transporter 4 regulates maturation and trafficking of CD147 to the plasma membrane in the metastatic breast cancer cell line MDA-MB-231. *Cancer Res* **2007**, *67* (9), 4182-9.
270. Persad, S.; Dedhar, S., The role of integrin-linked kinase (ILK) in cancer progression. *Cancer Metastasis Rev* **2003**, *22* (4), 375-84.
271. Wilson, K. S.; Roberts, H.; Leek, R.; Harris, A. L.; Geradts, J., Differential gene expression patterns in HER2/neu-positive and -negative breast cancer cell lines and tissues. *Am J Pathol* **2002**, *161* (4), 1171-85.
272. Yoganathan, N.; Yee, A.; Zhang, Z.; Leung, D.; Yan, J.; Fazli, L.; Kojic, D. L.; Costello, P. C.; Jabali, M.; Dedhar, S.; Sanghera, J., Integrin-linked kinase, a promising cancer therapeutic target: biochemical and biological properties. *Pharmacol Ther* **2002**, *93* (2-3), 233-42.
273. Shi, B.; Sepp-Lorenzino, L.; Prisco, M.; Linsley, P.; deAngelis, T.; Baserga, R., Micro RNA 145 targets the insulin receptor substrate-1 and inhibits the growth of colon cancer cells. *J Biol Chem* **2007**, *282* (45), 32582-90.
274. Patel, V.; Hood, B. L.; Molinolo, A. A.; Lee, N. H.; Conrads, T. P.; Braisted, J. C.; Krizman, D. B.; Veenstra, T. D.; Gutkind, J. S., Proteomic analysis of laser-captured paraffin-embedded tissues: a molecular portrait of head and neck cancer progression. *Clin Cancer Res* **2008**, *14* (4), 1002-14.
275. Maughan, K. L.; Lutterbie, M. A.; Ham, P. S., Treatment of breast cancer. *Am Fam Physician* **2010**, *81* (11), 1339-46.
276. Carey, L. A.; Metzger, R.; Dees, E. C.; Collichio, F.; Sartor, C. I.; Ollila, D. W.; Klauber-DeMore, N.; Halle, J.; Sawyer, L.; Moore, D. T.; Graham, M. L., American Joint Committee on Cancer tumor-node-metastasis stage after neoadjuvant chemotherapy and breast cancer outcome. *J Natl Cancer Inst* **2005**, *97* (15), 1137-42.

277. Amatschek, S.; Koenig, U.; Auer, H.; Steinlein, P.; Pacher, M.; Gruenfelder, A.; Dekan, G.; Vogl, S.; Kubista, E.; Heider, K. H.; Stratowa, C.; Schreiber, M.; Sommergruber, W., Tissue-wide expression profiling using cDNA subtraction and microarrays to identify tumor-specific genes. *Cancer Res* **2004**, *64* (3), 844-56.
278. Levental, K. R.; Yu, H.; Kass, L.; Lakins, J. N.; Egeblad, M.; Erler, J. T.; Fong, S. F.; Csiszar, K.; Giaccia, A.; Weninger, W.; Yamauchi, M.; Gasser, D. L.; Weaver, V. M., Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **2009**, *139* (5), 891-906.
279. Martin, L. J.; Boyd, N. F., Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence. *Breast Cancer Res* **2008**, *10* (1), 201.
280. Schuetz, C. S.; Bonin, M.; Clare, S. E.; Nieselt, K.; Sotlar, K.; Walter, M.; Fehm, T.; Solomayer, E.; Riess, O.; Wallwiener, D.; Kurek, R.; Neubauer, H. J., Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer Res* **2006**, *66* (10), 5278-86.
281. Pereira, M. B.; Dias, A. J.; Reis, C. A.; Schmitt, F. C., Immunohistochemical study of the expression of MUC5AC and MUC6 in breast carcinomas and adjacent breast tissues. *J Clin Pathol* **2001**, *54* (3), 210-3.
282. Rakha, E. A.; Boyce, R. W.; Abd El-Rehim, D.; Kurien, T.; Green, A. R.; Paish, E. C.; Robertson, J. F.; Ellis, I. O., Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer. *Mod Pathol* **2005**, *18* (10), 1295-304.
283. Rice, A.; Quinn, C. M., Angiogenesis, thrombospondin, and ductal carcinoma in situ of the breast. *J Clin Pathol* **2002**, *55* (8), 569-74.
284. Turashvili, G.; Bouchal, J.; Burkadze, G.; Kolar, Z., Differentiation of tumours of ductal and lobular origin: I. Proteomics of invasive ductal and lobular breast carcinomas. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* **2005**, *149* (1), 57-62.
285. Locopo, N.; Fanelli, M.; Gasparini, G., Clinical significance of angiogenic factors in breast cancer. *Breast Cancer Res Treat* **1998**, *52* (1-3), 159-73.
286. Rice, A. J.; Steward, M. A.; Quinn, C. M., Thrombospondin 1 protein expression relates to good prognostic indices in ductal carcinoma in situ of the breast. *J Clin Pathol* **2002**, *55* (12), 921-5.

287. Wong, S. Y.; Purdie, A. T.; Han, P., Thrombospondin and other possible related matrix proteins in malignant and benign breast disease. An immunohistochemical study. *Am J Pathol* **1992**, *140* (6), 1473-82.
288. Cowin, P.; Rowlands, T. M.; Hatsell, S. J., Cadherins and catenins in breast cancer. *Curr Opin Cell Biol* **2005**, *17* (5), 499-508.
289. Heimann, R.; Lan, F.; McBride, R.; Hellman, S., Separating favorable from unfavorable prognostic markers in breast cancer: the role of E-cadherin. *Cancer Res* **2000**, *60* (2), 298-304.
290. Rimm, D. L.; Sinard, J. H.; Morrow, J. S., Reduced alpha-catenin and E-cadherin expression in breast cancer. *Lab Invest* **1995**, *72* (5), 506-12.
291. De Leeuw, W. J.; Berx, G.; Vos, C. B.; Peterse, J. L.; Van de Vijver, M. J.; Litvinov, S.; Van Roy, F.; Cornelisse, C. J.; Cleton-Jansen, A. M., Simultaneous loss of E-cadherin and catenins in invasive lobular breast cancer and lobular carcinoma in situ. *J Pathol* **1997**, *183* (4), 404-11.
292. Gaspar, C.; Franken, P.; Molenaar, L.; Breukel, C.; van der Valk, M.; Smits, R.; Fodde, R., A targeted constitutive mutation in the APC tumor suppressor gene underlies mammary but not intestinal tumorigenesis. *PLoS Genet* **2009**, *5* (7), e1000547.
293. Wang, X.; Goode, E. L.; Fredericksen, Z. S.; Vierkant, R. A.; Pankratz, V. S.; Liu-Mares, W.; Rider, D. N.; Vachon, C. M.; Cerhan, J. R.; Olson, J. E.; Couch, F. J., Association of genetic variation in genes implicated in the beta-catenin destruction complex with risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* **2008**, *17* (8), 2101-8.
294. Prasad, C. P.; Mirza, S.; Sharma, G.; Prasad, R.; DattaGupta, S.; Rath, G.; Ralhan, R., Epigenetic alterations of CDH1 and APC genes: relationship with activation of Wnt/beta-catenin pathway in invasive ductal carcinoma of breast. *Life Sci* **2008**, *83* (9-10), 318-25.
295. Kruser, T. J.; Wheeler, D. L., Mechanisms of resistance to HER family targeting antibodies. *Exp Cell Res* **2010**, *316* (7), 1083-100.
296. Goldfarb, Y.; Ben-Eliyahu, S., Surgery as a risk factor for breast cancer recurrence and metastasis: mediating mechanisms and clinical prophylactic approaches. *Breast Dis* **2006**, *26*, 99-114.
297. Smith, M. J.; Culhane, A. C.; Killeen, S.; Kelly, M. A.; Wang, J. H.; Cotter, T. G.; Redmond, H. P., Mechanisms driving local breast cancer recurrence in a model of breast-conserving surgery. *Ann Surg Oncol* **2008**, *15* (10), 2954-64.

298. Kim, D. H.; Bae, J.; Lee, J. W.; Kim, S.; Kim, Y.; Bae, J.; Yi, J. K.; Yu, M.; Noh, D. Y.; Lee, C., Proteomic analysis of breast cancer tissue reveals upregulation of actin-remodeling proteins and its relevance to cancer invasiveness. *Proteomics - Clinical Applications* **2009**, *3* (1), 30-40.
299. Wagner, C.; Koury, M. J., S-Adenosylhomocysteine: a better indicator of vascular disease than homocysteine? *Am J Clin Nutr* **2007**, *86* (6), 1581-5.
300. Lukyanova, N. Y., Characteristics of homocysteine-induced multidrug resistance of human MCF-7 breast cancer cells and human A2780 ovarian cancer cells. *Exp Oncol* **32** (1), 10-4.
301. Kahle, P. J.; Waak, J.; Gasser, T., DJ-1 and prevention of oxidative stress in Parkinson's disease and other age-related disorders. *Free Radic Biol Med* **2009**, *47* (10), 1354-61.
302. Kim, R. H.; Peters, M.; Jang, Y.; Shi, W.; Pintilie, M.; Fletcher, G. C.; DeLuca, C.; Liepa, J.; Zhou, L.; Snow, B.; Binari, R. C.; Manoukian, A. S.; Bray, M. R.; Liu, F. F.; Tsao, M. S.; Mak, T. W., DJ-1, a novel regulator of the tumor suppressor PTEN. *Cancer Cell* **2005**, *7* (3), 263-73.
303. Le Naour, F.; Misek, D. E.; Krause, M. C.; Deneux, L.; Giordano, T. J.; Scholl, S.; Hanash, S. M., Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer. *Clin Cancer Res* **2001**, *7* (11), 3328-35.
304. Pucci-Minafra, I.; Cancemi, P.; Albanese, N. N.; Di Cara, G.; Marabeti, M. R.; Marrazzo, A.; Minafra, S., New protein clustering of breast cancer tissue proteomics using actin content as a cellularity indicator. *J Proteome Res* **2008**, *7* (4), 1412-8.
305. Gibson, S. L.; Ma, Z.; Shaw, L. M., Divergent roles for IRS-1 and IRS-2 in breast cancer metastasis. *Cell Cycle* **2007**, *6* (6), 631-7.
306. Chan, B. T.; Lee, A. V., Insulin receptor substrates (IRSs) and breast tumorigenesis. *J Mammary Gland Biol Neoplasia* **2008**, *13* (4), 415-22.
307. Rocha, R. L.; Hilsenbeck, S. G.; Jackson, J. G.; VanDenBerg, C. L.; Weng, C.; Lee, A. V.; Yee, D., Insulin-like growth factor binding protein-3 and insulin receptor substrate-1 in breast cancer: correlation with clinical parameters and disease-free survival. *Clin Cancer Res* **1997**, *3* (1), 103-9.
308. Loi, S., Molecular analysis of hormone receptor positive (luminal) breast cancers: what have we learnt? *Eur J Cancer* **2008**, *44* (18), 2813-8.

309. Voduc, K. D.; Cheang, M. C.; Tyldesley, S.; Gelmon, K.; Nielsen, T. O.; Kennecke, H., Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol* 28 (10), 1684-91.
310. Foulkes, W. D.; Reis-Filho, J. S.; Narod, S. A., Tumor size and survival in breast cancer--a reappraisal. *Nat Rev Clin Oncol* 7 (6), 348-53.
311. Lacroix, M., Significance, detection and markers of disseminated breast cancer cells. *Endocr Relat Cancer* **2006**, 13 (4), 1033-67.
312. Mainiero, M. B., Regional lymph node staging in breast cancer: the increasing role of imaging and ultrasound-guided axillary lymph node fine needle aspiration. *Radiol Clin North Am* 48 (5), 989-97.
313. Bouchal, P.; Roumeliotis, T.; Hrstka, R.; Nenutil, R.; Vojtesek, B.; Garbis, S. D., Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. *J Proteome Res* **2009**, 8 (1), 362-73.
314. Li, J.; Gromov, P.; Gromova, I.; Moreira, J. M.; Timmermans-Wielenga, V.; Rank, F.; Wang, K.; Li, S.; Li, H.; Wiuf, C.; Yang, H.; Zhang, X.; Bolund, L.; Celis, J. E., Omics-based profiling of carcinoma of the breast and matched regional lymph node metastasis. *Proteomics* **2008**, 8 (23-24), 5038-52.
315. Beasley, G. M.; Olson, J. A., Jr., What's new in neoadjuvant therapy for breast cancer? *Adv Surg* 44, 199-228.
316. Chen, R.; Li, A.; Zhu, T.; Li, C.; Liu, Q.; Chang, H. C.; Zhou, J., JWA--a novel environmental-responsive gene, involved in estrogen receptor-associated signal pathway in MCF-7 and MDA-MB-231 breast carcinoma cells. *J Toxicol Environ Health A* **2005**, 68 (6), 445-56.
317. Jung, D. J.; Na, S. Y.; Na, D. S.; Lee, J. W., Molecular cloning and characterization of CAPER, a novel coactivator of activating protein-1 and estrogen receptors. *J Biol Chem* **2002**, 277 (2), 1229-34.
318. Mercier, I.; Casimiro, M. C.; Zhou, J.; Wang, C.; Plymire, C.; Bryant, K. G.; Daumer, K. M.; Sotgia, F.; Bonuccelli, G.; Witkiewicz, A. K.; Lin, J.; Tran, T. H.; Milliman, J.; Frank, P. G.; Jasmin, J. F.; Rui, H.; Pestell, R. G.; Lisanti, M. P., Genetic ablation of caveolin-1 drives estrogen-hypersensitivity and the development of DCIS-like mammary lesions. *Am J Pathol* **2009**, 174 (4), 1172-90.

319. Chen, H.; Bai, J.; Ye, J.; Liu, Z.; Chen, R.; Mao, W.; Li, A.; Zhou, J., JWA as a functional molecule to regulate cancer cells migration via MAPK cascades and F-actin cytoskeleton. *Cell Signal* **2007**, *19* (6), 1315-27.
320. Chen, R.; Qiu, W.; Liu, Z.; Cao, X.; Zhu, T.; Li, A.; Wei, Q.; Zhou, J., Identification of JWA as a novel functional gene responsive to environmental oxidative stress induced by benzo[a]pyrene and hydrogen peroxide. *Free Radic Biol Med* **2007**, *42* (11), 1704-14.
321. Cao, H.; Xia, W.; Shen, Q.; Lu, H.; Ye, J.; Li, A.; Zou, C.; Zhou, J., Role of JWA in acute promyelocytic leukemia cell differentiation and apoptosis triggered by retinoic acid, 12-tetradecanoylphorbol-13-acetate and arsenic trioxide. *Chinese Science Bulletin* **2002**, *47* (10).
322. Wang, S.; Gong, Z.; Chen, R.; Liu, Y.; Li, A.; Li, G.; Zhou, J., JWA regulates XRCC1 and functions as a novel base excision repair protein in oxidative-stress-induced DNA single-strand breaks. *Nucleic Acids Res* **2009**, *37* (6), 1936-50.
323. Wang, N. P.; Zhou, J. W.; Li, A. P.; Cao, H. X.; Wang, X. R., [The mechanism of JWA gene involved in oxidative stress of cells]. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi* **2003**, *21* (3), 212-5.