NOVEL QUANTITATIVE PROTEOMIC APPROACHES TO REVEAL NEW DRIVING MECHANISMS AND BIOMARKERS OF TUMORIGENESIS

J. Astor Ankney

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics in the School of Medicine

Chapel Hill
2019

Approved by:

Xian Chen

Gaorav P Gupta

Leslie V Parise

Brian D Strahl

Gang Greg Wang

ABSTRACT

J. Astor Ankney: Novel quantitative proteomic approaches to reveal new driving mechanisms and biomarkers of tumorigenesis
(Under the direction of Xian Chen)

Understanding the biological processes and pathways important to healthy and disease states requires the probing of systems at the protein level. While genetic alterations may predict the likelihood of disease, it is the phenotypic changes which characterize disease onset. The growth of mass spectrometry-based proteomics over the past two decades has been essential for our ability to perform systems-level interrogations of both healthy and disease states. Here, we have employed quantitative proteomic techniques to investigate breast cancer and glioma biology.

Chapter 1 provides background on breast cancer heterogeneity and the challenges in classifying breast tumors in a clinically relevant manner. It additionally presents an overview of mass-spectrometry based proteomics and methods of proteomic quantitation.

Chapter 2 presents a novel workflow for the identification of new liquid biopsy biomarkers for breast cancer prognosis. We developed a multi-omic approach linking oncogenic secreted proteins to patient-specific transcriptomic data and clinical outcomes. Kaplan-Meier analysis of genes having a secretion correlated expression pattern was used to identify new liquid biopsy biomarkers for predicting individualized prognosis.

Chapter 3 explores the role of spliceosome protein SNRPD1 in breast cancer. SNRPD1 is often overexpressed in breast cancer and plays a seemingly important, but understudied, role in tumor aggressiveness. We have performed the first systems-level study of SNRPD1 function in breast cancer cells and identify previously unreported signaling pathways and biological functions impacted by SNRPD1.

Chapter 4 introduces a novel approach to the study of secreted proteins termed outside-in proteomics. We first perform an unbiased screen of the secretome to identify phenotypic-relevant markers and use those results to guide our investigation of the intracellular proteome. We applied this approach to analyze the role of EZH2 in the aggressive phenotype of triple negative breast cancer.

Chapter 5 presents a quantitative proteomic comparison of IDH1 mutant and wild type glioma cells. We performed a proteomic screen of IDH1 mutant versus wild type gliomas, and we report several new pathways and processes which may contribute to IDH1 mutant glioma pathogenesis or progression.

Overall, this work demonstrates both novel and traditional mass spectrometric proteomic techniques for the investigation of human disease.

For my wife, who didn’t sign up for this but never let me quit.

ACKNOWLEDGEMENTS

I cannot imagine getting this far without tremendous help from a large number of people. A few sentences are not enough to fully express my sincere gratitude to everyone who has contributed to my success, but in this space a few sentences are all I have. I would first like to thank my thesis advisor and mentor Xian Chen. I could not have asked for a better advisor. You have given me the support, freedom, and criticism I needed to grow as a scientist. I would also like to thank my committee members, Greg Wang, Leslie Parise, Gaorav Gupta, and Brian Strahl. You asked tough questions, provided honest feedback, and gave me confidence in my project and in myself. I thank my labmates for helping me work through issues of technique and theory. In particular, I appreciate the patience of Amey when helping me troubleshoot an experiment or teaching a new technique, and John, who wrote many of the analytical tools I have used, in explaining how to use his scripts and writing new ones at my request. Last, but in no way least, I must thank Ling Xie. If you were ever flustered by my constant questions, you never showed it. You have been an invaluable part of my training as a mentor and as a friend.

It is also important for me to thank the people who helped me get started in science. I thank Larry Eimers for giving me a free career aptitude test that set me on this path. Dr. Isabelle Lemasson gave me my first undergraduate research position and stoked my passion for research. Yu Yang invited me to join his lab and gave me confidence in my abilities. Geoff Mueller gave me a 10 week internship which challenged me and sparked my independent thought. I thank each of you for guiding me down this path.

Finally, having support at home has probably been the most important factor in my success. My children have chastised me for staying late in lab, given me hugs after a bad day, and been a constant reminder of why I started this in the first place. My parents and in-laws

have provided much-needed moral support (not to mention babysitting!), but my mom has been especially supportive and I will never be able to thank her appropriately. Each of the people named above has been instrumental to my adventure in science, but none has been as important as my wife. Christina, you have supported me unconditionally and there is no way to overstate how vital that has been to me. We have done this together. Now we see what's next together.

PREFACE

Portions of the work described in this dissertation were previously published:

Chapter 1 contains excerpts from J. Astor Ankney, Adil Muneer, and Xian Chen. Relative and Absolute Quantitation in Mass Spectrometry–Based Proteomics. *Annual Review of Analytical Chemistry*. 2018; 11:49-77. doi: 10.1146/annurev-anchem-061516-045357 reprinted with permission

Chapter 2 is reproduced in full from J. Astor Ankney, Ling Xie, John Wrobel, Li Wang, and Xian Chen. Novel secretome-to-transcriptome integrated or secreto-transcriptomic approach to reveal liquid biopsy biomarkers for predicting individualized prognosis of breast cancer patients. *BMC Medical Genomics*. 2019; 12:78. doi: 10.1186/s12920-019-0530-7 under the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

α-KG	α-ketoglutarate
APEX	Absolute protein expression
BC	Breast cancer
BLBC	Basal-like breast cancer
dNSAF	Distributed normalized spectral abundance factor
DRFS	Distant relapse free survival
emPAI	Exponential protein abundance index
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor
FDR	False discovery rate
GBM	Glioblastoma multiforme
GOBP	Gene Ontology biological process
GOCC	Gene Ontology cellular component
H3K27Me3	Histone H3 lysine 27 tri-methylation
HER2	Human epidermal receptor 2
IPA	Ingenuity Pathway Analysis
KD	Knockdown
KM	Kaplan-Meier
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LFQ	Label- free quantitative
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MS	Mass spectrometry
nanoLC	Nanoflow liquid chromatography
NSAF	Normalized spectral abundance factor

PAF	Protein abundance factor
PAI	Protein abundance index
PPI	Protein-protein interaction
PR	Progesterone receptor
PRC2	Polycomb repressive complex 2
R-2HG	R-2-hydroxyglutarate
SC	Spectral counting
SeCEP	Secretion-correlated mRNA expression pattern
snRNP	Small nuclear ribonucleoprotein
snRNA	Small nuclear RNA
SpC	Spectral counts
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TCGA	The Cancer Genome Atlas
TET	Ten eleven translocation
TNBC	Triple negative breast cancer
XIC	Extracted ion chromatogram

CHAPTER 1: BACKGROUND

Breast cancer epidemiology and etiology

Breast cancer (BC) is the most frequently diagnosed cancer among women in the United States and worldwide, and BC is the second leading cause of cancer death among women in the US.[1] However, despite decreased incidence in lesser developed countries, higher mortality rates in these areas make BC the leading cause of cancer death among women worldwide.[1] It is estimated that approximately 12%, or 1 in 8, women in the United States will be diagnosed in her lifetime.[2] Women of age 50 or older comprise approximately 80% of breast cancer diagnoses and almost 90% of BC deaths.[2] Although survival rates are higher in BC relative to many other cancer types, [3] this disease remains a significant public health concern.

The incidence rate of BC in the US has increased by an average of 0.4% per year since 2004, [2, 3] however the mortality rate has continued to decline due in part to earlier detection by increased rates of screening mammography and the development of more effective and targeted therapies.[4, 5] Notwithstanding improvement in survival rates, significant racial disparities in BC incidence and mortality have been observed, and black women have higher BC mortality rates versus white women despite having lower incidence rates.[2, 6] These disparities are partially due to later disease stage at the time of diagnosis, but black women also have worse survival than white women with the same disease stage.[7] Further, black women are more likely to be diagnosed with more aggressive forms of cancer versus other races.[2]

BC susceptibility is influenced by both genetic and non-genetic factors.[8, 9] Women with a close family member diagnosed with BC have an estimated two-fold increased risk of also developing BC.[10, 11] This risk increases depending on the number of close family members

diagnosed with BC and their age at first detection.[11] High susceptibility genes BRCA1 and BRCA2 account for less than 25% of this increased risk [12] and are rare in the population.[13] Other known genetic factors conferring intermediate risk, including ATM, PALB2, and CHEK2, comprise 2 to 5% of the increased familial risk.[9] Overall only 37% of excess familial risk is explained by known genetic mutations, with the remaining risk due to undiscovered genetic variants or other factors.[14] Environmental factors also play an important role in BC etiology. These include alcohol consumption, body mass index, and reproductive factors including ages at menarche, first birth, and menopause.[8, 14]

Breast cancer heterogeneity and classification

Breast cancer is not a single disease, but is a heterogeneous group of diseases with distinct clinical features and biological characteristics.[15-17] BC tumors vary widely in morphological structure, genetic mutation, biological behavior and metastatic potential.[18, 19] Even within an individual patient a tumor may exhibit heterogeneity in different regions of the tumor (spatial intratumor heterogeneity) or properties of the tumor may change over time (temporal intratumor heterogeneity).[20] For decades pathologists have recognized the need to classify tumors to guide therapeutic decisions and aid in prognosis.[21, 22]

Tumor classification may be defined by clinical, histological, or molecular features. An early classification based on clinical features was tumor stage, and this classification is still standard in pathology reports. Previously, tumor stage was defined solely by primary tumor size, involvement of regional lymph nodes, and distant metastasis.[23] Based on these features the patient was assigned a stage from 0 (non-invasive) to IV (metastatic).[23] In 2018, the American Joint Committee on Cancer updated the staging guidelines to also include molecular and histological features of the tumor.[24]

The histological tumor grade is based on the morphologic and proliferative characteristics of the tumor based on tubule formation, nuclear pleomorphism, and mitotic

index.[22] However, when initially introduced this system was susceptible to problems with reproducibility and interobserver variation in grading.[25, 26] Revisions to this grading system [21] increased objectivity and made grading semiquantitative, substantially improving reproducibility.[25, 26] Due to the prognostic value, this grading system is an important component of histopathological reports.[26, 27]

Histological tumor type is another longstanding method of classification still in use despite having limited prognostic value.[28] Based on cytological and architectural features, a tumor may be assigned to one of at least 17 subtypes.[20, 25] In this system, approximately 70% of all breast tumors are invasive ductal carcinomas not otherwise specified, which is a diagnosis of exclusion since these tumors cannot be classified as any special type.[15, 29] Another 5-15% are invasive lobular carcinomas which have lower risk of mortality.[25, 30] Although tumors of special histological type differ prognostically, this classification method has limited clinical utility because the vast majority of tumors are of no special type.[20]

While classification of BC tumors based on clinical and histological features provides a primarily prognostic benefit, molecular features can aid in therapeutic decision-making. The potential role of hormone signaling in breast cancer was first recognized in the late 19th century.[31] By the 1960's experiments using radiolabeled estrogens had established that these hormones accumulate and promote growth in specific tissues [32, 33] supporting the hypothesis that an estrogen receptor (ER) may exist in these tissues.[34] Subsequent experiments confirmed the existence of the estrogen receptor.[35, 36] Within a decade additional experiments established the link between ER expression and response to endocrine therapy, [37] and assays to quantify ER expression were being developed.[38] In addition to the therapeutic role, ER expression was also identified as an independent marker of better prognosis.[39]

Despite these encouraging advances, about 40% of ER+ tumors did not respond to endocrine therapy.[40] To explain this discrepancy, Horwitz and colleagues hypothesized that

some ER+ tumors may have a defect in the estrogen signaling pathway which prevents estrogen action in the cell.[41] Reasoning that the presence of an estrogen signaling product may indicate an intact signaling pathway, and therefore endocrine therapy responsiveness, the progesterone receptor (PR) was proposed as a suitable marker.[41] However, the use of PR as a predictor of response to endocrine therapy has been questionable due to differences in results between studies.[42, 43] Further, limited response is observed in some ER+/PR- tumors so that the clinician and patient must weigh the risks and benefits of undertaking endocrine therapy on an individual basis.[44] Endocrine therapy for the uncommon ER-/PR+ phenotype remains controversial.[45] Although PR status has questionable utility as a therapeutic predictor, PR+ patients typically have more favorable outcomes than PR- patients, making PR an important prognostic marker in BC.[45]

The existence of human epithelial growth factor receptor 2 (HER2) and its amplification in BC was first described by a series of reports in the mid-1980's [46-49] and its importance to relapse and survival were reported only two years later.[50] The prognostic significance of HER2 is well supported by subsequent studies.[51-53] HER2 is overexpressed by 15-20% of BC tumors, [54, 55] and these tumors are typically more aggressive with poorer differentiation and higher lymph node involvement than HER2- tumors.[56, 57] Further, although about half of HER2+ tumors are also ER+, these tumors are more likely to develop resistance to endocrine therapy, likely due to crosstalk between signaling pathways.[58] Encouragingly, since the introduction of HER2-targeted therapies including trastuzumab, [57] overall and relapse-free survival rates for HER2+ BC have improved significantly.[54]

Together, ER, PR, and HER2 are the most common molecular markers used to predict prognosis and response to therapy. Importantly, approximately 15% of BC tumors do not express any of these markers (ER-/PR-/HER2-) and comprise a critical additional class of BC called triple negative breast cancer (TNBC).[59, 60] TNBC is the most aggressive form of BC, with higher rates of metastasis and worse prognosis than other groups.[61, 62] Further, TNBC

cannot be treated with therapies that target HER2 or estrogen signaling, making systemic chemotherapy and radiation the most common treatment options.[61, 62] The development of improved therapies for TNBC is currently an area of intense research.[61]

Among broadly used classification systems, the most recently developed relies on DNA microarray analysis of tumor gene expression. In a seminal work, Perou et. al. identified four BC subtypes based on the expression of a 496-gene set.[63] Further refinement led to the establishment of a 50-gene set (PAM50) which identified five primary "intrinsic molecular subtypes": luminal A, luminal B, basal-like (BLBC), HER2-enriched, and normal-like.[64] Importantly, molecular subtypes exhibit important differences in prognosis [65] and response to therapy.[64, 66] Approximately half of tumors are classified as luminal A making this the most prevalent PAM50 subtype.[67, 68] About 20% of tumors are luminal B, 15-20% are BLBC, 10-15% are HER2-enriched, and >5% are normal-like.[67, 68] PAM50 subtypes roughly match molecular marker status as luminal A tumors are typically ER+/ER2-, luminal B are usually ER+/HER2+, the HER2-enriched group is most often ER-/HER2+, and the majority of basal-like tumors are triple-negative.[63] Although these groups are sometimes inappropriately equated [2, 69] the PAM50 subtype classification is distinct from receptor status.[70] Despite the dramatic advances that gene expression profiling has provided in prognostic prediction and overall understanding of tumor biology and heterogeneity, the lack of correlation between PAM50 subtype and tumor phenotype make this classification system unsuitable for therapeutic decision-making.[70]

A new generation of phenotypic markers is needed

The preceding classification systems vary in utility for prognosis and guidance of therapeutic decisions. The value of tumor stage assessment has been diminished with the introduction of more precise indicators of prognosis and response to therapy, however tumor stage remains a vital consideration in parts of the world where assessment of other markers is

unavailable.[24] Likewise, histological tumor type can be prognostically valuable in special-type tumors, however about 70% of tumors are of no special type and are therefore not benefitted by this system.[20] On the other hand, tumor receptor status has proven highly clinically valuable but cannot predict which tumors will be unresponsive to endocrine or HER2-directed therapies or will develop therapeutic resistance.[19, 71] Finally, the transcriptomic data used in gene expression profiling is unable to resolve the phenotypic changes which govern BC characteristics.[70] To prevent undertreatment of aggressive BC and avoid overtreatment of patients who will gain little benefit, there is a pressing need for a new generation of phenotypic BC markers which can better predict personalized prognosis and response to therapy.

The cancer cell secretome as a source of novel biomarkers

The most basic definition of a biomarker as set forth by the NIH Biomarkers Definitions Working Group is “A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.”[72] However, the ideal biomarker should be obtainable in a non-invasive manner and be easily and inexpensively assayed.[73-76] Because bodily fluids including blood can be obtained non-invasively, they are excellent biomarker reservoirs.[77, 78]

Many biological molecules and processes have been proposed as potential cancer biomarkers, e.g. microRNA, circulating tumor DNA, and DNA methylation patterns.[79-81] Alterations in any of these proposed markers usually results in an increase or decrease in the expression of various protein products.[82] Tumor cells secrete and shed characteristic proteins at a higher rate than healthy cells, and many of these proteins enter circulation to play extracellular regulatory roles.[83, 84] Secreted proteins make up approximately 10% of the human proteome and play an important role in normal physiological processes including cell signaling, immune defense, and blood coagulation.[77, 85] However, due to their regulation of cell-to-cell and cell-to-extracellular matrix interactions, secreted proteins are critical participants

in angiogenesis, invasion, and metastasis when deregulated in cancer.[77, 85] Secreted proteins have also been increasingly recognized for their role in the mechanisms of drug response.[86] Proteins secreted or shed by cancer cells, collectively referred to as the “cancer secretome”, could therefore be phenotypic biomarkers. More importantly, in clinical practice these tumor-characteristically secreted proteins may be detectable in blood or other bodily fluids in a non-invasive manner.[83, 84]

Challenges in blood-based proteomics and alternative approaches

Immunoassays are in wide clinical use for the detection and quantification of individual plasma proteins but are limited by isoform specificity and low multiplex capability.[87] Mass spectrometry (MS) –based proteomics is a commonly used technology which can overcome these limitations.[85, 87] However, a major challenge in biomarker discovery using blood-based proteomics is the complex diversity of blood. Thousands of proteins with varying isoforms and glycosylation patterns, have a dynamic range of abundance spanning 10 or more orders of magnitude.[85, 88] Albumin alone comprises over half of the protein mass in blood, and only a dozen proteins account for 95% of total blood protein mass.[88, 89] These highly abundant proteins may suppress the detection or quantification of lower abundance proteins which are potential biomarkers.[85] Several methods have been developed to address this obstacle by depletion of high abundance proteins from the plasma sample before MS analysis, including Cibacron blue dye to remove albumin, protein A and G resins to remove immunoglobulins, and various commercial products incorporating multiple antibodies to deplete panels of 7-50 highly abundant proteins.[90-93] Despite improving low abundance protein identification, these methods are inefficient and unreliable due to the depletion of unintended targets.[92, 94]

To avoid direct blood analysis, secretome studies have been performed on tumor interstitial fluid, [95] proximal body fluids such as nipple aspirate fluid, [96-98] and conditioned media from cell lines.[99, 100] Results using proximal body fluids have been mixed, partially due

to the need for collection protocol standardization.[77] Proximal fluid controls may also be insufficient due to discrepancies in fluid output from disease and healthy tissue.[101] Likewise, the use of tumor interstitial fluid is complicated by difficulty in sample availability, and proteomic results are directly influenced by the collection method.[102] On the other hand, cell lines are benefitted by the reduced complexity of conditioned media versus biological fluids and controlled experimental conditions which limits experimental variability.[77] Sampling of cell line conditioned media has therefore become a common approach to proteomic secretome analysis.[99, 103-106]

Label Free Quantitative Proteomics

Accurate measurement of molecular changes in disease associated processes and pathways is critical to better understand pathogenesis and to discover new biomarkers for diagnosis, prognosis, and treatment. While genetic alterations predict the likelihood of disease, it is the phenotypic changes which define disease onset. Emerging techniques of mass MS-based proteomics over the past two decades have enabled the genome-wide analysis of the phenotype-correlated protein changes in comparing healthy versus disease states. As proteomic technology has advanced, including improvements in computing resources, algorithm design, and nanoflow liquid chromatography, the field has shifted focus to accurate and reliable quantitation rather than mere protein identification. Because mass spectrometry is not inherently quantitative, myriad strategies have been developed to assist MS for protein quantitation. These strategies can often be distinguished not only by the method of quantitation, i.e., labelling versus label-free, but also by the quantitative goal, i.e., relative versus absolute quantitation and the scope of the experiment, i.e., targeted versus shotgun strategies. Stable isotope labeling strategies can generally be divided into two groups: postharvest tagging strategies (e.g., ICAT, iTRAQ, TMT) introduce labels at protein or peptide level after cells have been lysed, while metabolic labelling strategies (e.g., uniform N^{15} labeling, AACT/SILAC) involve endogenous incorporation of isotopic

labels in metabolically active cells. Even within these broadly general groupings, numerous options exist with each offering advantages and disadvantages. For example, labeling techniques can tag peptides chemically, metabolically, or enzymatically, while label-free quantitation (LFQ) techniques can be ion counting- or intensity-based. Multiple approaches are further available even within these narrower groupings.

Increasingly, LFQ approaches are being employed to replace or complement labeling methods due to the ease of experimental set-up and relative low cost. Two major methods have emerged for label free quantitation. The first, known as spectral counting (SC), provides relative quantitation by comparing the number of MS/MS spectra produced by the same protein across multiple liquid chromatography-tandem MS (LC-MS/MS) runs. The second is based on the precursor ion intensity as determined by the extracted ion chromatogram (XIC), which is the plot of intensity versus retention time of a particular m/z value. Both LFQ methods offer several benefits over labeling methods. (i) LFQ approaches typically identify larger number of proteins with a wider dynamic range of detection compared to labeling methods.[107] (ii) Both LFQ approaches are less expensive because no special media or expensive reagents are required to perform an experiment.[107] (iii) LFQ workflows are simpler and faster because there is no need to determine labeling efficiency.[108] (iv) LFQ approaches offer the ability to compare a large number of experiments, whereas labeling methods are limited in multiplexing capabilities.

Ion counting-based label-free quantitation

One of the earliest attempts to quantify proteins without labeling was based on peptide counting.[109, 110] The protein abundance index (PAI) provided only a rough estimate of quantitation by dividing the number of peptides identified for a protein by the total number of theoretical peptides the protein could produce.[109, 110] Later, one of the two groups who simultaneously developed the method refined it to emPAI, the exponential form of PAI minus one.[111] Although emPAI shows better correlation to known protein amounts, it is still less

accurate than other available methods.[112] However, unlike most other LFQ methods emPAI can be used to estimate total protein abundance.[111]

In the spectral counting method, the number of MS/MS spectra identified for a given protein across multiple LC-MS/MS runs provides relative quantification.[113] The relationship between spectral counts and protein abundance was first explored based on the observation that more abundant proteins will produce a larger number of proteolytic peptides.[113] These peptides are more likely to be sampled, yielding a higher number of spectra.[113] Indeed, protein abundance is strongly correlated (r2=0.9997) to spectral count but only weakly correlated to peptide count or sequence coverage.[113] However, the common use of dynamic exclusion to aid in the identification of low abundance peptides has a suppressive effect on quantitation by spectral counting. [107] Dynamic exclusion parameters which maximize identification of low abundance peptides while minimizing undersampling of highly abundant peptides have been reported, but this remains a consideration when designing an experiment.[114]

Another concern with spectral counting is based on the observation that the number of peptides generated by proteolytic cleavage is dependent on the length of the protein. Hence, quantitation of proteins less than 20 kDa tends to be less accurate than for larger proteins.[115, 116] Several methods attempt to resolve this discrepancy by normalizing spectral count data based on protein length, molecular mass, or machine learning classification. Most commonly, Normalized Spectral Abundance Factor (NSAF) is employed. NSAF is calculated by dividing the number of spectral counts (SpC) belonging to a given protein by the length of the protein (L), then dividing that value by the sum of (SpC/L) values for all proteins in the experiment.[116] Normalization by NSAF offers the added benefit of improving the minimum fold change detectable by SC.[116] While NSAF normalizes SC data based on protein length, Protein Abundance Factor (PAF) focuses on protein mass by normalizing the total number of non-redundant spectra to the molecular mass of the intact protein.[117] A third method of handling this discrepancy is the

Absolute Protein Expression (APEX) algorithm which is a machine learning classification system that corrects spectral counts for the likelihood that a spectrum might be detected.[115]

SC shows some limitation in the ability to quantify small abundance changes. Typically, larger numbers of spectra must be identified in order to calculate smaller abundance changes.[107] The minimum detectable fold change and the number of spectra required can be improved by increasing the scoring requirements for spectrum identification.[118] However, these improvements come at the expense of dynamic range and the accurate quantification of lowly abundant proteins.[118] A final limitation of SC is the handling of peptides that can be assigned to more than one protein. Typically these spectra would be assigned to all possible matching proteins.[119] This has an obvious negative impact on the accuracy of the quantitation for those proteins, since any peptide can only have been generated by a single protein.[119] To address this issue the distributed Normalized Spectral Abundance Factor (dNSAF) was developed, which divides shared spectra proportionally between the possible contributing proteins based on the distribution of the other identified unique peptides.[119]

Intensity-based label-free quantitation

The other major method used for label free quantitation is based on the precursor signal intensity from the extracted ion chromatogram. As spectral counts are strongly correlated with peptide abundance, so too is precursor signal intensity.[120] Conceptually, quantitation is an area under the curve or peak height calculation for each peptide as it elutes from the LC column at an expected retention time. Because LC retention time is an important aspect of the quantitation, a robust and consistent LC program is necessary to allow for more accurate quantitation and fewer incorrect peptide assignments. Retention time alignment is also a crucial data processing step.[121] Intensity normalization based on total ion count is another necessary step to account for bias in signal intensity.[122]

One concern with XIC-based quantitation is that XIC mapping is based on the full MS spectrum while peptide identification is based on MS2 spectra. However, MS is typically operated in a data dependent manner such that MS1 and MS2 spectra are obtained in the same run with MS2 acquisition being based on MS1 results. Because identification and quantitation are competing for scan time, a balance must be achieved between these goals. Excess time dedicated to peptide identification will negatively impact quantitation. Conversely, too much focus on quantitation will reduce the number of MS2 scans leading to a lower number of peptides identified. Another concern with XIC-based quantitation versus SC is that peak identification, noise reduction, retention time alignment, and peak intensity calculations including normalization require significantly more computational power than the simpler spectral counting process. Several algorithms requiring varying levels of computational strength have been published. (Reviewed in Ref. [123])

Numerous comparisons between SC and XIC-based quantitation have been performed (Reviewed in Ref. [124]). XIC-based methods are consistently more sensitive and more accurate than SC when using high resolution mass spectrometers because these machines are able to discriminate between co-eluting peptides of similar mass while allowing the accurate mapping of XIC to peptide.[125] XIC-based quantitation also offers the advantage of accurately discerning fold-changes as low as 1.1 versus the SC limit of approximately 1.4. [112, 116] Some reports have also noted that SC tends to overestimate low abundance peptides.[126, 127]

Limitations of single-omics approaches

While proteomics is commonly used for the identification of potential biomarkers, at the current level of sensitivity, mass spectrometry can detect only very limited regions of each individual protein, limiting information about patient-specific alterations in secreted proteins.[128, 129] Further, due to the sampling limitations of blood-based proteomics, there is no database which contains patient-specific proteomic data and associated clinical data. There are many

databases which include transcriptomic and clinical data, however the correlation bbetween transcriptome and phenotype can be very low, which is problematic because phenotype matters most. Therefore, due to either the low phenotypic accuracy of genomics/transcriptomics or the low phenotypic coverage of proteomics, approaches that employ only single-omics methods will necessarily fail to identify biomarkers of patient-specific alterations which distinguish patient subpopulations having different clinical outcomes or prognoses.

CHAPTER 2: NOVEL SECRETOME-TO-TRANSCRIPTOME INTEGRATED OR SECRETO-TRANSCRIPTOMIC APPROACH TO REVEAL LIQUID BIOPSY BIOMARKERS FOR PREDICTING INDIVIDUALIZED PROGNOSIS OF BREAST CANCER PATIENTS

INTRODUCTION

Breast cancer (BC) is the most prevalent type of cancer among women in the United States, with over 200,000 new diagnoses of invasive BC per year.[130] However, significant heterogeneity among BC tumors contributes to highly variable clinical pathology and patient outcomes, ultimately confounding efforts toward precision diagnosis and prognosis.[131] A 50-gene expression pattern has been used to classify five molecular subtypes or PAM50-subtypes, including basal-like/triple-negative (BLBC/TNBC), luminal-A and –B, Her2+, and normal-like BC.[64] Within these molecular subtypes, the luminal subtype accounts for approximately half of all tumors, [132] and BLBC/TNBC is the most aggressive form of the disease with the overall worst survival rate.[65] However, these gene-expression signatures are inadequate to resolve interpatient heterogeneity and patient subpopulations with different clinical outcomes cannot be stratified within each BC or PAM50 subtype.[133] These limitations arise because disease onset is directly governed by phenotype-specific proteomic changes [134, 135] which cannot be measured using genomic/transcriptomic tools or data alone. Because prognoses of BC patients cannot be easily discerned, there is an urgent need for individualized/personalized biomarkers that predict patient-specific survival rates and therapeutic response so that standard chemotherapy may be replaced by more effective and precise treatment. Because tumor cells secrete and shed characteristic proteins at a higher rate than healthy cells, and many of these proteins enter circulation to play extracellular regulatory roles, [83, 84] proteins secreted or shed

by cancer cells (the “cancer secretome”), could be phenotypic biomarkers. More importantly, in clinical practice these tumor-characteristically secreted proteins may be detectable in blood or other bodily fluids in a non-invasive manner.[84]

Secreted proteins, which constitute approximately 10% of the human proteome, play an important role in normal physiological processes including cell signaling, immune defense, and blood coagulation.[77] Further, when deregulated, secreted proteins are critical participants in pathological processes such as cancer angiogenesis, invasion, and metastasis.[77] Also, secreted factors have been increasingly recognized for their role in the mechanisms of drug response.[86] The studies of BC secretomes using nanoliter liquid chromatograph tandem mass spectrometry (nanoLC-MS/MS) to sample conditioned medium from cell lines, tumor/tissue interstitial fluid, or tumor proximal body fluids have been reported.[83, 99, 136-138] However, few proteins identified in these BC secretomes have established correlation with patient-specific clinical outcomes.

Technically, at the current level of sensitivity, mass spectrometry can detect only very limited regions of each individual protein, further limiting information about patient-specific alterations in these secreted proteins.[128, 129] Therefore, due to either the low phenotypic accuracy of genomics/transcriptomics or the low phenotypic coverage of proteomics, approaches that employ only single-omics methods will necessarily fail to identify biomarkers of patient-specific alterations which distinguish patient subpopulations having different clinical outcomes or prognoses. To overcome these single-omics limitations for identifying new biomarkers to predict individualized prognosis in non-invasive, blood-based tests, we developed a new multi-omics method, termed *secreto-transcriptomics*, to identify the BC-subtypic secreted proteins that are encoded by genes bearing patient-specific mRNA expression patterns of prognostic significance.

Strategically in advance, our secreto-transcriptomic approach bypasses both the inability of conventional MS to connect genotype to phenotype and the inability of MS to fully identify

patient-specific proteomic alterations, integrating oncogenic (tumorigenic) multi-omics data for efficient *de novo* discovery of personalized/individualized prognostic markers. In the clinic, these markers may be used to stratify patients, within single PAM50 subtypes, into different prognostic groups, and predict treatment benefit and/or outcome with patient-specific or individualized sensitivity and specificity before any therapeutic decision for newly diagnosed breast cancer.

METHODS

Chemicals and reagents. Cell culture media and fetal bovine serum (FBS) were obtained from Gibco. All other components of cell culture media and protease inhibitor cocktails were purchased from Sigma (St, Louis, MO). Trypsin was purchased from Promega. All chemicals were HPLC-grade unless specifically indicated. All cell lines including MCF10A, MCF7, MDA-MB-231, T47D, and HCC1806 were purchased from ATCC (Manassas,VA).

Secreted protein extraction from cell lines. MCF10A cells were cultured in DMEM/F12 supplemented with 5% horse serum, 20 ng/mL epidermal growth factor, 50 ng/mL cholera toxin, 500 ng/mL hydrocortisone, and 2 µg/mL insulin. MCF7 and MDA-231 cells were cultured in DMEM containing 10% fetal bovine serum. T47D and HCC1806 cells were maintained in RPMI supplemented with 10% fetal bovine serum. When cells reached approximately 70% confluence, the growth media was removed, cells were washed twice with PBS, and serum-free media without phenol red was added to the plate. After 24 hours the conditioned media were collected and centrifuged at 500 x g for 5 minutes to remove cellular debris, then the supernatant was syringe-filtered with 0.2 µm 13 mm diameter PTFE filters (VWR International) and transferred to fresh tubes. Samples were stored at -80 °C until further processed. After thawing, proteins were concentrated by trichloroacetic acid/sodium deoxycholate precipitation. Briefly, 1/10 of the sample volume of 0.15% sodium deoxycholate was added to each sample, then tubes were incubated on ice for 15 minutes. Next, 1/10 of the original sample volume of cold 72% TCA was added and the tubes were incubated on ice for 15 minutes. Samples were

centrifuged for 10 minutes at max speed, 4° C. The pellets were washed in cold acetone and air dried until no residual odor was detected. Next, the pellets were resuspended in 50 µl buffer (8 M Urea, 50 mM Tris-HCl pH 8.0, 150 mM NaCl), reduced with dithiothreitol (5 mM final) for 30 minutes at room temperature, and alkylated with iodoacetamide (15 mM final) for 45 minutes in the dark at room temperature. Alkylation was quenched with dithiothreitol (10 mM final). Samples were diluted 4-fold with 25mM Tris-HCl pH 8.0, 1mM $CaCl_2$ and digested with 500 ng trypsin overnight at room temperature. Peptides were desalted on a StageTip containing a 4 × 1 mm C18 extraction disk (3M) and dried.[139]

LC-MS/MS analysis. Desalted peptides were dissolved in 20 µl 0.1% formic acid (Thermo-Fisher), of which 2 µl was injected and analyzed by an Easy nanoLC 1000 coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) Peptides were loaded on to a 15 cm C18 RP column (15 cm × 75 µm ID, C18, 2 µm, Acclaim Pepmap RSLC, Thermo-Fisher) and eluted with a gradient of 2-30% buffer B at a constant flow rate of 300 nl/min for 70 min followed by 30% to 80% B in 5 min and 80% B for 10 min. The Q-Exactive was operated in the positive-ion mode but using a data-dependent top 20 method. Survey scans were acquired at a resolution of 70,000 at m/z 200. Up to the top 20 most abundant isotope patterns with charge ≥ 2 from the survey scan were selected with an isolation window of 2.0 m/z and fragmented by HCD with normalized collision energies of 27. The maximum ion injection time for the survey scan and the MS/MS scans was 250 ms and 120 ms, respectively and the ion target values were set to 1e6 and 2e5, respectively. Selected sequenced ions were dynamically excluded for 20 seconds.

Mass spec data and LFQ analysis. Mass spectral processing and peptide identification were performed on the Andromeda search engine in MaxQuant software (Version 1.5.3.17) against a human UniProt database. All searches were conducted with a defined modification of cysteine carbamidomethylation, with methionine oxidation and protein amino-terminal acetylation as dynamic modifications. Peptides were confidently identified using a target-decoy

approach with a peptide false discovery rate (FDR) of 1% and a protein FDR of 5%. A minimum peptide length of 7 amino acids was required, maximally two missed cleavages were allowed, initial mass deviation for precursor ion was up to 8 ppm and the maximum allowed mass deviation for fragment ions was 0.5 Da. LFQ-based LC-MS/MS experiments were conducted in multiple replicates (three biological replicates each with two technical replicates). For LFQ analysis, a match between runs option was enabled and time window at 0.7 minutes. Data processing and statistical analysis were performed on Perseus (Version 1.5.1.6)[140].

Analysis of functional category and networks of subtype-specific secreted proteins. The biological processes and molecular functions of the G9a-interacting proteins were categorized by Ingenuity Pathway Analysis (IPA) [141] and STRING [142] similar to previously described. [143]

TCGA and METABRIC data sets. All TCGA and METABRIC data used in our study were retrieved from cBioPortal [144, 145] using the 'cgdsr' R package (version 1.2.6)[146]. For our analysis of TCGA data, we used complete samples (case list id = brca_tcga_pub2015_3way_complete / brca_tcga_pub2015_freeze) with mutation, copy-number, and mRNA expression data provided (N = 816) from the TCGA cancer study brca_tcga_pub2015 (Breast Invasive Carcinoma)[147]. The mRNA expression data sets for G9a interactor genes were obtained from TCGA Genetic Profile: brca_tcga_pub2015_rna_seq_v2_mrna_median_Zscores, containing mRNA expression Z-scores compared with diploid tumors (diploid for each gene). The copy-number data sets for G9a interactor genes were obtained from TCGA Gene Profiles: brca_tcga_pub2015_gistic for putative copy-number from GISTIC 2.0 for each gene. Clinical data were obtained from TCGA case list id: brca_tcga_pub2015_3way_complete / brca_tcga_pub2015_freeze, which included information on patient survival, estrogen receptor (ER) status, progesterone receptor (PR) status, and HER2 enrichment (HER2) status. Additional TCGA clinical information, including the PAM50 subtype assigned to each patient, was obtained from "Table S1" in the TCGA

publication.[147] For the METABRIC [148, 149] data we used: case list id = brca_metabric_cnaseq (samples with mRNA, GISTIC, mutational data), gene profile = brca_metabric_mrna_U133_Zscores (for mRNA expression), and gene profile = brca_metabric_cna (for GISTIC data). We used 1866 of the 2051 samples in the METABRIC case list for which there was survival data.

Heatmap construction. The heatmaps of mRNA expression levels SeCEP genes from the TCGA BRCA datasets were constructed using the 'ComplexHeatmap' R package (version 1.12.0)[150]. Hierarchical clustering was performed using the euclidean distance method and the Ward clustering method (option ward.D2 in R's hclust function).

Statistical Analyses. Kaplan-Meier curve plotting and statistical analysis for differences of overall survival (OS) based on mRNA expression of SeCEP genes among TCGA BRCA samples, and distant relapse free survival (DRFS) for the GSE25066 samples, were performed using the 'survival' R package (2.41.3).[151] Kaplan-Meier estimator and log-rank tests were performed using the survival functions Surv, survfit, and survdiff. Cox proportional hazard survival analysis was performed using the survival function coxph. Mann-Whitney-Wilcoxon Test was performed using the wilcox.test function in R. In the iCEP analysis, we included all genes in the TCGA (n = 18,097) or METABRIC (n = 16,555) data sets in order to obtain an adjusted p-value for multiple comparisons for the SeCEP genes. The adjusted p-values were calculated by submitting all the p-values from the individual Mann-Whitney-Wilcoxon Tests determined for each gene to the p.adjust function in R using the 'fdr' method.

RESULTS

The secreto-transcriptomic workflow for discovering candidate biomarkers for non-invasive, individualized prognosis.

As shown in Figure 2.1, LFQ proteomics was first used to determine the compositional differences in the secretomes isolated from different BC subtypes versus non-malignant cells.

The proteins showing BC or PAM50 subtype-specific or subtypic secretion were identified by LFQ-based Perseus analysis.[152] Taking advantage of the databases of two large patient cohorts, TCGA [147] and METABRIC (Molecular Taxonomy of BC International Consortium) [148], which contain clinic-pathologically correlated gene-expression or transcriptomic data, we retrospectively established the proteo-transcriptomic links between BC subtypically secreted proteins and the patient-specific mRNA-expression alterations of the genes that encode these proteins. As a result, this proteo-transcriptomic approach identified the secretome-encoding genes that showed a secretion-correlated mRNA expression pattern (SeCEP), wherein the patient-specific mRNA expression of these genes was positively correlated with increased secretions of the proteins encoded by these genes in similar BC subtypes. This expression-to-secretion correlation not only indicated those genes that are fully translated into extracellularly functional, oncogenically active proteins, but also identified new phenotypic markers that describe the similar PAM50-classified BC subtype. Further, Kaplan-Meier (KM) survival analyses were used to distinguish, from >18,000 genes, patient-specific mRNA expression patterns of select SeCEP genes of prognostic value, indirectly identifying those proteins showing BC subtypic secretion as candidate markers for non-invasive prognostic prediction.

LFQ secretome screening identified particular protein clusters showing BC-subtypic secretion.

We used a similar LFQ proteomic approach [153, 154] to comparatively profile the extracellular proteins secreted from different cell lines, respectively representing the BLBC/TNBC subtypes (MDA-231 and HCC1806), luminal subtypes (MCF-7 and T47D), and non-malignant mammary control (MCF10A). (Figure 2.2a) A total of 2,345 proteins were identified in these five cell lines. Using existing databases of secreted proteins we then examined the purity of our secretome isolation. The analysis of Gene Ontology Cellular Component (GOCC) indicated that 685 or 29% of the totally identified proteins were previously

known for their locations in the extracellular space or plasma membrane. Also, 503 proteins were previously known as secreted or highly likely secreted proteins in the MetazsecKB database that is generated by multiple bioinformatics tools including SignalP4, TMHMM, and TargetP.[155] By comparing our identifications to a number of experimentally identified secretomes 832 proteins were found in common in the secretome from the LPS-stimulated macrophages [154] and 1042 proteins were also identified in a breast cancer secretome.[156] These results in combination validated the high quality of our secretome preparation and analysis.

To identify the proteins showing BC-subtypic secretion we used the Profile Plot function of the Perseus software platform [152] to determine the relative abundances of individual secreted proteins across different cell lines, which correlate with the LFQ ratios of identified proteins. Profile Plot performs pattern matching but does not perform statistical testing on the identified proteins, therefore the statistical significance of protein abundance changes between BC subtypes was validated by one-way ANOVA. As a result we identified clusters that contain the proteins showing increased or decreased secretion only in either BLBC- or luminal-subtypic cell lines, respectively. (Figure 2.2 b-e)

For example, we identified 55 proteins as having BLBC-specific secretion in both BLBC cell lines, including 35 proteins with increased secretion and 20 proteins with decreased secretion compared to the luminal and non-malignant control cell lines. Meanwhile, there were 86 additional proteins showing either increased or decreased secretion in one of the BLBC cell lines. (Figure 2.3a) In view of BC-related function of these BLBC-specifically secreted proteins, several factors involved in tumor progression and metastasis showed increased secretion, including CD44, [157] HSPA5, [158] and HSP90B1.[159] Meanwhile, some proteins such as E-cadherin (CDH1) and DDB1 that were known to be down-regulated in BLBC [160, 161] showed BLBC-specific reduction in secretion.

Similarly, we identified a total of 274 proteins that showed luminal-specific secretion changes in one or both of the luminal cell lines (Figure 2.3b), including decreased secretion of several members of the cathepsin family of globular proteases such as CTSB, CTSL, and CTSZ that were known to promote breast cancer progression and metastasis, [162] and the increased secretion of metastasis suppressor PEBP1 (a.k.a. RKIP) that showed luminal-specific intracellular expression.[163]

Immunoblotting of some of BC-subtypic secreted proteins showed consistent results with LFQ secretome screening (Figure 2.4).

Using bioinformatics tools including Ingenuity Pathway Analysis (IPA) [141] and the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [142] we then studied the biological processes and pathways in which these BC-subtypic secreted proteins are involved. In the IPA annotation, greater than 95% of all identified BC-subtypic secreted proteins, *i.e.,* 136 of 141 BLBC-specific and 269 of 274 luminal-specific secreted proteins were respectively cancer-related.

Figure 2.5a shows the biological processes that are over-presented by the BLBC- and luminal-specific proteins. Although major biological processes are comparable between subtypes, more detailed analysis of these broad categories highlighted the differences between subtype characteristics. BLBC-specific proteins were involved in increased cell movement or migration, invasiveness of breast cancer cells, and cell survival, while luminal-specific proteins were associated with decreases in cell movement and vascularization, indicating the aggressiveness differences between these two BC subtypes (Figure 2.5b).

IPA analysis of BLBC-specific secreted proteins (Figure 2.5c) indicated the activation of a few known BC-driving signaling pathways, including PI3K-Akt signaling, [164] protein kinase A signaling, [165] signaling by Rho family GTPases, [166] the 14-3-3-mediated signaling associated with BC oncogenesis, [167] and the actin cytoskeleton signaling involved in the epithelial-mesenchymal transition (EMT).[168] Meanwhile, the altered secretion of other BLBC-

specific proteins indicated that the activity of the HIPPO signaling was suppressed in BLBC cells, which could lead to a more invasive tumor phenotype.[169] On the other hand, the luminal-specific secreted proteins revealed activation of HIPPO and mTOR signaling along with the suppression of eIF2 signaling, G2/M DNA damage checkpoint regulation, and ILK signaling (Figure 2.5c).

To further determine the functional networks involving BC subtypic secreted proteins we performed protein-protein interaction (PPI) analysis using STRING, which revealed statistically significant enrichment of PPIs among the proteins secreted in both BLBC-specific ($p<1e-16$) and luminal-specific ($p<1e-16$) manners (Figure 2.6). The Gene Ontology Biological Process enrichment of the proteins with BLBC-specific increased secretion identified multiple subnetworks associated with protein folding, regulation of cell communication, regulation of apoptosis, cell development, regulation of cell motility, blood coagulation, and proteolysis. Analysis of the proteins with decreased secretion in BLBC cells also revealed particular subnetworks/pathways with suppressed activities, including DNA damage response, regulation of actin depolymerization, and regulation of cell-cell adhesion. In contrast, the proteins showing luminal-specific increases of secretion over-represented the subnetworks associated with regulation of growth, and cell differentiation while the proteins with decreased secretion in luminal cells were involved in positive regulation of apoptotic process, angiogenesis, extracellular matrix disassembly, and cell motility. These results showed that the proteins secreted or secretomes are characteristic of distinct BC subtypes.

The genes that encode increasingly secreted proteins showed secretion-correlated mRNA over-expression patterns in BC patients in a PAM50-subtypic manner.

To determine the clinicopathological relevance of the proteins showing BC subtypic secretion**,** in the databases of TCGA [147] and METABRIC [148] we retrospectively examined patient mRNA expression patterns for the genes that encode the proteins showing either BLBC-

or luminal-specific secretion. These databases contain large cohorts of > 2600 BC patients that were classified by PAM50 as the BLBC/TNBC, luminal-A and luminal-B, Her2+, and normal-like BC subtypes, [170] and the information about mRNA expression, mutations, copy-number variations, and associated clinical/pathological data (stages/grades and relapse status).

First, to determine the mRNA expression differences between BLBC and luminal A/B TCGA patients we performed a Mann-Whitney-Wilcoxon Test on the z-scored expression values (downloaded from the cBioPortal for Cancer Genomics [144, 145]) of the two PAM50-subtypic populations. This test was performed on all secreted protein-encoding genes in each subtype-specific dataset, and p-values were adjusted by the Benjamini Hochberg procedure for multiple testing. Using this multi-testing scheme, a secreted protein-encoding gene was classified as BLBC if the expression level between the two PAM50-subtypic populations was significantly different (adjusted p-value < 0.05) and the median mRNA expression was greater among BLBC patients. Likewise, a gene was classified as luminal if its median mRNA expression level was higher for luminal patients and the gene showed a statistically significant difference between luminal and BLBC patients (adjusted p-value < 0.05).

On a systems view, a heat map of unsupervised hierarchical clustering showed patient mRNA expression patterns for those genes that encode the proteins demonstrating PAM50-subtypic secretion in TCGA patients (Figure 2.7). In a statistically significant manner, we identified a secretome-to-patient transcriptome or secreto-transcriptomic link for some genes that encode PAM50-subtypic secreted proteins, *i.e*., we found a secretion-correlated mRNA overexpression pattern or SeCEP wherein the PAM50-subtypic secretion of some proteins showed *cis*-mRNA expression of their encoding genes in patients with the corresponding PAM50-subtypes (Figure 2.8). For example, mRNA overexpression of 57 genes that encode BLBC-specific secreted proteins clustered BLBC patients while 60 genes that encode luminal-specific proteins with increased secretion showed mRNA overexpression in luminal patients. These results indicated that these secretome genes are fully translated into the oncogenically

active secretome in a BC-subtypic manner.

Further, we observed interpatient heterogeneity in the mRNA expression pattern of secretome genes within each PAM50 subtype, i.e., not all SeCEP genes were simultaneously overexpressed at the mRNA level in each individual BLBC or luminal patient. Bearing in mind that mRNA expression patterns of PAM50 genes are insufficient to stratify the patient subsets with different clinical outcomes or prognoses, we reasoned that, within a single PAM50-classified subtype, these patient mRNA expression variations of select SeCEP genes can mark the patient subpopulations with distinct prognoses.

Patient-specific mRNA co-overexpression patterns of select secretome-encoding genes mark the high-risk subpopulations of PAM50-subtypic patients with poor prognosis.

To identify BC-subtypic secreted proteins of prognostic significance, we performed Kaplan-Meier (KM) analysis on PAM50-subtypic patients in the two independent datasets TCGA and METABRIC [148, 149] for any combination of up to five SeCEP genes having mRNA overexpression (z-score > median z-score) for all genes in the combination. The statistical significance of each gene combination was determined by a multi-parameter threshold including log-rank p value < 0.05 and lower 95 confidence interval for the hazard ratio > 1 in both the TCGA and METABRIC datasets.

For example, we identified approximately 8% or more subpopulations of BLBC patients who showed mRNA co-overexpression of four BLBC-specific SeCEP genes, YWHAZ, GDA, MFAP2, and PRKCSH in correlation with poor survival (Figure 2.9 a,b). YWHAZ, which encodes the 14-3-3ζ protein, was characterized as a promoter of cell survival which, when overexpressed, is associated with poor prognosis and disease-free survival.[171, 172] Another SeCEP gene combination indicating the co-overexpression-correlated poor prognosis was ADM, PSMB6, SERPINH1, and SFN (Figure 2.9 c,d). ADM was known to promote angiogenesis, cell survival, and metastasis, [173, 174] and was associated with poor prognosis

in ovarian cancer patients.[175] Interestingly, although SFN (14-3-3σ or stratifin) was considered as a tumor suppressor, overexpression in BLBC was reported.[176] Recently, overexpression of SFN was found to be associated with tumor invasion and migration.[177] Another BLBC subpopulation showed co-overexpression of GAL, MMP12, MSLN, and a multifunctional oncoprotein SET.[178] (Figure 2.9 e,f).

Similarly, this secreto-transcriptomic approach enabled identifications of the distinct subpopulations of luminal patients with poor prognosis. Further, as an example of how the co-overexpression of multiple SeCEP genes improves the specificity and sensitivity in predicting personalized prognosis, as shown in Figure 2.10 a,b, overexpression of CLEC3A alone indicated modest differences in the overall survival rate of two major luminal patient subpopulations. However, the luminal patient subsets showing co-overexpression of CLEC3A with CTTN, IGFBP5, NRCAM were statistic-significantly correlated with worse prognosis and can be readily discriminated from other luminal patients (Figure 2.10 c,d). Among these SeCEP genes, CLEC3A is a C-type lectin that promotes tumor adhesion in breast cancer [179] and was recently found to enhance plasminogen activation by tissue-type plasminogen activator.[180] CTTN encodes cortactin, an actin cytoskeleton regulator that promotes metastasis in breast cancer.[181] Meanwhile, co-overexpression of CLEC3A with ALDOA, EEA1, and FKBP4 was also associated with substantially worse prognosis than CLEC3A alone. (Figure 2.10 e,f)

Importantly, the use of co-overexpressed SeCEP genes can further resolve individual luminal subtypes among luminal patients to identify the high-risk subpopulations of luminal-A or luminal-B patients. For example, the luminal-A subpopulations overexpressing CAPZA2, FKBP4, KRT18, and OLFML3 exhibited poor prognosis with decreased overall survival, but this combination did not distinguish any subpopulation of luminal-B patients or combined luminal A/B groups (Figure 2.11 a,b). Also, co-overexpression of FASN, IGFBP5, ISOC1, and PIP was likewise specific to the luminal-A group (Figure 2.11 c,d). Interestingly, although each of these genes has been reported to play a role in breast cancer development, progression, or

metastasis, [182-185] over-expression of these genes individually did not provide subtype-specific prognostic value. Overall, we found 52 gene combinations with co-overexpression that showed poor survival among luminal-A patients but not in other BC subtypes.

Several gene co-overexpression patterns specifically correlated with luminal-B patient prognosis were also identified. In the co-overexpression pattern involving HSP90B1, EEF1A2, EIF4B, and KRT18, (Figure 2.12 a,b) HSPB1 was known to play a role in epithelial-mesenchymal transition and tumors overexpressing HSPB1 demonstrated enhanced drug resistance.[186] Similarly, luminal-B patients overexpressing a combination of AGR2, CYFIP2, KRT18, and RAB1B exhibited worst overall survival while luminal-A patients and the combined luminal A/B group showed no significant differences in survival (Figure 2.12 c,d). In total we identified 39 gene combinations that, when overexpressed, indicated poor overall survival specifically among luminal-B patients. Our combined results demonstrated that patient-specific co-overexpression of SeCEP genes can resolve the interpatient heterogeneity within different PAM50-subtypes, confirming that these gene expression alteration patterns are prognostically meaningful in distinguishing the subsets of BLBC or luminal patients with distinct clinical outcomes with multi-testing of large patient cohorts.

Notably, unsupervised hierarchical clustering mRNA expression of luminal and basal SeCEP genes using the TCGA patient cohort revealed that HER2-overexpressing or -enriched patients did not cluster together but were interspersed among primarily luminal A/B patients. (Figure 2.7) This result implied that various clinical outcomes of HER2-enriched patients could be represented by select luminal SeCEP genes. We therefore searched for altered mRNA expression patterns of luminal SeCEP genes in correlation with destinct clinical outcomes of HER2+ patients. Generally, patients with the HER2-enriched subtype show overall poor survival similar to BLBC patients.[70] In KM analysis, we identified unique gene combinations associated with poor survival among HER2-enriched patient subpopulations which were not prognostically significant among luminal patients. For example, HER2-enriched patients showing co-

overexpression of CAPZA2, CBX1, G6PD, and NQO1 had worse survival (Figure 2.13 a,b). NQO1 was highly expressed in BC patients with high HER2 expression and was linked to increased metastasis.[187] High expression of CYFIP1, DDR1 and GYG1 was also associated with worse survival (Figure 2.13 c,d), and DDR1 was linked to BC invasion and drug resistance.[188, 189] Another HER2-enriched subpopulation with poor survival showed co-overexpressed G6PD, CYFIP1, PSMC2, and KYNU (Figure 2.13 e,f), the latter of which has been implicated in increased metastasis and tumor aggressiveness.[190] In sum, these results indicate that altered mRNA expression patterns of select luminal SeCEP genes can be used to distinguish the distinct subpopulations of HER2-enriched patients with poor prognosis.

More importantly, the majority of the genes encoding BLBC- or luminal-specific secretome in networks showed statistically significant, secretion-correlated cis-mRNA expression in some BC patients. Further, by identifying their co-overexpressed patterns in BC-subtypic patients, we revealed the pathological or prognostic significance of these secreted proteins in multiple interactive sub-networks (Figure 2.14a). Strikingly, the majority of the BLBC-specific proteins involved in the interactive subnetworks associated with unfolded protein response, cell migration, and negative regulation of cell death. Specifically, the glycoprotein THBS1 promoted BC invasion and metastasis and was associated with disease recurrence in BC patients.[191, 192] Similarly, higher serum levels of metallopeptidase inhibitor TIMP-1 were associated with increased likelihood of BC metastasis.[193] Further, the disulfide isomerase PDIA6 promoted tumor immune evasion [194] and enhanced cell proliferation by activating Wnt/β-catenin signaling.[195] Three of four genes in the combination of CORO1C, MSN, ICOSLG, and HIST1H1B are in this subnetwork, and KM analysis reveals a significant decrease in the overall survival rate of BLBC patients overexpressing these genes. (Figure 2.14 b,c)

Luminal-specific subnetworks were also identified, (Figure 2.15a) however there was no biological process enrichment observed. These oncogenically active interacting proteins included BAG3 which reduced BC cell adhesion and increased motility.[196] This network also

included RHOC, a small GTPase that regulates cytoskeletal architecture [197] and is associated with increased rates of metastasis.[198] SNCG, a neuronal protein overexpressed in BC was also associated with higher likelihood of metastasis.[199] TCGA and METABRIC patients exhibiting a four gene co-overexpression pattern involving SNCG, CLEC3A, DNPEP, and KRT18, three of which are members of the interacting subnetworks, had lower overall survival rates (Figure 2.15 b,c). Together, these results indicate the coordinated, extracellular oncogenic activity of the networked proteins.

Secreto-transcriptomic analysis identified patient-specific co-overexpression patterns of select secreted proteins as prognostic markers to predict personalized response to therapy.

Nether the TCGA nor the METABRIC study was designed to answer specific clinical questions. To assess the clinical significance of altered mRNA expression of multiple SeCEP genes in predicting the response to specific therapeutic interventions, we next looked for distinct combinations of SeCEP genes showing statistically significant changes in distant relapse free survival (DRFS) among patients receiving neoadjuvant taxane-anthracycline therapy in the clinical trial GSE25066.[200] Following similar procedures to those described above for TCGA and METABRIC, we performed KM analysis on the BLBC-SeCEP genes in combinations of up to five genes having mRNA overexpression (z-score > median z-score) for all genes in the combination.

Among BLBC patients, we found 12 combinations with >10% of both GSE25066 and TCGA patients overexpressing each gene in the combination and having a significant difference in DRFS. Examples are shown in Figure 2.16 a,b. One such combination was ANXA2, CALR, MFAP2, and SERPINH1. ANXA2 has been reported as an independent predictor of poor prognosis in breast cancer patients receiving neoadjuvant therapy, [201] however overexpression of ANXA2 alone did not have a statistically significant impact on DRFS among

GSE25066 patients. Likewise, co-overexpression of ADM, MAGEA4, and PRKCSH was also associated with a statistically significant change in DRFS. Similar analysis of luminal-SeCEP gene combinations yielded five combinations with at least 10% of patients in both the GSE25066 and TCGA datasets overexpressing all genes in the combination and p value < 0.05. One combination was BLVRB, EIF4B, and ISOC1 (Figure 2.16c). Importantly, BLVRB is associated with the development of chemotherapeutic resistance, though overexpression of BLVRB alone did not predict worse patient outcomes.[202]

Overall, these analyses identified the subpopulations within each PAM50 subtype with resistance to neoadjuvant anthracycline-taxane therapy along with the correlation to their poorer overall survival. Thus, we demonstrate the potential clinical uses of the analysis to aid the clinician in determining the appropriate therapeutic intervention to be employed.

DISCUSSION

The development of a novel secreto-transcriptomic approach underlies our innovation in the identification of liquid biopsy biomarkers capable of discriminating between patient subpopulations having variable outcomes. Recognizing that single-omics approaches are insufficient for making these distinctions, due to the negligible data on the oncogenic phenotype provided by genomics/transcriptomics and the inadequate phenotypic coverage of patient-specific proteomics offers, our secreto-transcriptomic workflow is a multi-omic integrated method which offers a robust and efficient scheme to distinguish patient subpopulations within each BC subtype. First, by using a LFQ-based nanoLC-MS/MS approach for secretome profiling, [153, 154] we comparatively analyzed the compositional differences in the extracellular proteins secreted from a series of BC cell lines representing various BC- or PAM50-subtypes. We then determined the clinicopathological relevance of the proteins showing subtype-specific or subtypic secretion by a retrospective proteo-transcriptomic analysis [203] of the BC patient datasets (> 2600 patients) from The Cancer Genome Atlas (TCGA) [147] and METABRIC

(Molecular Taxonomy of BC International Consortium). [148] We found a PAM50-subtypic Secretion-Correlated mRNA Expression Pattern (SeCEP) wherein the PAM50-subtypic secretion of some proteins showed statistically significant *cis*-mRNA expression of the genes that encode them in patients with the corresponding PAM50-subtypes. This expression-to-secretion correlation highlighted those genes that are fully translated into the oncogenically active secretome in a PAM50-subtypic manner. Further, we observed that patient-to-patient mRNA expression variations of individual secretome genes describe the interpatient heterogeneity within each single PAM50 subtype. In this regard, patient-specific co-overexpression of distinct SeCEP genes were found in correlation with specific prognoses within distinct subsets of BLBC or luminal-A and luminal-B patients. Currently, available blood-based tests for cancer prognosis or diagnosis are often based on a single gene or protein marker, therefore lacking the specificity and sensitivity in determining individualized clinical outcomes.[204] Our identification of multi-gene or multi-protein panels as systems signatures can precisely describe the predominant tumor phenotype with significantly improved phenotype accuracy. Because our workflow starts with the identification of tumor-phenotypic alterations and work back to the genotypic data with the coverage of patient-specific alterations, we are able to bypass the need for extensive modeling [205] or analysis of a large number of patients [206] by pinpointing a few prognostically significant marker genes.

Our dissection of the BC-subtypic secretomes highlights the differences between subtype-characteristic extracellular functions reflecting the divergent underlying pathologies of each subtype. We found that although both luminal and basal secreted proteins fall into the same broad categories (e.g. cell motility), the functional roles of these proteins are significantly different between subtypes. For example, cellular movement was a highly overrepresented category in the secretomes of both subtypes, but the basal-specific proteins were promotive of cell motility while the luminal-specific proteins were inhibitive, which is in line with the more

aggressive nature of BLBC subtype in general. Similarly, pathway activation analysis found the same pathways enriched in both subtypes, but with opposite activation states.

We identified multiple over-secreted proteins exhibiting a BC-subtypic SeCEP consistently in both TCGA and METABRIC databases with > 2600 BC patients which constitute the subtype-specific fully translated oncogenic-active secretome. Additionally, the co-overexpression of multiple SeCEP genes in unique combinations was prognostic of differential survival rates of subpopulations within each PAM50-subtype. Further, these co-overexpressed gene combinations were distinct for each subtype, *i.e.* combinations showing decreased overall survival in one subtype did not exhibit altered survival rates in other subtypes.

There are some important considerations to note in this study. First, due to the significant breast tumor heterogeneity our work cannot, and is not intended to, identify all of the secreted proteins relevant to the characteristics of a given tumor. We used breast cell lines as a model system to identify potential markers and reinforce these identifications with a broad set of patient data. In order to efficiently connect patient outcomes to potential markers, we must select practical criteria. The high expression/high secretion correlation provides a reasonable and straightforward link between the secretome and the transcriptome. Alternate expression/secretion patterns are observed, however these are harder to quantify and correlate. Importantly, it is not necessary to identify every gene combination of relevance in order to identify specific patient subpopulations with poorer outcomes. The present study also has some limitations which preclude the ability to identify all such subgroups, including the number of basal and luminal cell lines examined and the exclusion of the HER2-enriched subtype from the secretome analysis. Despite these limitations our work names several noteworthy gene combinations which define specific patient subpopulations, but more importantly provides a template for the further identification of combinations defining other subgroups.

Because multiple SeCEP genes showing prognostic-significant mRNA co-overexpression in a marker panel were identified as over-secreted proteins in a single BC subtype, these gene-

coded proteins are putative liquid biopsy markers to distinguish high-risk populations within PAM50-subtypic classification. Importantly, we also demonstrated the clinical utility of this method in identifying patient subpopulations with the worst outcomes in response to specific therapeutic interventions.

CONCLUSIONS

In summary, our novel secreto-transcriptomic method efficiently and precisely delineated high-risk subpopulations within each PAM50-subtype by linking oncogenically secreted proteins to patient-specific transcriptomic alterations that correlate with distinct clinical outcomes. This multi-omics approach leverages the discrimination of a few tumorigenic/oncogenic alterations in broad transcriptomic profiles of > 18,000 genes, which provide an advantage over any single omics approaches. These multi-gene prognostic markers offer individualized specificity and sensitivity which may guide the clinician to optimize the treatment plan for distinct patient subsets in blood test.

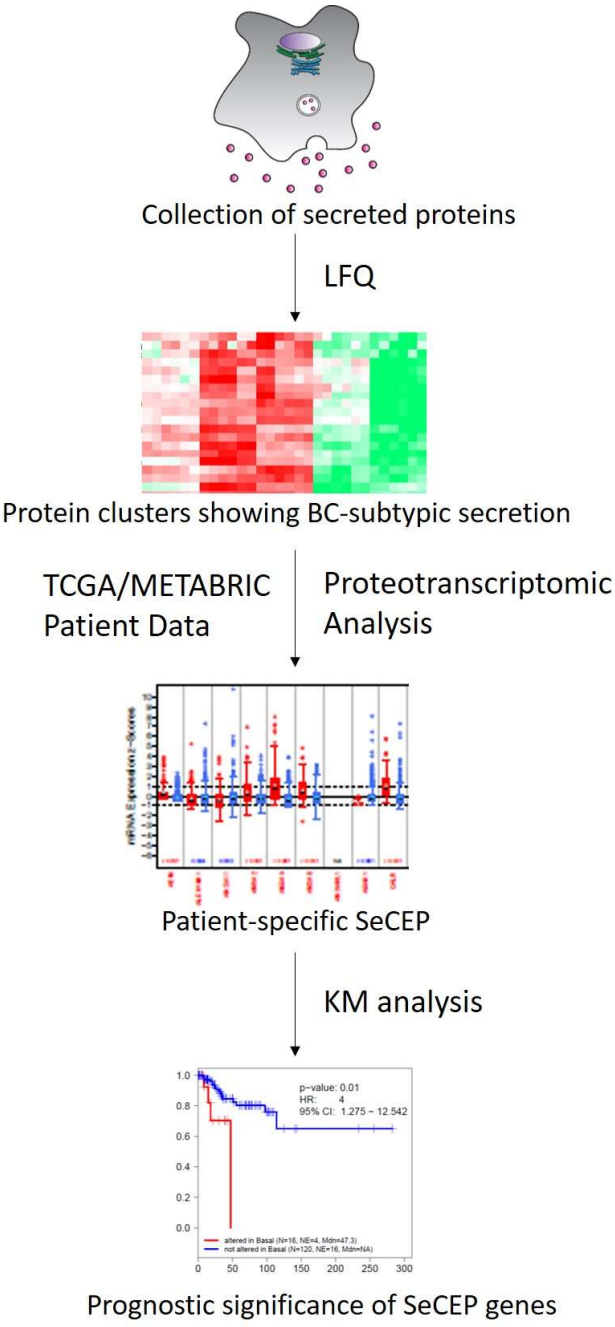


Figure 2.1 - Schematic of secreto-transcriptomic approach for identifying putative liquid biopsy prognostic markers. Label-free quantitative (LFQ) proteomics is used to resolve the compositional differences in cell line secretomes. Then, TCGA/METABRIC patient data are analyzed to determine the PAM50-subtypic transcription of genes encoding the proteins exhibiting either BLBC- or luminal-specific secretion. Kaplan-Meyer (KM) analysis is performed on combinations of genes showing a secretion-correlated mRNA expression pattern of increased transcription positively correlating to increased secretion to identify putative liquid biopsy markers for subpopulations within each subtype having reduced overall survival.

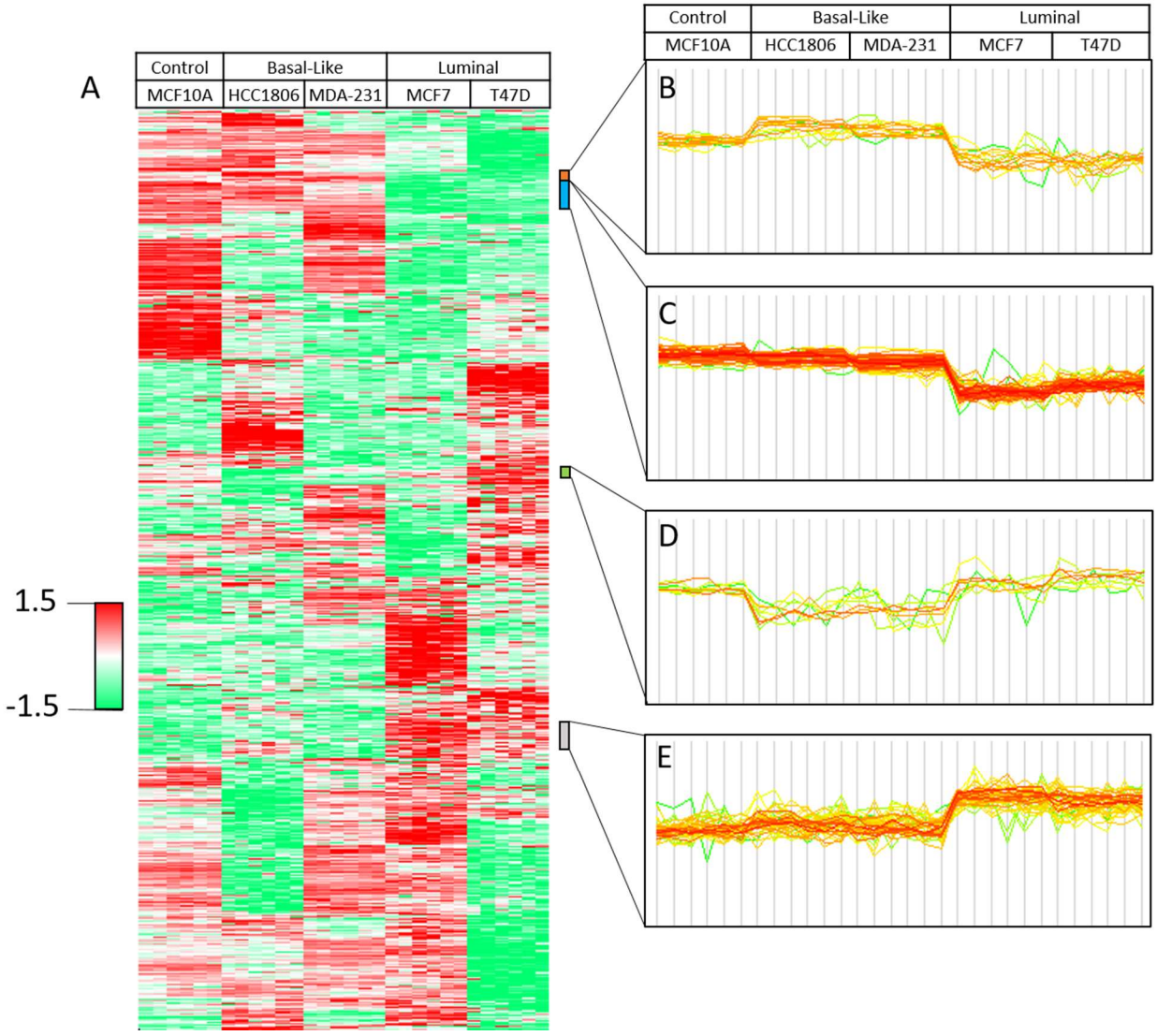


Figure 2.2 - A) A heatmap of unsupervised hierarchical clustering analysis of z-score normalized protein secretion by one non-malignant control (MCF10A), two basal breast cancer (MDA-231 and HCC1806), and two luminal breast cancer (MCF7 and T47D) cell lines highlights intra-subtype heterogeneity. Each cell line is represented by 3 biological replicates and 2 technical replicates. Red indicates higher secretion, green indicates lower secretion, and white indicates mean secretion. Profile plots of 4 representative clusters are shown which demonstrate proteins B) higher in BLBC, C) lower in luminal BC, D) lower in BLBC, and E) higher in luminal BC. Each line represents one protein and the color indicates the density of proteins with similar expression levels.

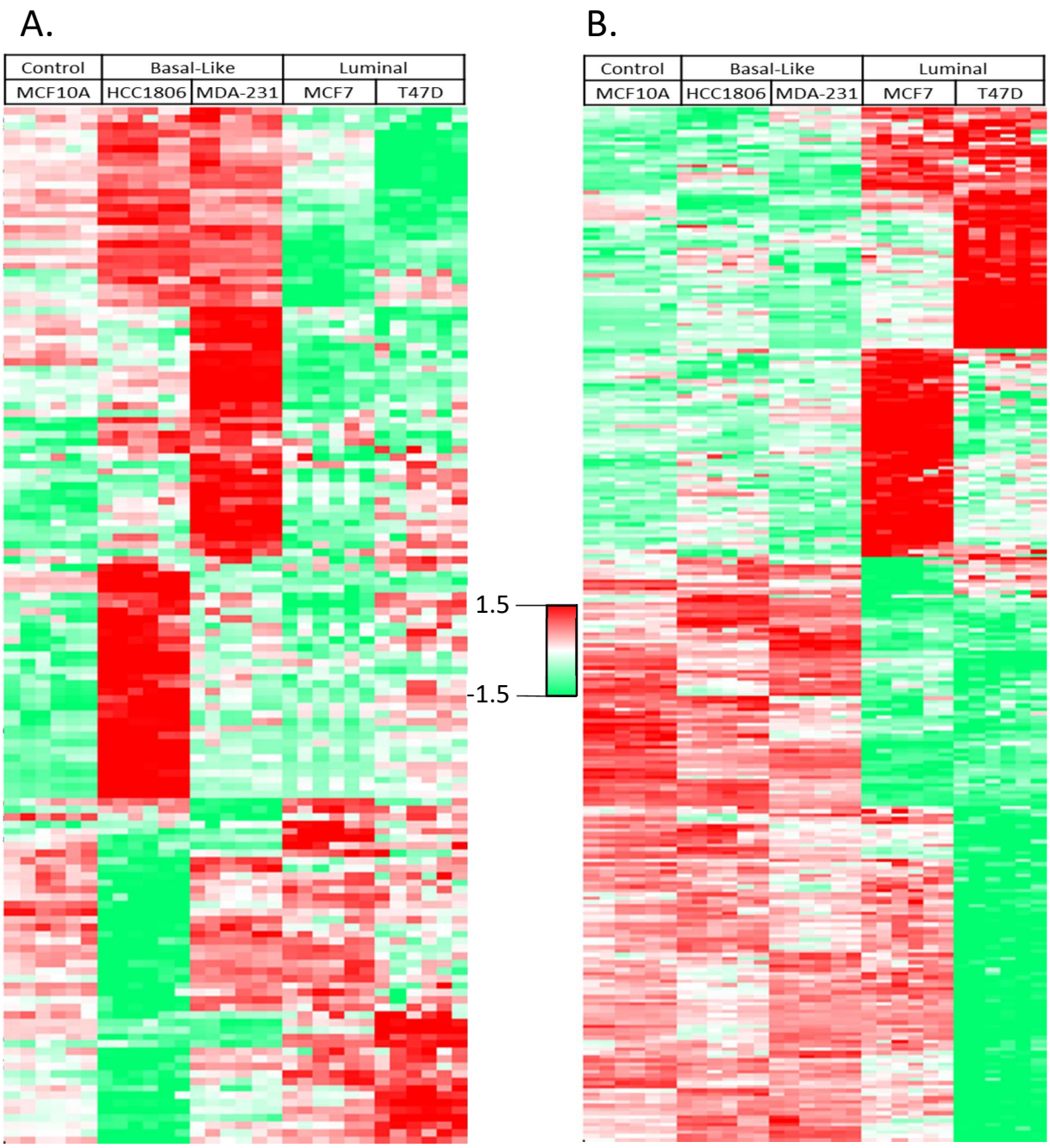


Figure 2.3 - A heatmap of unsupervised hierarchical clustering analysis of the z-scored basal-specific (A) and luminal-specific (B) proteins (rows) secreted by five cell lines (columns). Each cell line is represented by 3 biological replicates and 2 technical replicates. Red indicates higher secretion, green indicates lower secretion, and white indicates mean secretion.

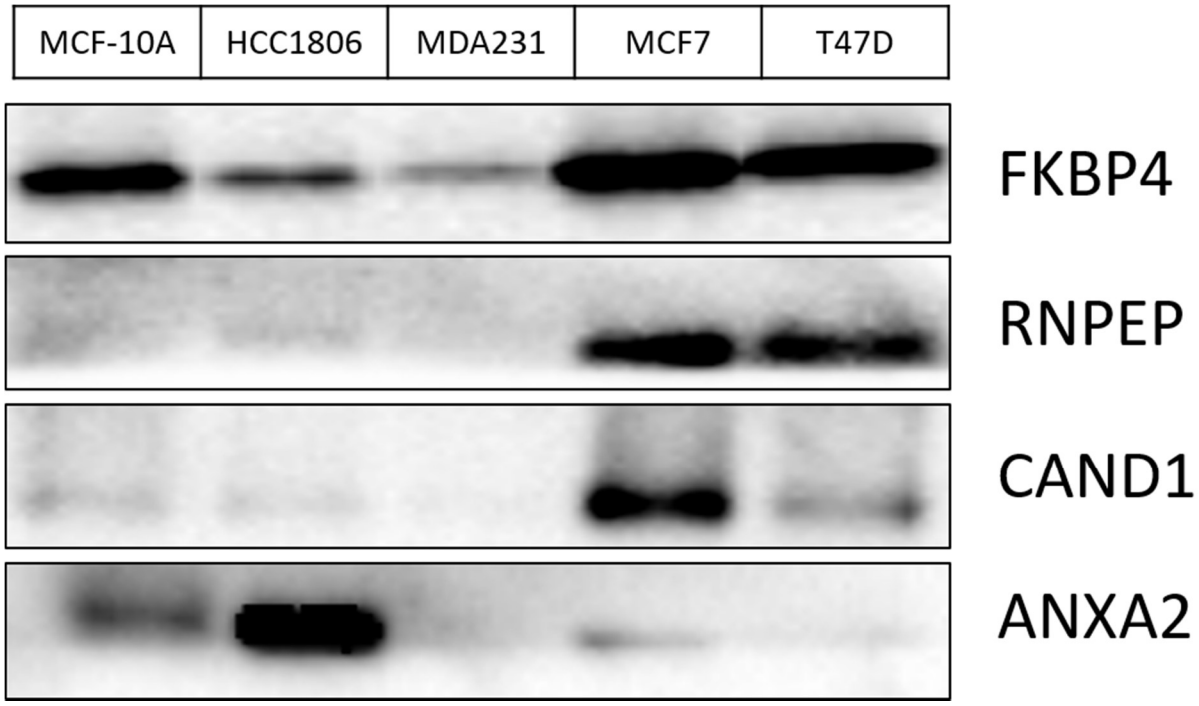


Figure 2.4 - Western blot validation of LFQ data.

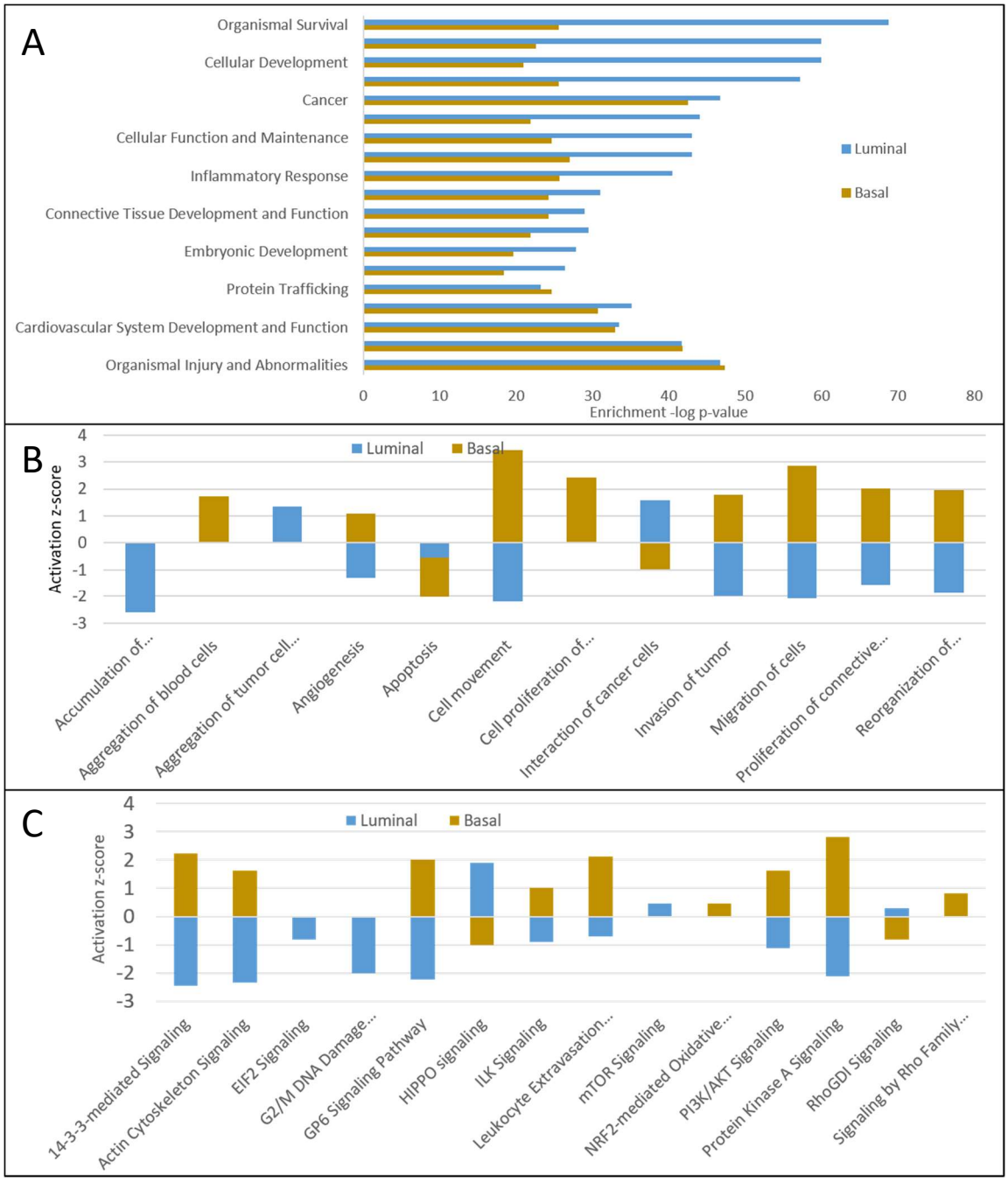


Figure 2.5 - A) Over-represented biological processes in the basal-like and luminal BC subtypes. Bars representing the negative log p-values of BLBC process enrichment are displayed in orange and luminal values are in blue. B) Biological functions activated (positive z-score) or suppressed (negative z-score) in the BLBC (orange) and luminal (blue) PAM50 subtypes. C) Pathway activation analysis of PAM50 subtypes. Orange bars represent BLBC values and blue bars represent luminal values. Positive z-scores indicate pathway activation; negative z-scores indicate pathway suppression.

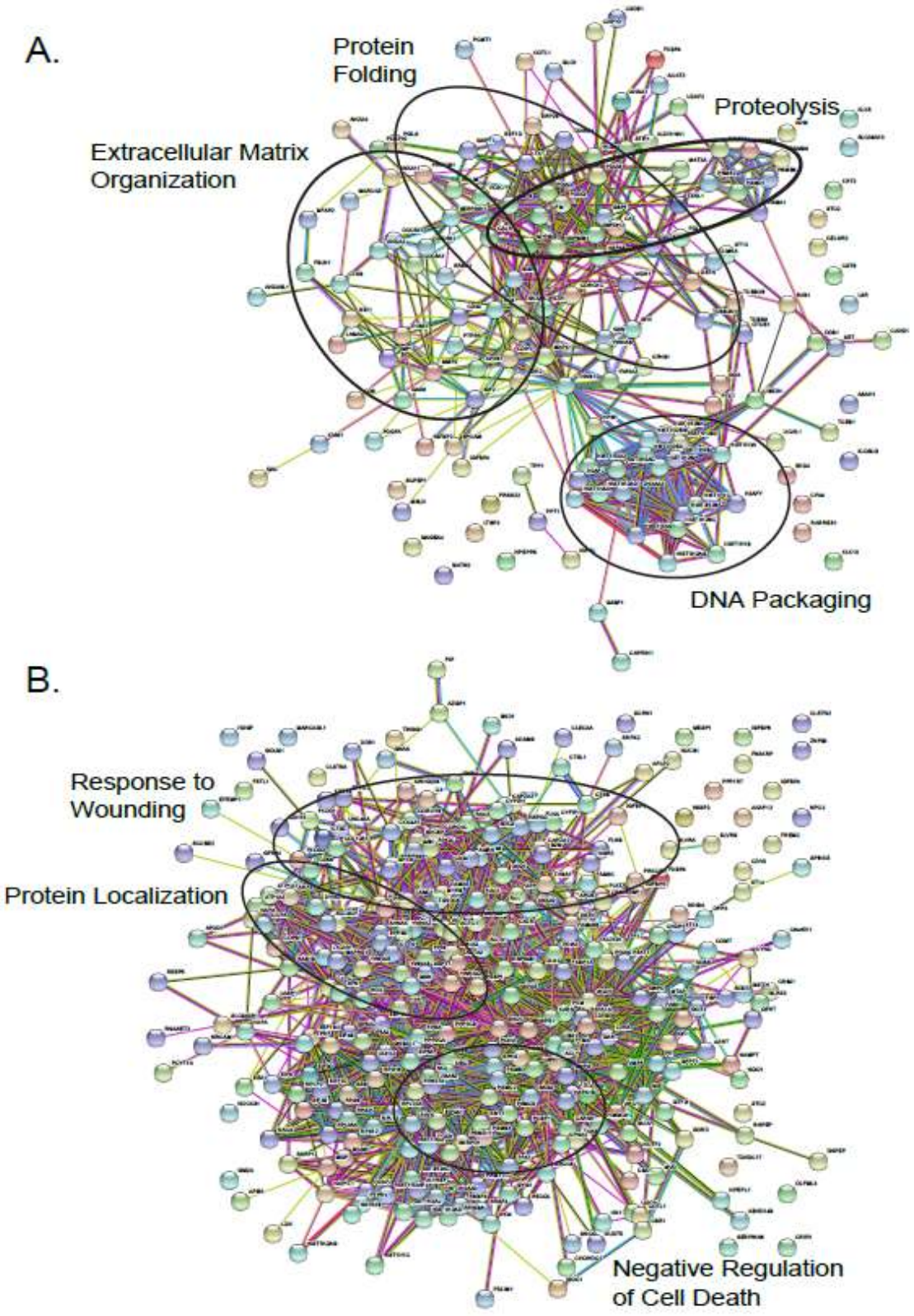


Figure 2.6 - Protein-protein interaction analysis of A) basal-specific and B) luminal-specific secreted proteins.

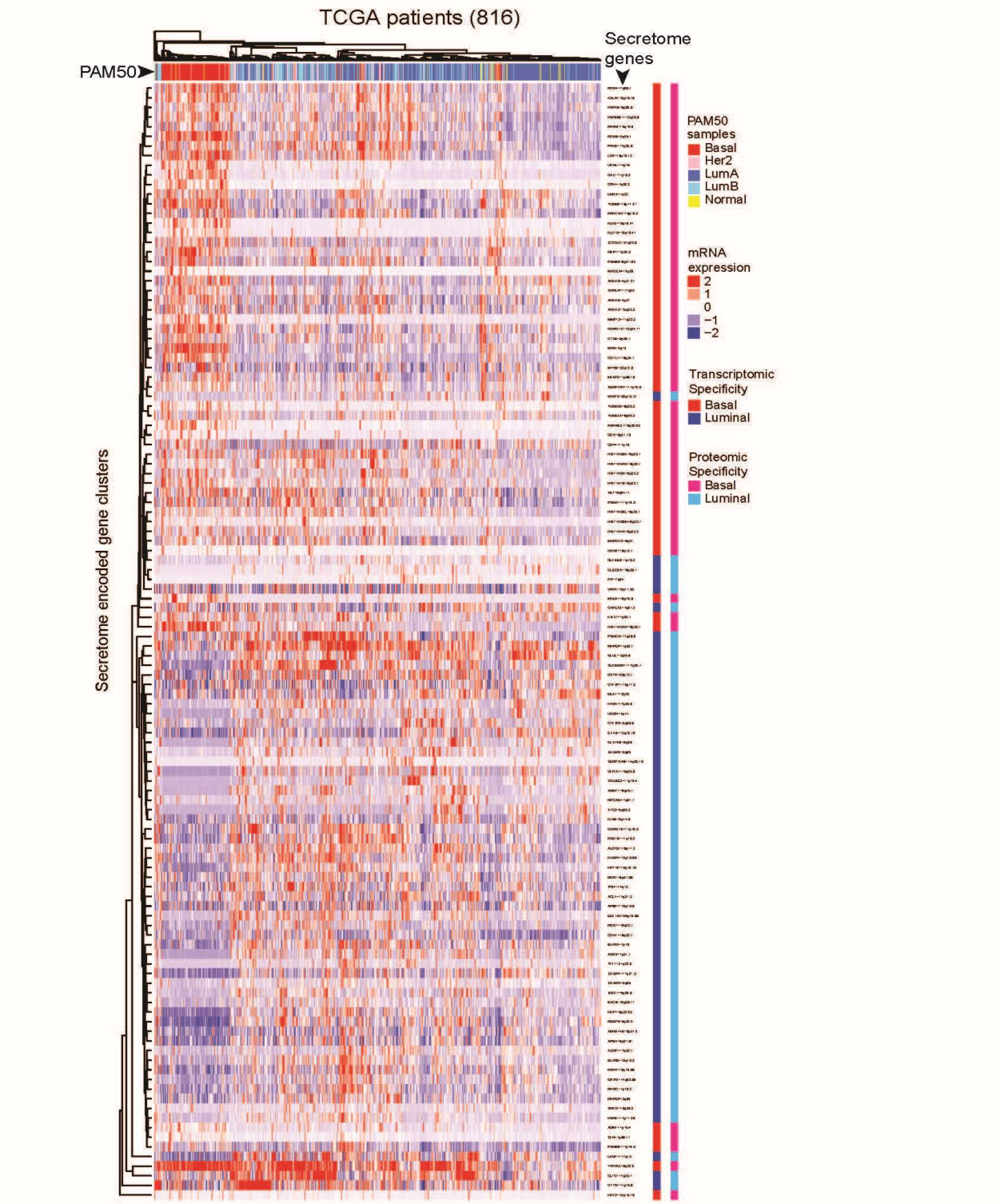


Figure 2.7 - Heatmap of the unsupervised hierarchical clustering of mRNA expression of SeCEP genes (rows) in the TCGA patient cohort (columns). The PAM50 subtype of each patient is indicated by the row above the heatmap. Proteomic (left) and transcriptomic (right) subtype-specificity are indicated by the columns to the right of the heatmap. Within the heatmap red represents higher expression, blue represents lower expression, and white represents mean expression.

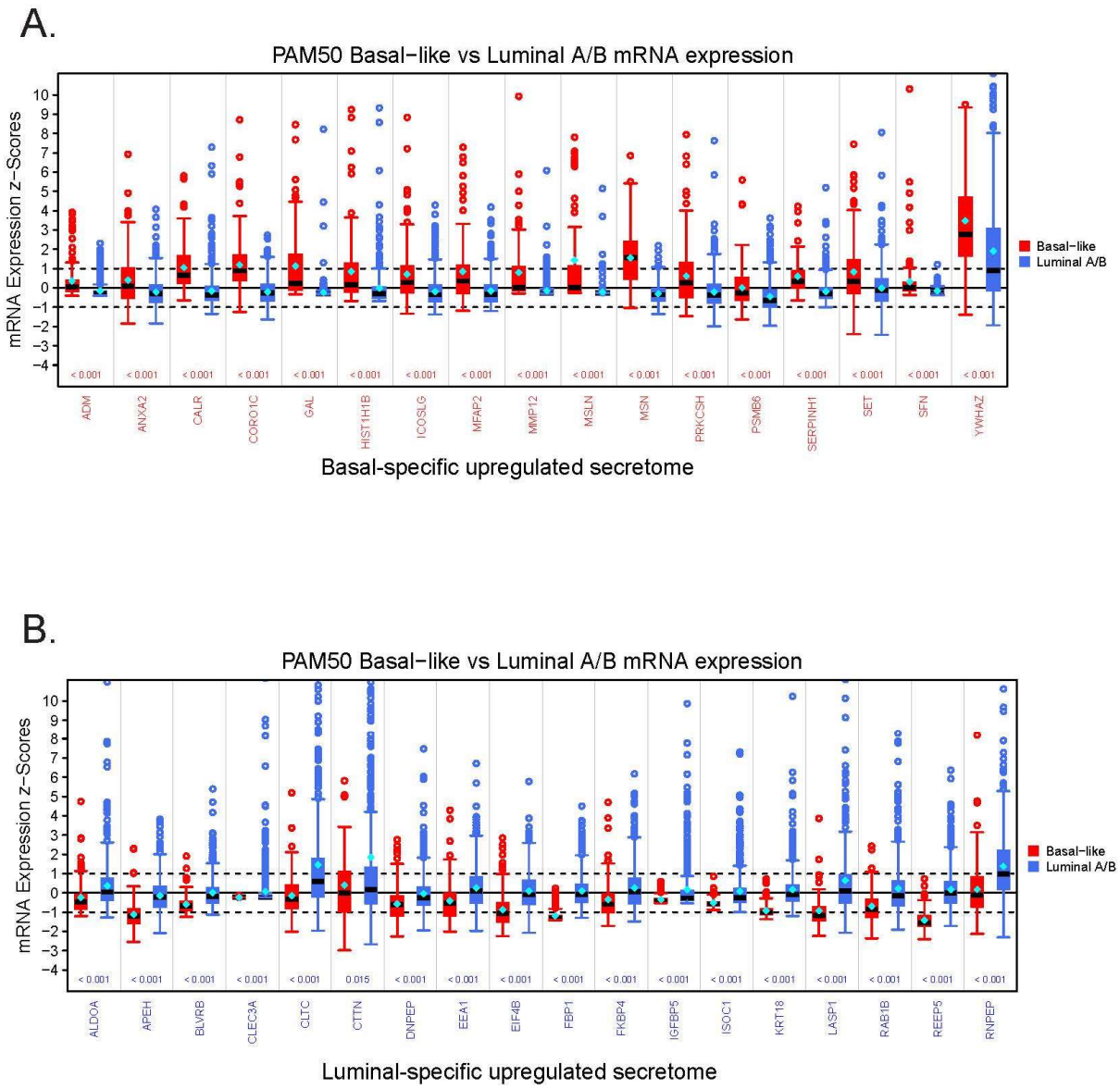
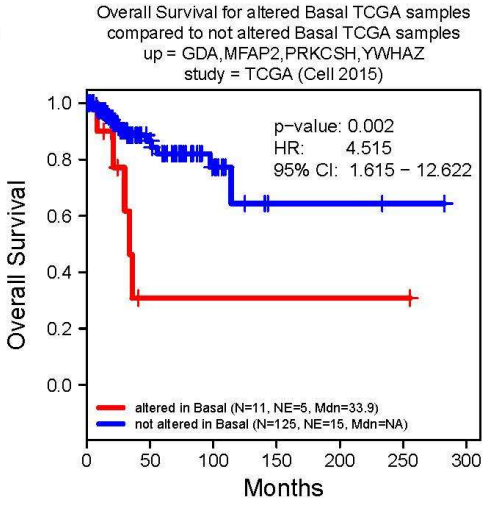
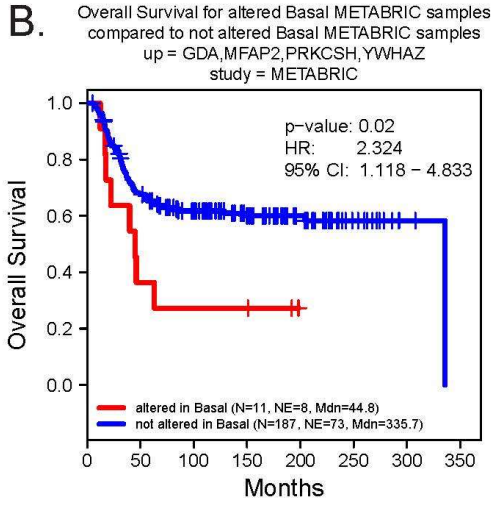


Figure 2.8 - Box plots showing the statistically significant altered mRNA expression (x-axis) for BLBC and luminal secreted proteins among TCGA patients. The distribution of mRNA expression for BLBC patients is shown in red on the left, and for luminal A/B patients in blue on the right. The z-scored mRNA expression is displayed on the x-axis. Values > 1 are considered to be significantly up-regulated, values < -1 to be significantly downregulated, and values between -1 and 1 are considered “not altered”. The median value is displayed as a black bar inside the box. A Mann-Whitney-Wilcoxon Test to ascertain the expression differences between the two PAM50-subtypic populations was performed. The p-value is displayed above the x-axis with p-values < 0.05 colored red if expression is higher for BLBC and blue if expression is higher for luminal samples.

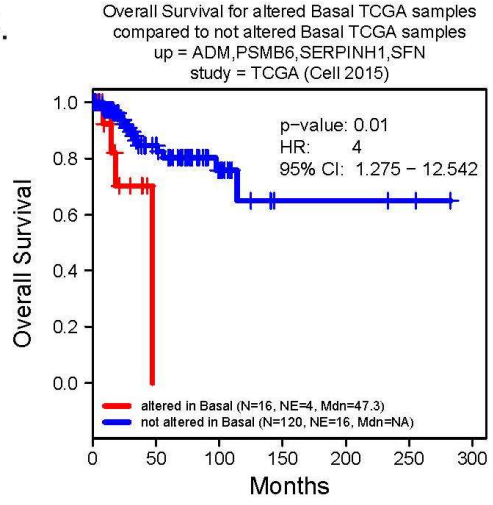
A.


Overall Survival for altered Basal TCGA samples
compared to not altered Basal TCGA samples
up = GDA,MFAP2,PRKCSH,YWHAZ
study = TCGA (Cell 2015)
p−value: 0.002
HR: 4.515
95% CI: 1.615 − 12.622
Overall Survival
Months
altered in Basal (N=11, NE=5, Mdn=33.9)
not altered in Basal (N=125, NE=15, Mdn=NA)

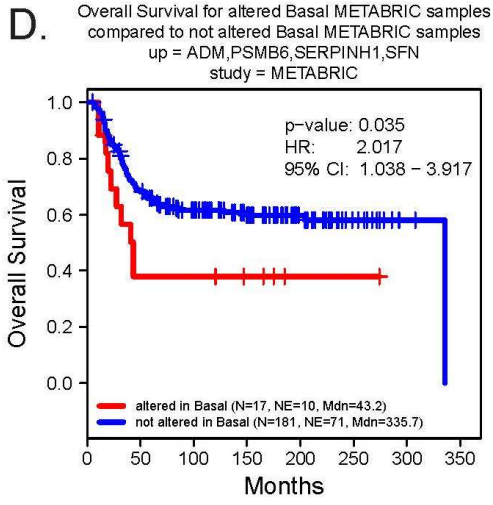
B.


Overall Survival for altered Basal METABRIC samples
compared to not altered Basal METABRIC samples
up = GDA,MFAP2,PRKCSH,YWHAZ
study = METABRIC
p−value: 0.02
HR: 2.324
95% CI: 1.118 − 4.833
Overall Survival
Months
altered in Basal (N=11, NE=8, Mdn=44.8)
not altered in Basal (N=187, NE=73, Mdn=335.7)

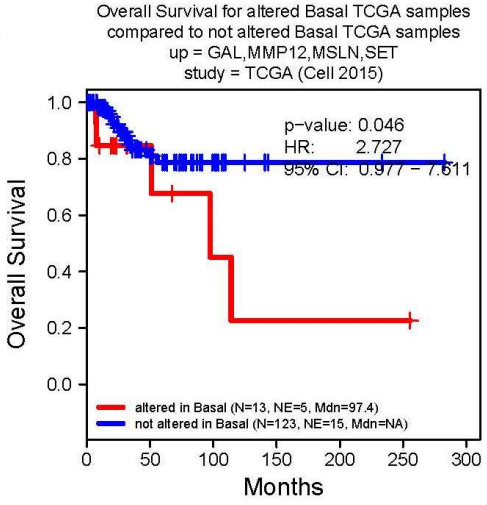
C.


Overall Survival for altered Basal TCGA samples
compared to not altered Basal TCGA samples
up = ADM,PSMB6,SERPINH1,SFN
study = TCGA (Cell 2015)
p−value: 0.01
HR: 4
95% CI: 1.275 − 12.542
Overall Survival
Months
altered in Basal (N=16, NE=4, Mdn=47.3)
not altered in Basal (N=120, NE=16, Mdn=NA)

D.


Overall Survival for altered Basal METABRIC samples
compared to not altered Basal METABRIC samples
up = ADM,PSMB6,SERPINH1,SFN
study = METABRIC
p−value: 0.035
HR: 2.017
95% CI: 1.038 − 3.917
Overall Survival
Months
altered in Basal (N=17, NE=10, Mdn=43.2)
not altered in Basal (N=181, NE=71, Mdn=335.7)

E.


Overall Survival for altered Basal TCGA samples
compared to not altered Basal TCGA samples
up = GAL,MMP12,MSLN,SET
study = TCGA (Cell 2015)
p−value: 0.046
HR: 2.727
95% CI: 0.977 − 7.611
Overall Survival
Months
altered in Basal (N=13, NE=5, Mdn=97.4)
not altered in Basal (N=123, NE=15, Mdn=NA)

F.

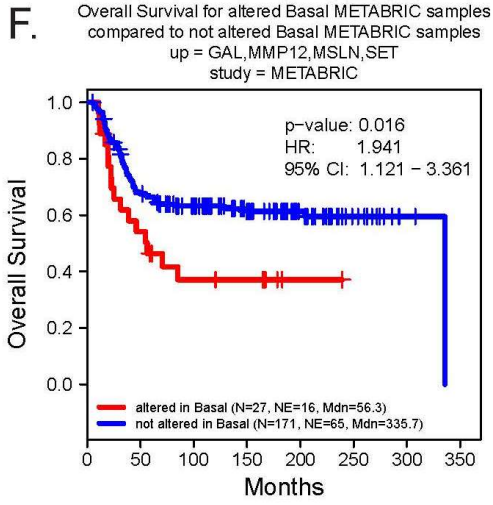
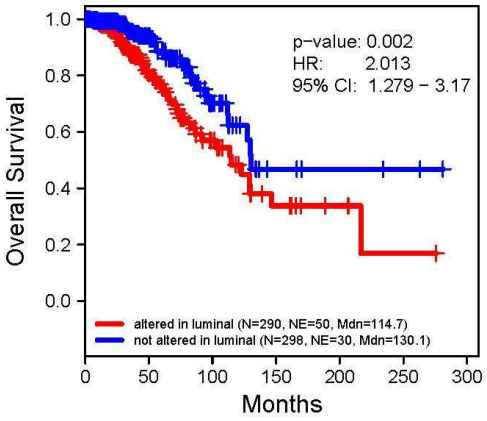
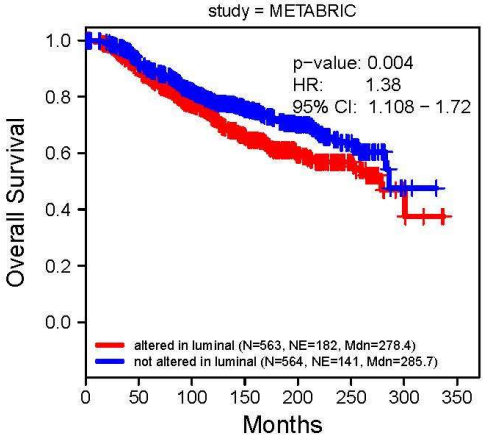

Overall Survival for altered Basal METABRIC samples
compared to not altered Basal METABRIC samples
up = GAL,MMP12,MSLN,SET
study = METABRIC
p−value: 0.016
HR: 1.941
95% CI: 1.121 − 3.361
Overall Survival
Months
altered in Basal (N=27, NE=16, Mdn=56.3)
not altered in Basal (N=171, NE=65, Mdn=335.7)

Figure 2.9 - Correlation between Kaplan-Meier survival plots of the clinical outcomes and mRNA co-overexpression of indicated basal SeCEP genes based on TCGA (left column) and METABRIC (right column) patient data. "N" refers to "Number of patients," and "NE" refers to "Number of Events (Overall Survival status = DECEASED)". Each plot shows the log-rank p-value and Hazard Ratio (HR) with 95% Confidence Interval (CI) between the two groups. The red line designates the patient subpopulation showing statistically significant overexpression of the indicated basal-specific genes (“altered”). The blue line designates the group of patients not showing statistically significant overexpression of the indicated basal-specific genes (“not altered”).

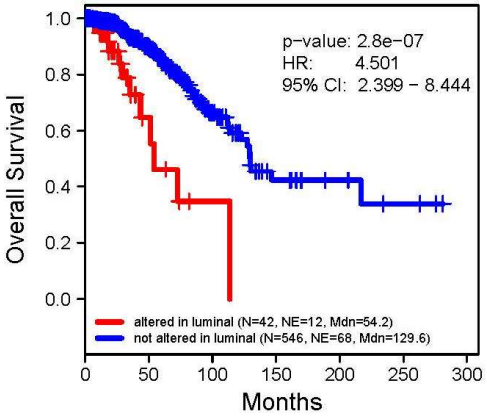
A.


Overall Survival for altered luminal TCGA samples
compared to not altered luminal TCGA samples
up = CLEC3A
study = TCGA (Cell 2015)
p−value: 0.002
HR: 2.013
95% CI: 1.279 − 3.17
Overall Survival
Months
altered in luminal (N=290, NE=50, Mdn=114.7)
not altered in luminal (N=298, NE=30, Mdn=130.1)

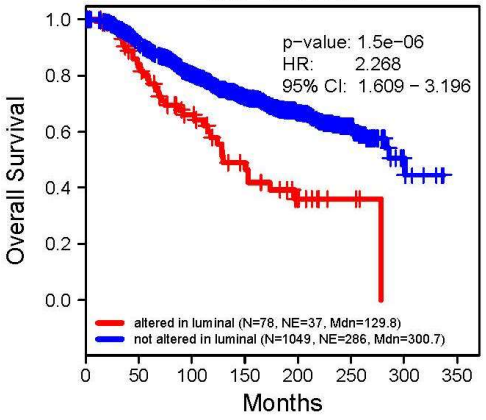
B.


Overall Survival for altered luminal TCGA samples
compared to not altered luminal TCGA samples
up = CLEC3A
study = METABRIC
p−value: 0.004
HR: 1.38
95% CI: 1.108 − 1.72
Overall Survival
Months
altered in luminal (N=563, NE=182, Mdn=278.4)
not altered in luminal (N=564, NE=141, Mdn=285.7)

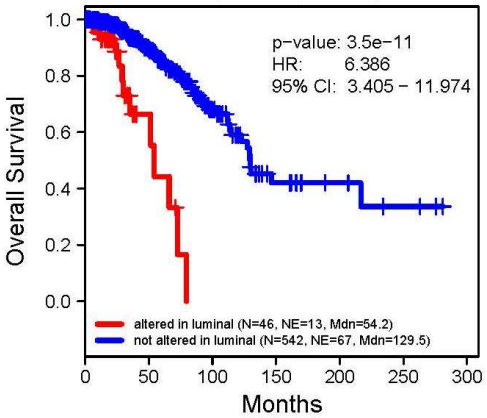
C.


Overall Survival for altered luminal TCGA samples
compared to not altered luminal TCGA samples
up = CLEC3A,CTTN,IGFBP5,NRCAM
study = TCGA (Cell 2015)
p−value: 2.8e−07
HR: 4.501
95% CI: 2.399 − 8.444
Overall Survival
Months
altered in luminal (N=42, NE=12, Mdn=54.2)
not altered in luminal (N=546, NE=68, Mdn=129.6)

D.


Overall Survival for altered luminal TCGA samples
compared to not altered luminal TCGA samples
up = CLEC3A,CTTN,IGFBP5,NRCAM
study = METABRIC
p−value: 1.5e−06
HR: 2.268
95% CI: 1.609 − 3.196
Overall Survival
Months
altered in luminal (N=78, NE=37, Mdn=129.8)
not altered in luminal (N=1049, NE=286, Mdn=300.7)

E.


Overall Survival for altered luminal TCGA samples
compared to not altered luminal TCGA samples
up = ALDOA,CLEC3A,EEA1,FKBP4
study = TCGA (Cell 2015)
p−value: 3.5e−11
HR: 6.386
95% CI: 3.405 − 11.974
Overall Survival
Months
altered in luminal (N=46, NE=13, Mdn=54.2)
not altered in luminal (N=542, NE=67, Mdn=129.5)

F.

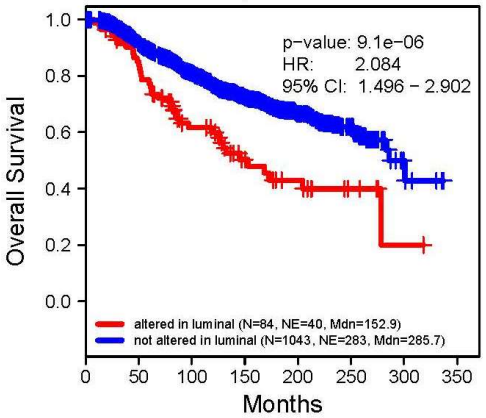

Overall Survival for altered luminal TCGA samples
compared to not altered luminal TCGA samples
up = ALDOA,CLEC3A,EEA1,FKBP4
study = METABRIC
p−value: 9.1e−06
HR: 2.084
95% CI: 1.496 − 2.902
Overall Survival
Months
altered in luminal (N=84, NE=40, Mdn=152.9)
not altered in luminal (N=1043, NE=283, Mdn=285.7)

Figure 2.10 - Correlation between Kaplan-Meier survival plots of the clinical outcomes and mRNA co-overexpression of indicated luminal SeCEP genes based on TCGA (left column) and METABRIC (right column) patient data. "N" refers to "Number of patients," and "NE" refers to "Number of Events (Overall Survival status = DECEASED)". Each plot shows the log-rank p-value and Hazard Ratio (HR) with 95% Confidence Interval (CI) between the two groups. The red line designates the patient subpopulation showing statistically significant overexpression of the indicated luminal-specific genes (“altered”). The blue line designates the group of patients not showing statistically significant overexpression of the indicated luminal-specific genes (“not altered”). Overexpression of CLEC3A provides less prognostic value than overexpression of CLEC3A in combination with other secreted factors.

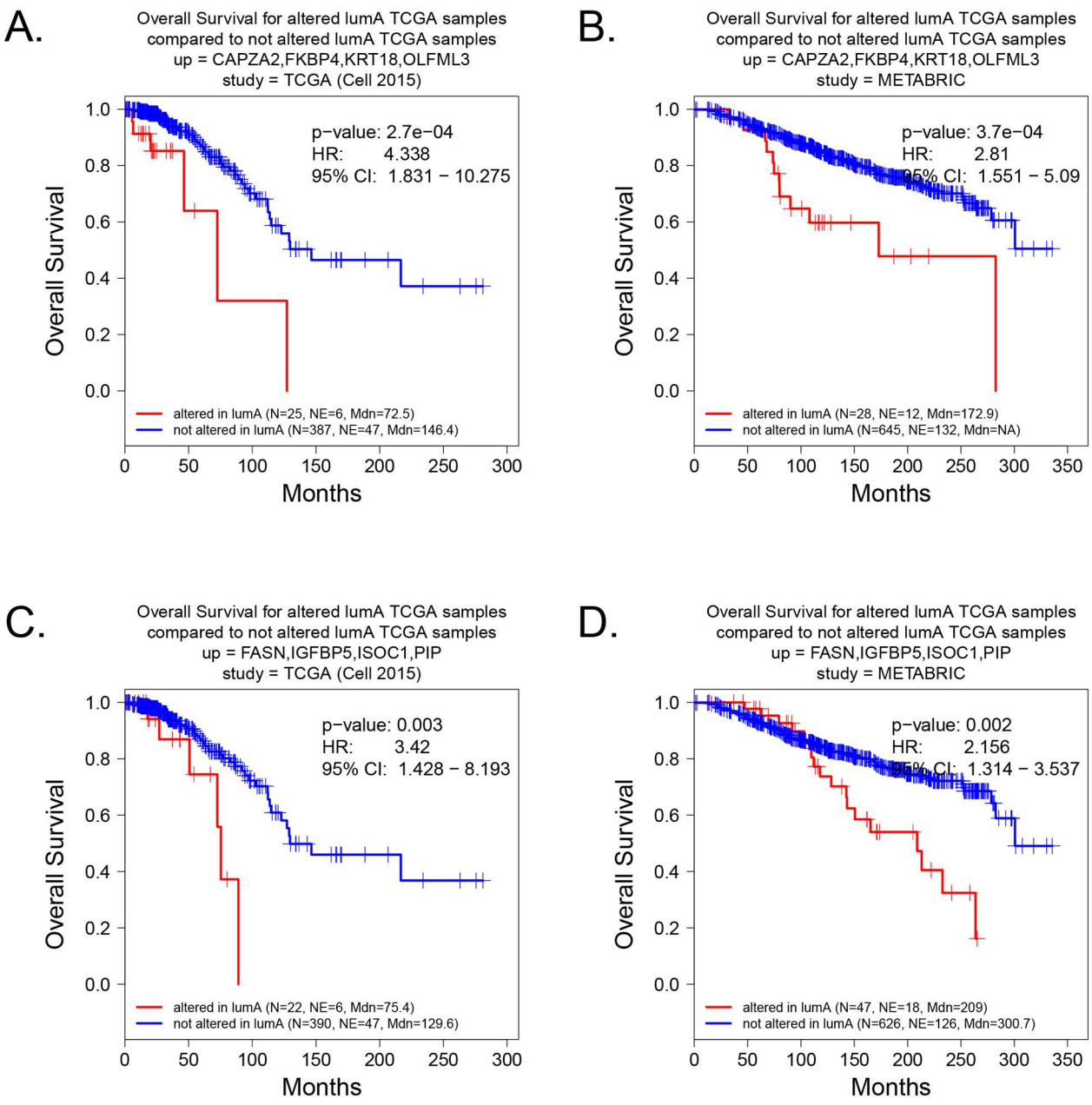


Figure 2.11 - Correlation between Kaplan-Meier survival plots of the clinical outcomes and mRNA co-overexpression of indicated luminal SeCEP genes based on TCGA (left column) and METABRIC (right column) patient data. "N" refers to "Number of patients," and "NE" refers to "Number of Events (Overall Survival status = DECEASED)". Each plot shows the log-rank p-value and Hazard Ratio (HR) with 95% Confidence Interval (CI) between the two groups. The red line designates the patient subpopulation showing statistically significant overexpression of the indicated luminal-specific genes (“altered”). The blue line designates the group of patients not showing statistically significant overexpression of the indicated luminal-specific genes (“not altered”). Co-overexpression of distinct sets of genes correlate with statistically significant changes in overall survival in Luminal A patients but not Luminal B or other BC subtypes.

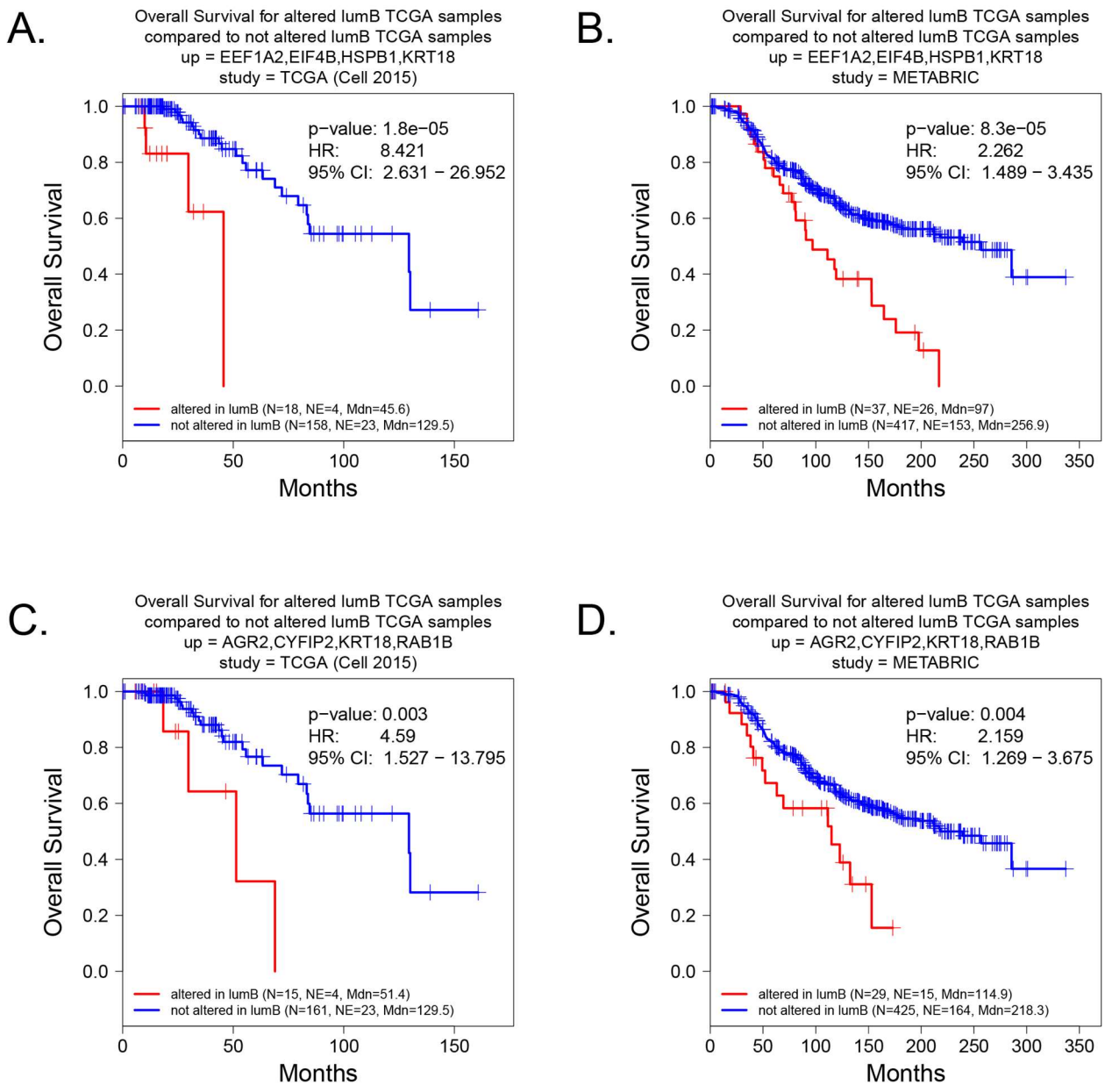
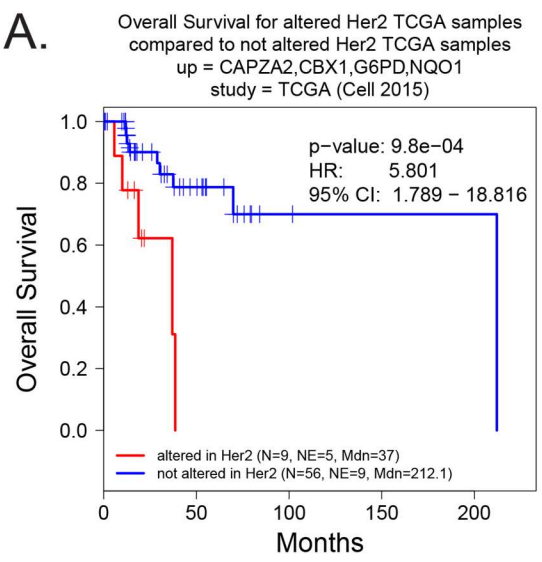
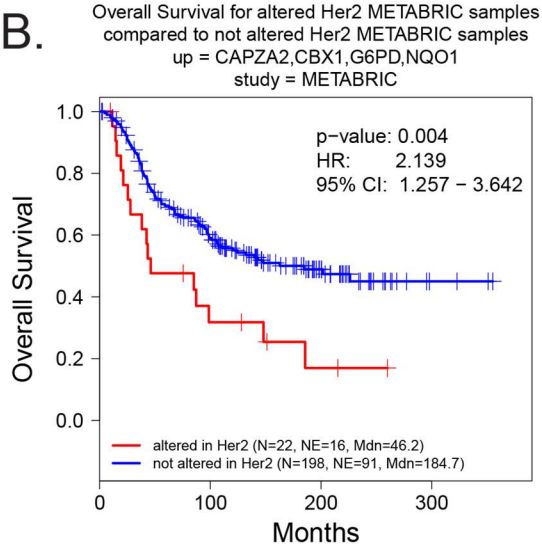


Figure 2.12 - Correlation between Kaplan-Meier survival plots of the clinical outcomes and mRNA co-overexpression of indicated luminal SeCEP genes based on TCGA (left column) and METABRIC (right column) patient data. "N" refers to "Number of patients," and "NE" refers to "Number of Events (Overall Survival status = DECEASED)". Each plot shows the log-rank p-value and Hazard Ratio (HR) with 95% Confidence Interval (CI) between the two groups. The red line designates the patient subpopulation showing statistically significant overexpression of the indicated luminal-specific genes (“altered”). The blue line designates the group of patients not showing statistically significant overexpression of the indicated luminal-specific genes (“not altered”). Co-overexpression of distinct sets of genes correlate with statistically significant changes in overall survival in Luminal B patients but not Luminal A or other BC subtypes.

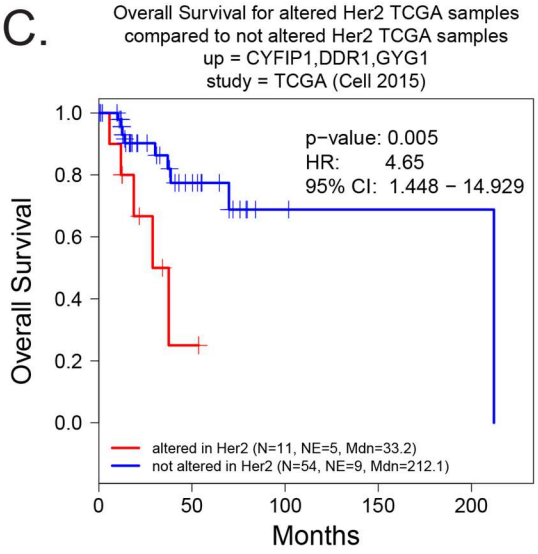
A.


Overall Survival for altered Her2 TCGA samples
compared to not altered Her2 TCGA samples
up = CAPZA2,CBX1,G6PD,NQO1
study = TCGA (Cell 2015)
p−value: 9.8e−04
HR: 5.801
95% CI: 1.789 − 18.816
altered in Her2 (N=9, NE=5, Mdn=37)
not altered in Her2 (N=56, NE=9, Mdn=212.1)
Overall Survival
Months

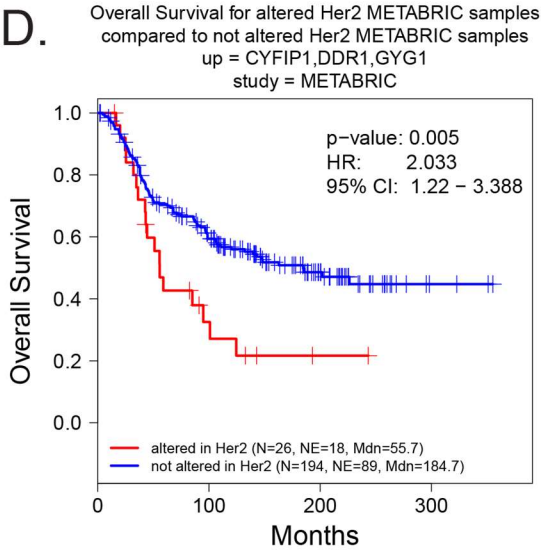
B.


Overall Survival for altered Her2 METABRIC samples
compared to not altered Her2 METABRIC samples
up = CAPZA2,CBX1,G6PD,NQO1
study = METABRIC
p−value: 0.004
HR: 2.139
95% CI: 1.257 − 3.642
altered in Her2 (N=22, NE=16, Mdn=46.2)
not altered in Her2 (N=198, NE=91, Mdn=184.7)
Overall Survival
Months

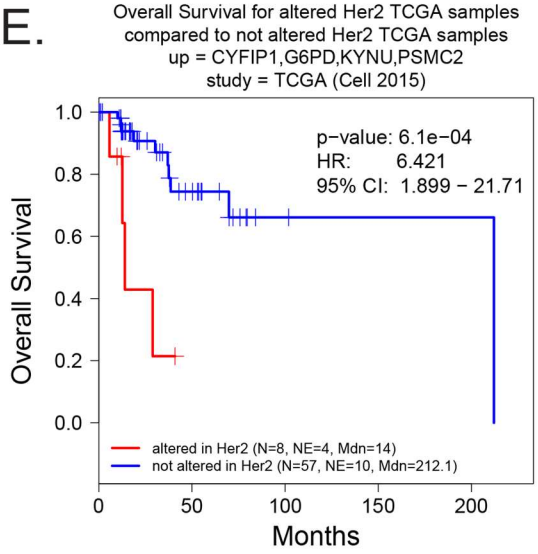
C.


Overall Survival for altered Her2 TCGA samples
compared to not altered Her2 TCGA samples
up = CYFIP1,DDR1,GYG1
study = TCGA (Cell 2015)
p−value: 0.005
HR: 4.65
95% CI: 1.448 − 14.929
altered in Her2 (N=11, NE=5, Mdn=33.2)
not altered in Her2 (N=54, NE=9, Mdn=212.1)
Overall Survival
Months

D.


Overall Survival for altered Her2 METABRIC samples
compared to not altered Her2 METABRIC samples
up = CYFIP1,DDR1,GYG1
study = METABRIC
p−value: 0.005
HR: 2.033
95% CI: 1.22 − 3.388
altered in Her2 (N=26, NE=18, Mdn=55.7)
not altered in Her2 (N=194, NE=89, Mdn=184.7)
Overall Survival
Months

E.


Overall Survival for altered Her2 TCGA samples
compared to not altered Her2 TCGA samples
up = CYFIP1,G6PD,KYNU,PSMC2
study = TCGA (Cell 2015)
p−value: 6.1e−04
HR: 6.421
95% CI: 1.899 − 21.71
altered in Her2 (N=8, NE=4, Mdn=14)
not altered in Her2 (N=57, NE=10, Mdn=212.1)
Overall Survival
Months

F.

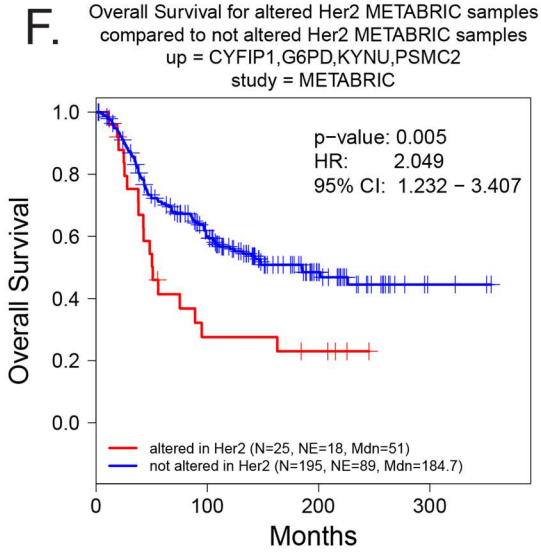

Overall Survival for altered Her2 METABRIC samples
compared to not altered Her2 METABRIC samples
up = CYFIP1,G6PD,KYNU,PSMC2
study = METABRIC
p−value: 0.005
HR: 2.049
95% CI: 1.232 − 3.407
altered in Her2 (N=25, NE=18, Mdn=51)
not altered in Her2 (N=195, NE=89, Mdn=184.7)
Overall Survival
Months

Figure 2.13 - Correlation between Kaplan-Meier survival plots of the clinical outcomes and mRNA co-overexpression of indicated luminal SeCEP genes based on HER2+ patient data. "N" refers to "Number of patients," and "NE" refers to "Number of Events (Overall Survival status = DECEASED)". Each plot shows the log-rank p-value and Hazard Ratio (HR) with 95% Confidence Interval (CI) between the two groups. The red line designates the patient subpopulation showing statistically significant overexpression of the indicated luminal-specific genes (“altered”). The blue line designates the group of patients not showing statistically significant overexpression of the indicated luminal-specific genes (“not altered”).

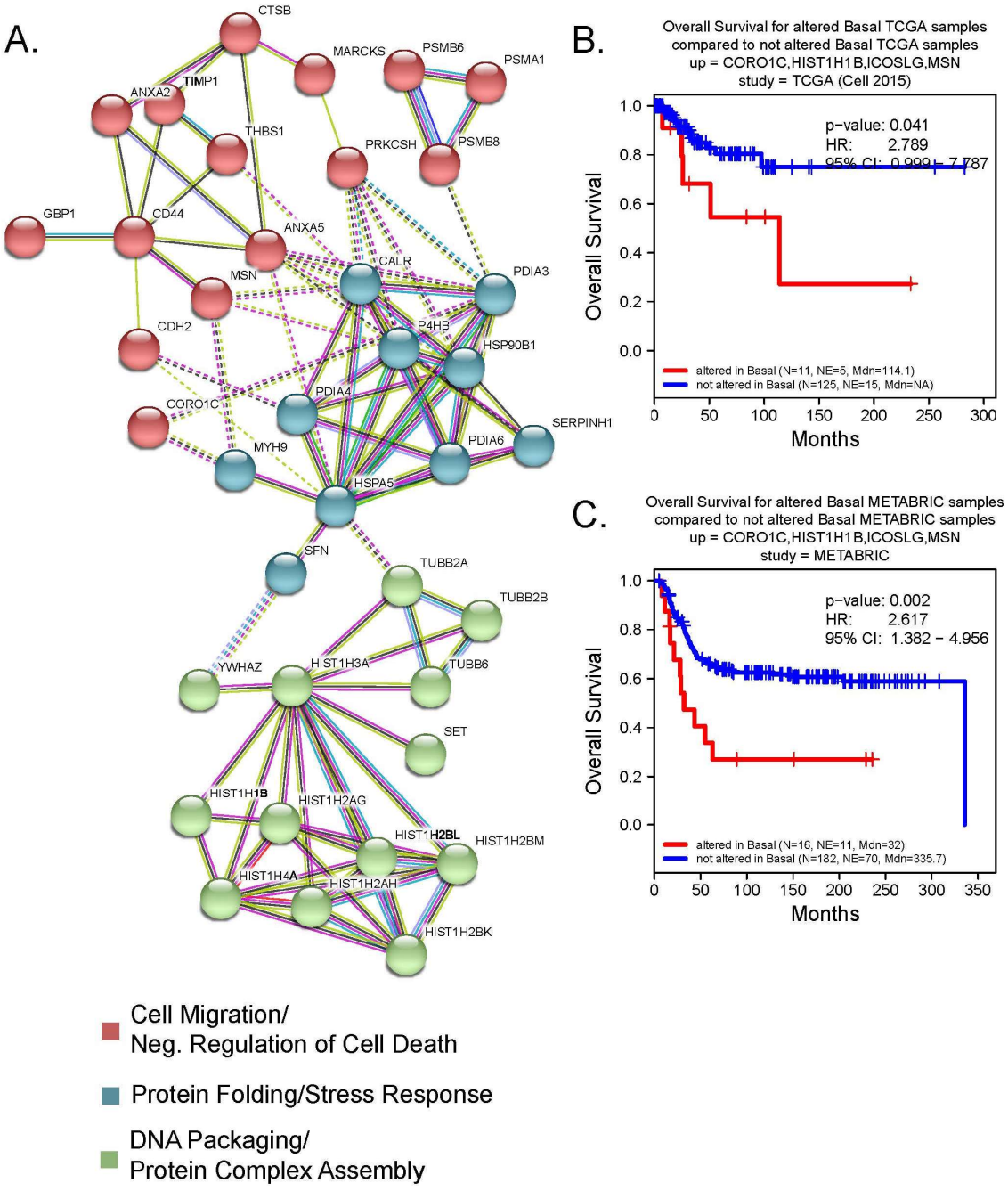


Figure 2.14 - A) The interactive subnetworks of basal SeCEP genes. B,C) Correlation between Kaplan-Meier survival plots of the clinical outcomes and mRNA co-overexpression of indicated luminal SeCEP genes based on TCGA (B) and METABRIC (C) patient data. "N" refers to "Number of patients," and "NE" refers to "Number of Events (Overall Survival status = DECEASED)". Each plot shows the log-rank p-value and Hazard Ratio (HR) with 95% Confidence Interval (CI) between the two groups. The red line designates the patient subpopulation showing statistically significant overexpression of the indicated basal-specific genes (“altered”). The blue line designates the group of patients not showing statistically significant overexpression of the indicated basal-specific genes (“not altered”). Three of the four overexpressed genes are members of basal-specific interactive subnetworks.

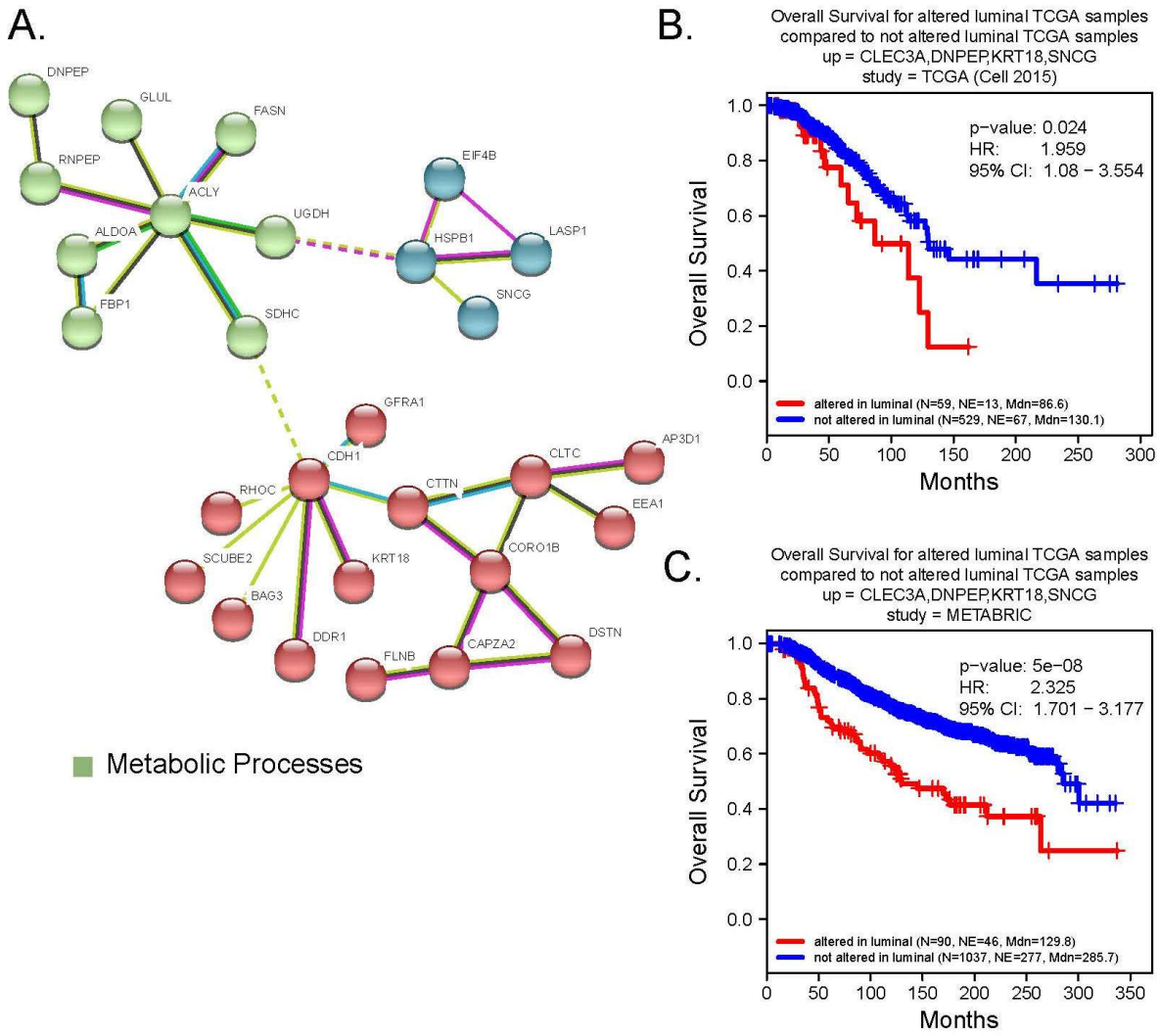


Figure 2.15 – A) The interactive subnetworks of luminal SeCEP genes. B,C) Correlation between Kaplan-Meier survival plots of the clinical outcomes and mRNA co-overexpression of indicated luminal SeCEP genes based on TCGA (B) and METABRIC (C) patient data. "N" refers to "Number of patients," and "NE" refers to "Number of Events (Overall Survival status = DECEASED)". Each plot shows the log-rank p-value and Hazard Ratio (HR) with 95% Confidence Interval (CI) between the two groups. The red line designates the patient subpopulation showing statistically significant overexpression of the indicated luminal-specific genes (“altered”). The blue line designates the group of patients not showing statistically significant overexpression of the indicated luminal-specific genes (“not altered”). Three of the four overexpressed genes are members of luminal-specific interactive subnetworks.

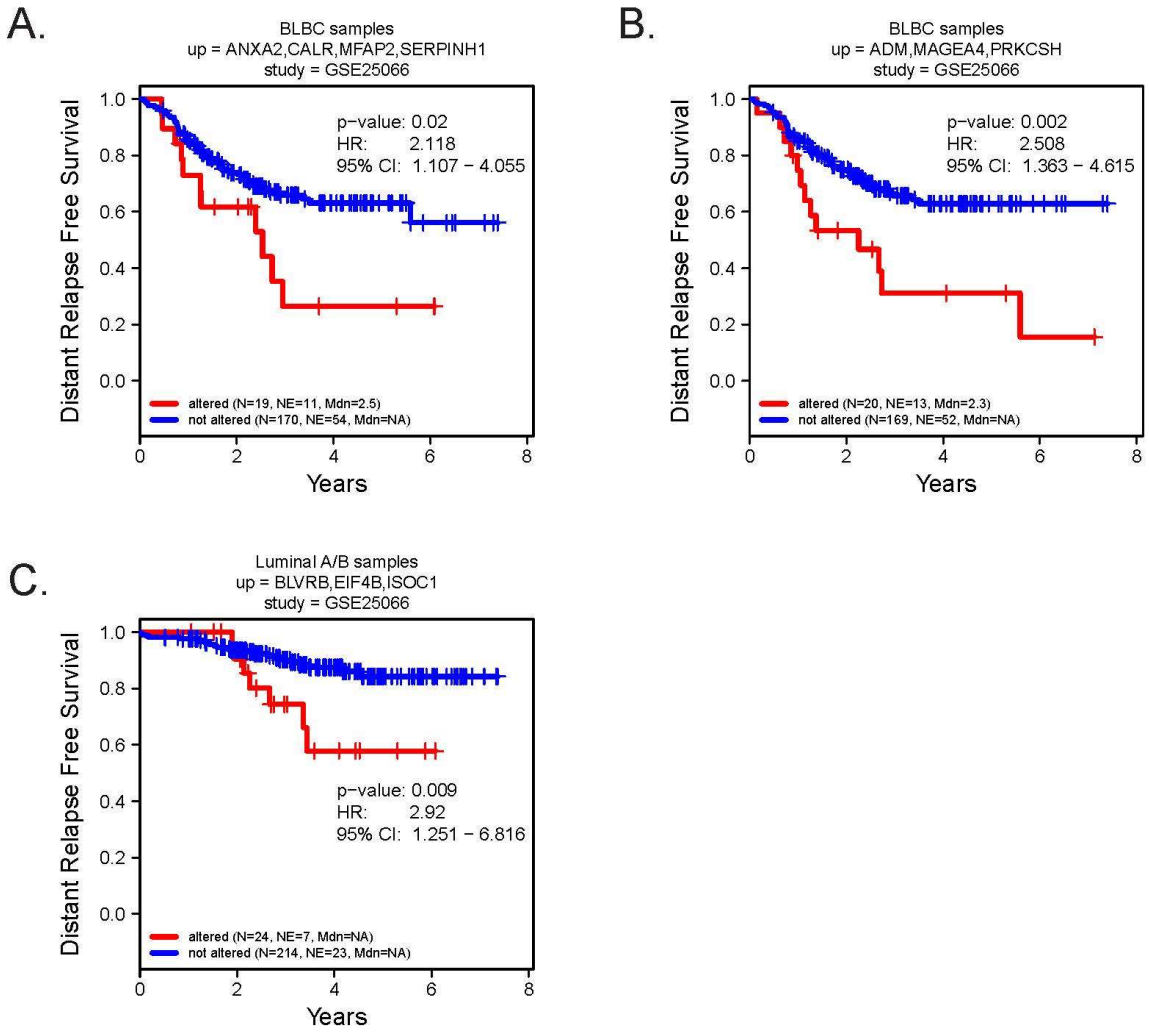


Figure 2.16 - Correlation between Kaplan-Meier survival plots of the clinical outcomes and mRNA co-overexpression of indicated SeCEP genes based on GSE25066 patient data. "N" refers to "Number of patients," and "NE" refers to "Number of Events (Overall Survival status = DECEASED)". Each plot shows the log-rank p-value and Hazard Ratio (HR) with 95% Confidence Interval (CI) between the two groups. The red line designates the patient subpopulation showing statistically significant overexpression of the indicated subtype-specific genes (“altered”). The blue line designates the group of patients not showing statistically significant overexpression of the indicated subtype-specific genes (“not altered”).

CHAPTER 3: A novel “outside-in” proteomic approach reveals signaling pathways promoting the EZH2-dependent aggressive phenotype of triple negative breast cancer

INTRODUCTION

Breast cancer (BC) is the second most common type of cancer among women in the United States, with over 200,000 new diagnoses of invasive breast cancer per year.[130] Many BC-associated genes have been identified by global genomic sequencing of metastatic and primary lesions,[147, 207, 208] and six BC PAM50-subtypes have been classified by a 50-gene transcriptome model.[64] However, these genetic and transcriptomic signatures are insufficient to distinguish, within single PAM50-subtypes, BC patient subpopulations having highly variable clinical outcomes/phenotypes. This limitation implies that evolution of malignant and metastatic breast pre-neoplasia is mostly due to post-genomic and post-transcriptional alterations. Among all BC subtypes, triple-negative breast cancer (TNBC) is the most difficult to treat and is generally associated with faster tumor growth rates and poorer prognosis.[83, 209] Additionally, TNBC is more likely to acquire drug-resistance than other types of breast cancer, further hindering treatment efforts.[210, 211] Conventional prognostic parameters, such as tumor size and grade, are often inaccurate to specify a TNBC pathological stage, and there is a lack of TNBC-specific biomarkers. For these reasons, some TNBC patients are diagnosed only after reaching advanced pathologic stages, with approximately 30% of patients undergoing distant relapse and some acquiring therapeutic resistance. Others are unnecessarily over-treated with toxic chemotherapies. Due to high TNBC mortality, there is an urgent need to develop highly sensitive and specific markers to predict prognosis and drug-responsiveness of TNBC patients and to identify new-generation targets for precision therapy. Protein biomarkers represent a

potentially paradigm-shifting advance in the detection and clinical evaluation of breast cancer. Tumor cells are a potentially rich source of biomarkers because they secrete and shed proteins at a higher rate than normal cells, and many of these proteins enter circulation.[83, 84]

One epigenetic driver that has been linked to both the aggressive and drug-resistant phenotypes of TNBC is the chromatin modifier EZH2.[212-214] As the only catalytic subunit of Polycomb Repressive Complex 2 (PRC2), EZH2 acts as a master transcriptional regulator through the catalysis of tri-methylation of lysine 27 on histone H3 (H3K27me3) within the chromatin associated with transcriptionally silenced genes.[215] In particular, genes related to tumor suppression, cell apoptosis, and DNA repair are silenced.[216] Because of the clear relationship between EZH2 and disease characteristics, several phase I and II trials targeting EZH2 and the PRC2 complex are underway.[217] Although aberrant overexpression of EZH2 in breast tissue correlates with more aggressive breast cancer, higher rates of metastasis, and ultimately poorer clinical outcome, [218, 219] there is lack of mechanistic details with respect to exactly how aberrantly expressed EZH2 drives tumorigenesis or therapeutic resistance. Identification of the biological processes and pathways associated with EZH2 dysregulation will reveal the underlying mechanisms and mechanistically derive new phenotypic, non-invasive biomarkers for precision prognosis and therapeutic response for precision medicine.

In correlation with distinctive pathological stage/grade, tumor cells secrete specific extracellular proteins (‘outer proteome’) that originate from the aberrantly regulated intracellular proteome and associated pathways (‘inside proteome’). We hypothesized that TNBC cells secrete characteristic proteins that promote the aggressive and metastatic nature of TNBC, and these characteristic proteins are derived from EZH2-dependent aberrant intracellular signaling pathways. To test this hypothesis we treated MDA-231 breast cancer cells with the EZH2 inhibitor UNC1999 and used our label-free quantitative proteomic (LFQ) method to compare the protein secretomes between treated and untreated cells. We further dissected those EZH2-dependent biological processes and pathways that contribute to breast cancer tumorigenesis

and aggressiveness using a novel “outside-in” LFQ proteomic approach. Our novel workflow first identifies potential biomarkers of EZH2-active tumors in the extracellular phenotype and identifies factors which may contribute to metastasis and proliferation in TNBC. It then explores the intracellular mechanisms involved in the production and secretion of those factors. We discovered that signaling intracellular signaling pathway activation was strongly reflected in the secretome and found that tools designed for intracellular analysis can also be effective for secretome screening.

METHODS

Chemicals and reagents. Cell culture media and fetal bovine serum were obtained from Gibco. All other components of cell culture media and protease inhibitor cocktails were purchased from Sigma (St, Louis, MO). Trypsin was purchased from Promega. All chemicals were HPLC-grade unless specifically indicated. MDA-231 cells were purchased from ATCC (Manassas,VA).

Cell culture and secreted protein collection. MDA-231 cells were cultured in DMEM containing 10% fetal bovine serum. For intracellular analysis, cells were treated with 0.5 μM UNC1999 for 72 hours and lysed with 5x cell pellet volume of RIPA buffer (50 mM Tris pH 8.0, 150mM NaCl, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 1 μM EDTA, 1 μM phenylmethylsulfonyl fluoride). The lysates were sonicated 5 x 1 s at 30%. After 10 minutes of centrifugation at maximum speed, the supernatants were transferred to a fresh tube and the pellets were discarded. Cold acetone (4x volume) was added to the lysates, mixed, and stored overnight at -80°. After centrifugation at maximum speed for 10 minutes, the pellets were resuspended in 50 μl buffer (8 M Urea, 50 mM Tris-HCl pH 8.0, 150 mM NaCl), reduced with dithiothreitol (5 mM final) for 30 minutes at room temperature, and alkylated with iodoacetamide (15 mM final) for 45 minutes in the dark at room temperature. Alkylation was quenched with dithiothreitol (10 mM final). Samples were diluted 4-fold with 25mM Tris-HCl pH

8.0, 1mM $CaCl_2$ and digested with 1:200 (wt/wt) trypsin overnight at room temperature. Peptides were desalted on a StageTip containing a 4 × 1 mm C18 extraction disk (3M) and dried.[139]

For secretome analysis, cells were treated with 0.5 or 1 μM UNC1999 for 48 hours, then growth media was removed and cells were washed twice with PBS. Serum-free media without phenol red containing 0.5 or 1 μM was then added to the plate. After 24 hours the conditioned media were collected and centrifuged at 500 x g for 5 minutes to remove cellular debris, then the supernatant was syringe-filtered with 0.2 μm 13 mm diameter polytetrafluoroethylene filters (VWR International) and transferred to fresh tubes. Samples were stored at -80 °C until further processed. After thawing, proteins were concentrated by trichloroacetic acid/sodium deoxycholate precipitation. Briefly, 1/10 of the sample volume of 0.15% sodium deoxycholate was added to each sample, then tubes were incubated on ice for 15 minutes. Next, 1/10 of the original sample volume of cold 72% trichloroacetic acid was added and the tubes were incubated on ice for 15 minutes. Samples were centrifuged for 10 minutes at max speed, 4° C. The pellets were washed in cold acetone and air dried until no residual odor was detected. Next, the pellets were resuspended in 50 μl buffer (8 M Urea, 50 mM Tris-HCl pH 8.0, 150 mM NaCl), and reduction, alkylation, and trypsinization were performed as with intracellular samples.

LC-MS/MS analysis. LC-MS/MS analysis was performed as previously described.[203] Briefly, desalted peptides were dissolved to a concentration of 1 μg/μl 0.1% formic acid (Thermo-Fisher) for intracellular analysis or in 20 μl 0.1% formic acid for secretome analysis. An injection of 2 μl was analyzed by an Easy nanoLC 1000 with a 15 cm C18 reverse phase column (15 cm × 75 μm ID, C18, 2 μm, Acclaim Pepmap RSLC, Thermo-Fisher) coupled to a Q-Exactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Peptides were eluted at a constant flow rate of 300 nl/min with a gradient of 5-30% buffer B (acetonitrile and 0.1% formic acid) for 75 min, 30%-45% buffer B for 15 min, 45%-100% buffer B for 1 min, and 100% B for 9 min for intracellular analysis or 5-30% buffer B (acetonitrile and 0.1% formic acid) for 17 min, 30%-40% buffer B for 3 min, 40%-100% buffer B for 1 min, and 100% B for 9 min for secretome

analysis. The Q-Exactive was operated in the positive-ion mode but using a data-dependent top 20 method. Survey scans were acquired at a resolution of 70,000 at m/z 200. Up to the top 20 most abundant isotope patterns with charge ≥ 2 from the survey scan were selected with an isolation window of 2.0 m/z and fragmented by HCD with normalized collision energies of 27. The maximum ion injection time for the survey scan and the MS/MS scans was 250 ms and 120 ms, respectively and the ion target values were set to 1e6 and 2e5, respectively. Selected sequenced ions were dynamically excluded for 20 seconds.

Mass spec data and LFQ analysis. Mass spectral processing and peptide identification were performed on the Andromeda search engine in MaxQuant software (Version 1.5.3.17) against a human UniProt database. Cysteine carbamidomethylation was set as a defined modification, and methionine oxidation and protein amino-terminal acetylation were set as dynamic modifications. Peptide inference was made with a false discovery rate (FDR) of 1% and peptides were assigned to proteins with FDR of 5%. At least 7 amino acids were required with no more than two missed cleavages. The precursor ion mass tolerance was 8 ppm and the fragment ion mass tolerance was 0.5 Da. Experiments were conducted in multiple replicates (three biological replicates each with two technical replicates) using a match between runs option enabled and time window at 0.7 minutes. Data processing and statistical analysis were performed on Perseus (Version 1.5.1.6).[140]

Analysis of functional category and networks of subtype-specific secreted proteins. The biological processes and molecular functions of secretome proteins were categorized by Ingenuity Pathway Analysis (IPA) [141] and STRING [142] similar to previously described.[143]

RESULTS

EZH2 inhibition suppresses the extracellular aggressive phenotype of TNBC

To probe the EZH2-mediated secretome, we used a label-free quantitative (LFQ) proteomic approach to analyze MDA-231 breast cancer cells treated with 0.5 or 1 μM of the

EZH2 inhibitor UNC1999 versus treatment with DMSO. The effective inhibition of EZH2 was confirmed by the dose-dependent reduction of the H3K27me3 marker as determined by western blot (Figure 3.1a). Three biological replicates were analyzed by MS with three technical replicates each. We identified a total of 1,877 proteins, with a range of 1,334 proteins identified in the untreated control group to 1,793 proteins in the cells treated with 1 μM UNC1999.

Unsupervised hierarchical clustering revealed two clusters which showed clear trends for the dose dependent response to EZH2 inhibition. (Figure 3.1b) Cluster 1 consisted of 147 proteins which showed decreased abundance after treatment with UNC1999. (Figure 3.1c) Interestingly, analysis by STRING found that proteins involved in glycosylation were overrepresented in this cluster, and aberrant glycosylation of cell surface proteins promotes invasion and metastasis.[220] (Figure 3.2) This cluster also contained a group of lysosomal proteins, and the lysosome typically increases in volume and protein content during oncogenic transformation.[221] Increased secretion of lysosomal proteins leading to ECM degradation has also been reported in several types of cancer including BC.[221, 222] Proteins involved in adhesion, angiogenesis, and regulation of growth provide also characterized this cluster.

Cluster 2 contained 183 proteins which increased in a dose dependent fashion upon treatment with UNC1999. (Figure 3.1d) Interestingly, there were a significant number of proteins involved in functions such as translation and RNA splicing. (Figure 3.3) Tumors produce eukaryotic initiation factor (EIF) and eukaryotic elongation factor (EEF) proteins at a higher rate than normal cells, and the activity of some of these proteins has been linked to cancer development and progression.[223] This cluster was also enriched in proteins which regulate actin cytoskeleton reorganization and proteasomal proteins. The epithelial to mesenchymal transition causes dysregulation of actin cytoskeleton organization, leading to increased cell motility and invasion.[168] Further, TNBC exhibits proteasome “addiction,” and proteasomal inhibitors kill cancer cells at a higher rate than normal cells.[224]

To identify activated or suppressed signaling pathways which may promote TNBC-characteristic extracellular functions we examined canonical signaling pathways using Ingenuity Pathway Analysis (IPA). (Figure 3.4a) EZH2 inhibition with UNC1999 resulting in increases in HIPPO and RhoGDI signaling, which are tumor suppressive, [225, 226] and inhibition of matrix metalloproteinases (MMPs). MMPs are extracellular matrix (ECM) remodelers that can cleave almost all components of the ECM, thereby promoting tumor cell invasion and metastasis.[227] EZH2 inhibition led to suppression of adherens junction remodeling, ephrin receptor signaling, and actin cytoskeleton signaling. All of these pathways promote cancer progression, invasion, and/or metastasis in breast cancer.[168, 228, 229] Overall, changes to signaling pathways due to EZH2 inhibition highlight the role of EZH2 in promoting an aggressive tumor phenotype. Likewise, IPA analysis of biological functions indicated increased cell attachment and apoptosis in UNC1999 treated cells, whereas cell proliferation and cell movement were among the decreased functions. (Figure 3.4b)

Finally, we compared proteins which were identified as TNBC-characteristic (see chapter 2) to those proteins which were significantly altered by EZH2 inhibition. Of the 141 basal-specific proteins we identified, nine showed statistically significant differential secretion after treatment with UNC1999. All members of this group are known to be secreted, but only PCBP1 has previously been reported to be EZH2-mediated.[230] No pathways or biological processes are overrepresented within this group, however these proteins represent a preliminary set of TNBC-characteristic, EZH2-mediated secreted proteins which may be indictors of aggressive metastatic disease.

Intracellular pathways and processes mirror the secretome

To investigate the intracellular mechanisms associated with the extracellular role of EZH2, we compared proteomic alterations in MDA-231 cells treated with UNC1999 versus control. The effective inhibition of EZH2 was verified by western blot using a probe against

H3K27Me3. (Figure 3.5a) One biological replicate was analyzed for each condition, with two technical replicates performed per sample. We identified 6,030 proteins and a heatmap of protein quantitation is shown in Figure 3.5b. Proteins identified in only one technical replicate were removed from the dataset, leaving 4893 proteins for analysis. Treatment with UNC1999 substantially altered the intracellular proteomic landscape, with over 10% of proteins (539 of 4893) having statistically significant changes in abundance.

To examine the role of these significantly altered proteins, we used STRING to analyze their Gene Ontology Biological Process enrichment. (Figure 3.6) Many of the most enriched processes involved the mitochondria, including mitochondrial gene expression, electron transport chain assembly, oxidative phosphorylation, and mitochondrion organization. EZH2 inhibition induces apoptosis through the mitochondria dependent cell death pathway in glioma and head and neck squamous cell carcinoma, [231, 232] but there have not been similar reports in breast cancer. We also found a high number of transport proteins in the UNC1999 treated cells which is consistent with the higher number of proteins we identified in the secretome of treated cells.

Among significantly downregulated proteins we found decreases in proteins associated with cell cycle progression and nucleotide biosynthesis, consistent with UNC1999 treatment evoking a less aggressive cancer phenotype. (Figure 3.7) We also found a decrease in proteins involved in translation. With the increased abundance of these proteins in the secretome, this may indicate a need for the cell to expel extraneous proteins that were essential to the malignancy but expendable upon EZH2 inhibition. Interestingly, we observed a decrease in the chromatin organization process. This is likely due to inhibition of EZH1, which represses transcription through chromatin compaction and is also a target of UNC1999, rather than EZH2 which represses transcription by depositing the H3K27Me3 mark.[233]

To determine the intracellular signaling pathways which may exert an extracellular response promoting an aggressive phenotype, we analyzed signaling pathway activation with

IPA. (Figure 3.8a) We observed suppression of glycolysis and activation of the TCA cycle and oxidative phosphorylation consistent with a shift toward a normal balance of energy production and away from the phenotype known as the Warburg effect. The Warburg metabolic phenotype is characterized by increased glucose metabolism uncoupled from the TCA cycle and oxidative phosphorylation, resulting in most glycolysis-derived pyruvate undergoing lactic acid fermentation.[234] Increases to HIPPO signaling and RhoGDI signaling were also observed, reflecting our findings in the secretome pathway analysis.[169, 226] Further, we discovered decreased actin cytoskeleton skeleton signaling and remodeling of adherens junctions which were similarly downregulated in the secretome analysis. Indeed, we found strikingly similar pathway activation and suppression when comparing the secretome and intracellular analyses. (Figure 3.8b) Recognizing that many of these pathways were previously identified as altered in the basal-specific secretome (Chapter 2), we found the majority of pathways showed the same activation/suppression status when comparing UNC1999 treated versus untreated MDA-231 cells to MCF10A non-malignant cells versus MDA-231. This was also true, with somewhat less consistency, in the IPA analysis of biological processes. (Figure 3.9a) EZH2 inhibition cause increased cell death and DNA damage response, while cell invasion, survival, and transformation were decreased. (Figure 3.9b)

DISCUSSION

Our novel “outside-in” method is innovative in its approach to performing secretome analysis. We first perform an unbiased, discovery-driven screen of the secretome to identify those biological processes and signaling pathways which drive the extracellular phenotype. We then analyze the intracellular proteome to discover the mechanisms within the cell that are exerting an extracellular influence. Using EZH2 inhibitor UNC1999 as a probe for EZH2 activity, this strategy allowed us to: 1) identify specific secreted proteins that are characteristic of EZH2-driven TNBC and contribute to the aggressive phenotype; 2) determine those EZH2-related

intracellular pathways which lead to the TNBC-characteristic secretome; and 3) ascertain how the EZH2 inhibitor intervenes in these pathways.

Our secretome analysis found that EZH2 has a dramatic influence on the abundance and type of proteins secreted. Nearly 20% of the proteins identified showed a statistically significant change in abundance when EZH2 was inhibited. This caused a shift away from an aggressive-promotive phenotype, where proteins promoting sell survival, ECM degradation, and tumor invasion were prevalent, toward an aggressive-suppressive phenotype, where cell attachment to substrate and apoptosis prevailed. This shift was reflected in the activation and suppression of specific BC-associated signaling pathways. The proteins identified as having EZH2-mediated secretion have the potential to be liquid biopsy biomarkers of aggressive BC, and work is ongoing in our lab to explore this possibility.

The intracellular analysis bore strikingly similar results to the secretome. A lesser percentage (~10%) of intracellular proteins showed abundance alterations with EZH2 inhibition versus extracellular, indicating the potent influence EZH2 has in the extracellular space. Despite this discrepancy, the intracellular results were highly reflective of the shift away from an aggressive phenotype observed in the secretome analysis. This included the same signaling pathways and biological processes typically being altered in the same way. However, intracellular analysis identified a shift in metabolism after UNC1999 treatment which suggests a role for EZH2 in aberrant energy metabolism, which has been reported in glioblastoma [235] and pancreatic cancer [236] but not breast cancer. Additional work is being performed to determine whether the intracellular EZH2-dependent proteins may be potential prognostic markers.

Importantly, our outside-in approach found significant overlap between signaling pathway activation status and biological process enrichment in the intracellular and extracellular space. This is significant because many bioinformatics tools, including IPA, are based on evidence from intracellular studies. Although IPA is commonly used for secretome analysis,

[237-240] to our knowledge this is the first study suggesting that this use of IPA is valid. Our results provide support to previous studies in which an intracellular/extracellular correlation was assumed but not tested.

CONCLUSIONS

In sum, we developed a novel approach to the identification of putative biomarkers and the cellular pathways promoting their production and secretion. We identified a set of EZH2-mediated secreted proteins which characterize the aggressive nature of EZH2-active tumors. We then looked inside the cell and found highly similar signaling pathway activation driving the aggressive phenotype.

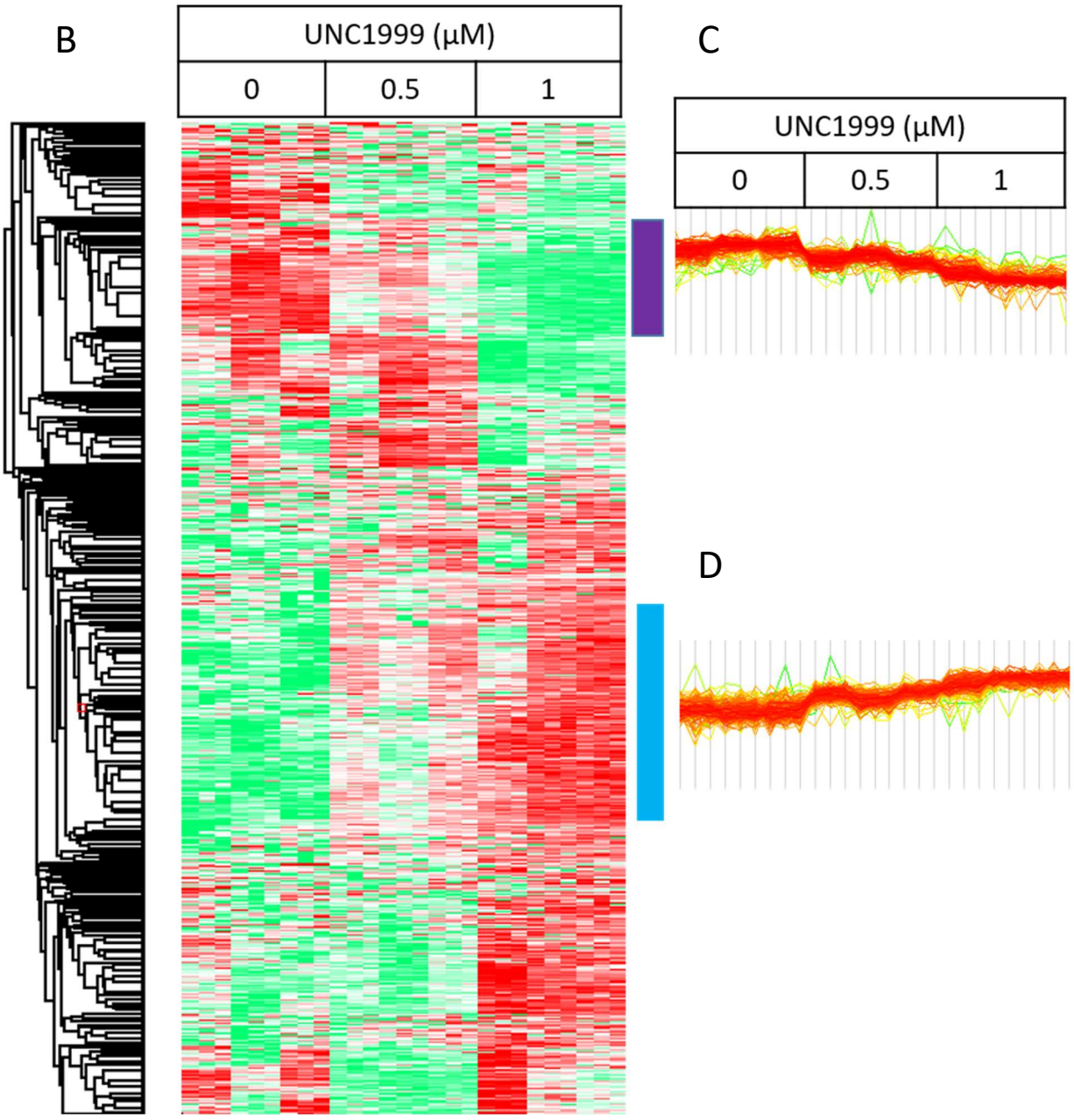


Figure 3.1 – A) Western blot showing a dose-dependent decrease in H3K27Me3 indicating effective inhibition of EZH2 B) A heatmap of unsupervised hierarchical clustering analysis of z-score normalized protein secretion MDA231 cells treated for 72 hours with 0.5 or 1 μM UNC1999 or DMSO. Each condition is represented by 3 biological replicates and 3 technical replicates. Red indicates higher secretion, green indicates lower secretion, and white indicates mean secretion. Profile plots of UNC1999 dose-dependent C) lower secretion or D) higher secretion. Each line represents one protein and the color indicates the density of proteins with similar expression levels.

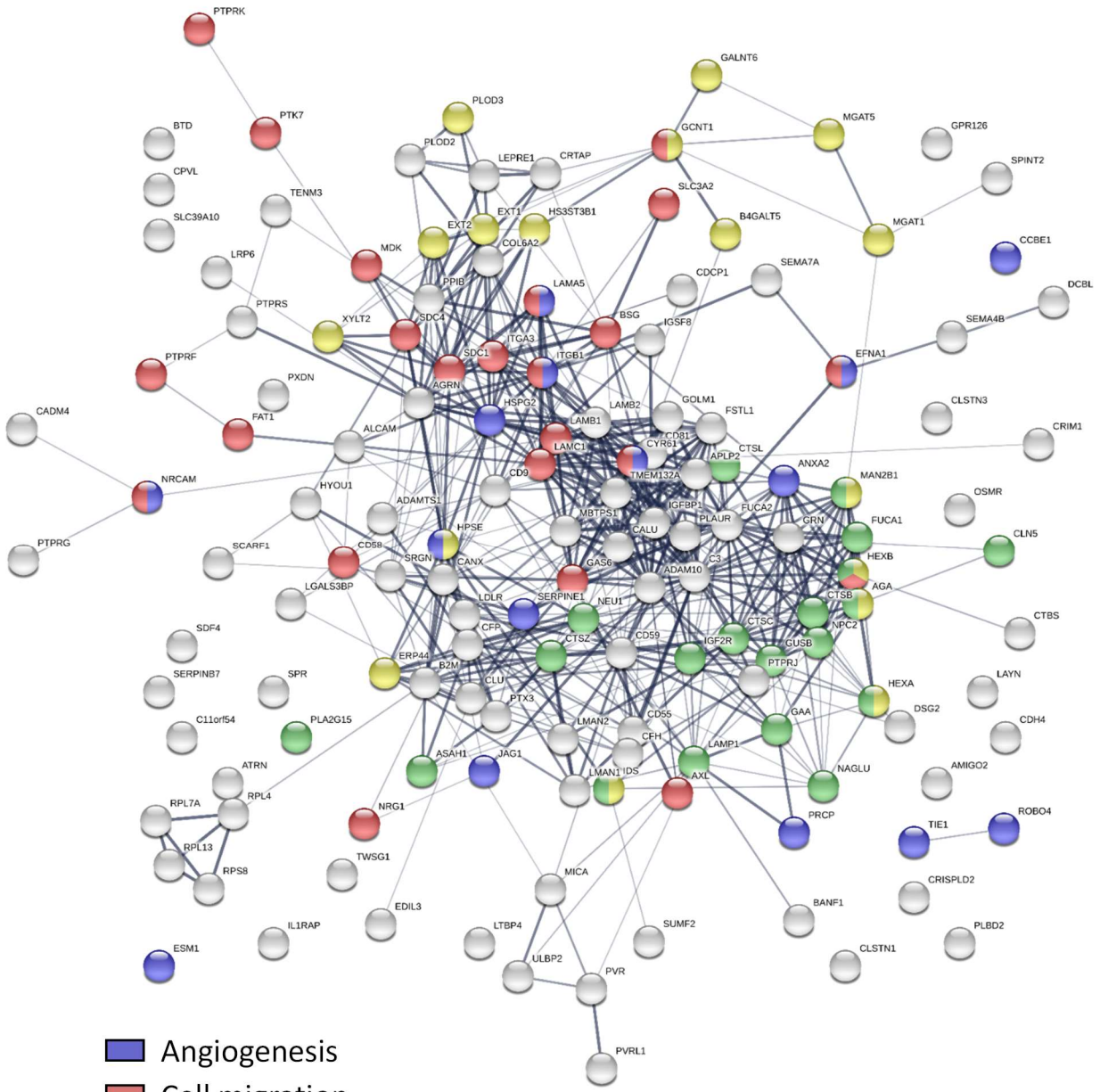


Figure 3.2 – STRING analysis of proteins with dose-dependent decreased secretion after treatment with UNC1999. Select biological processes are highlighted.

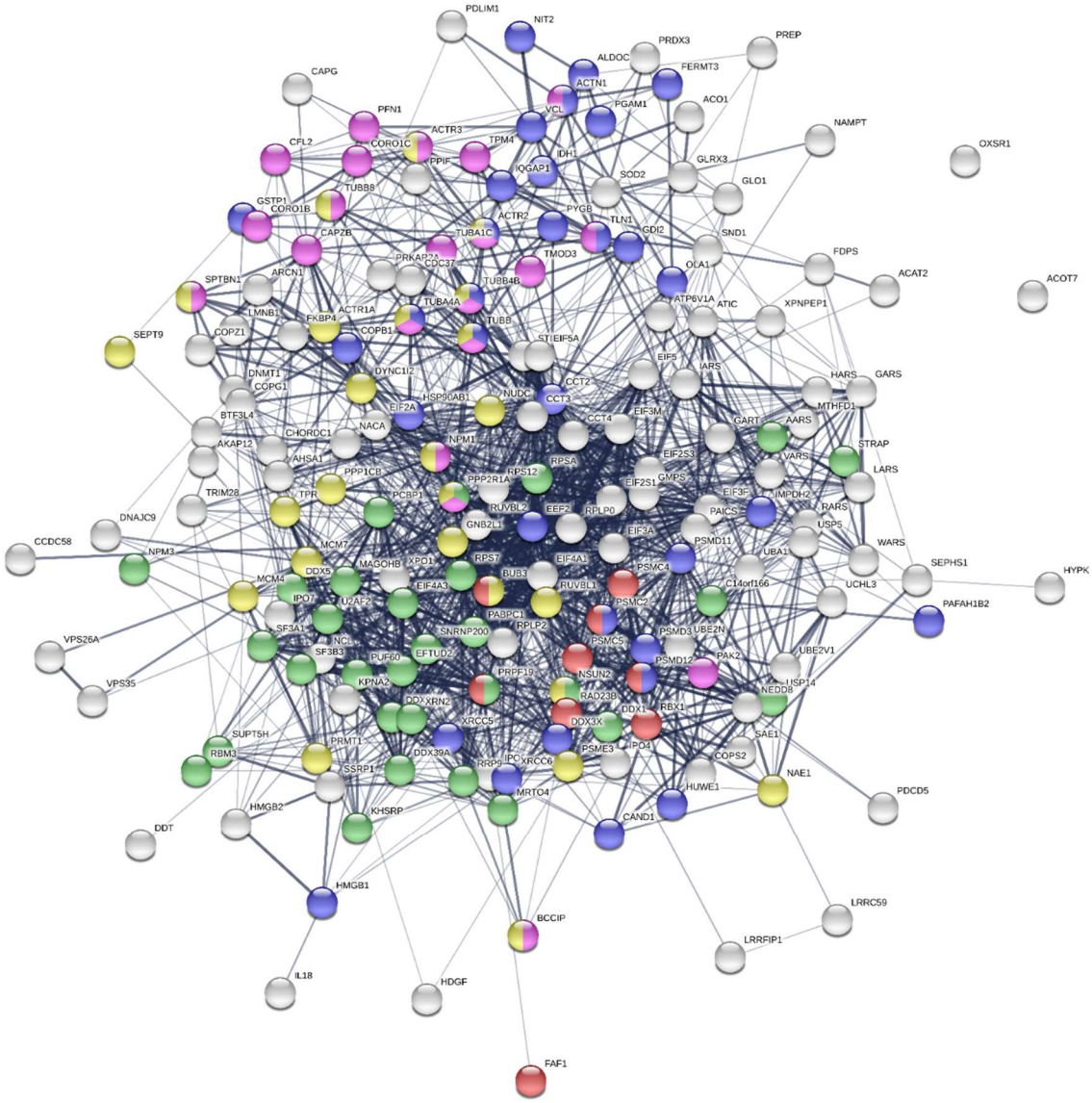


Figure 3.3 – STRING analysis of proteins with dose-dependent increased secretion after treatment with UNC1999. Select biological processes are highlighted.

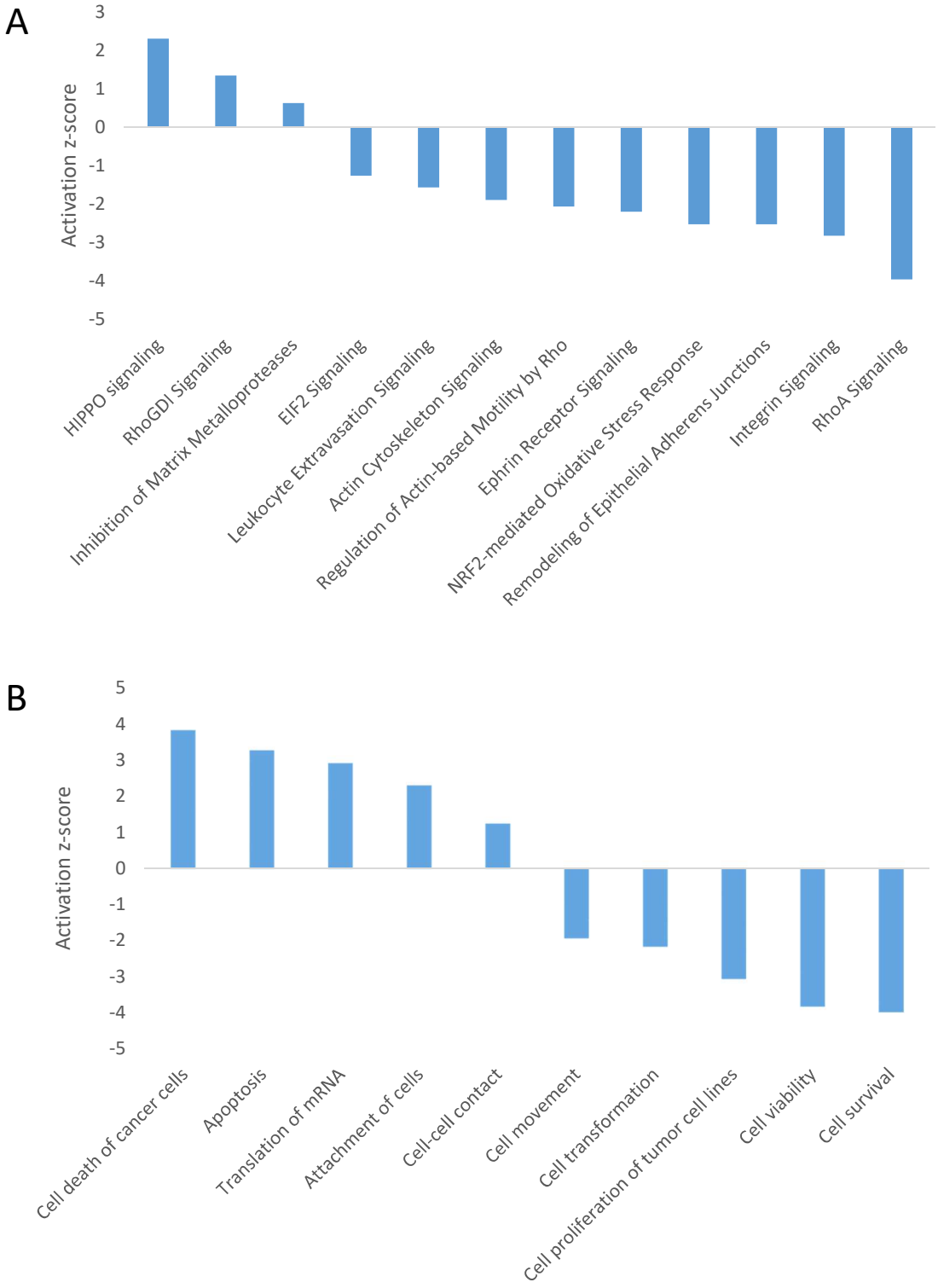


Figure 3.4 - A) Pathway activation analysis of EZH2-dependent signaling. Positive z-scores indicate pathway activation; negative z-scores indicate pathway suppression. B) Biological functions activated (positive z-score) or suppressed (negative z-score) after treatment with UNC1999.

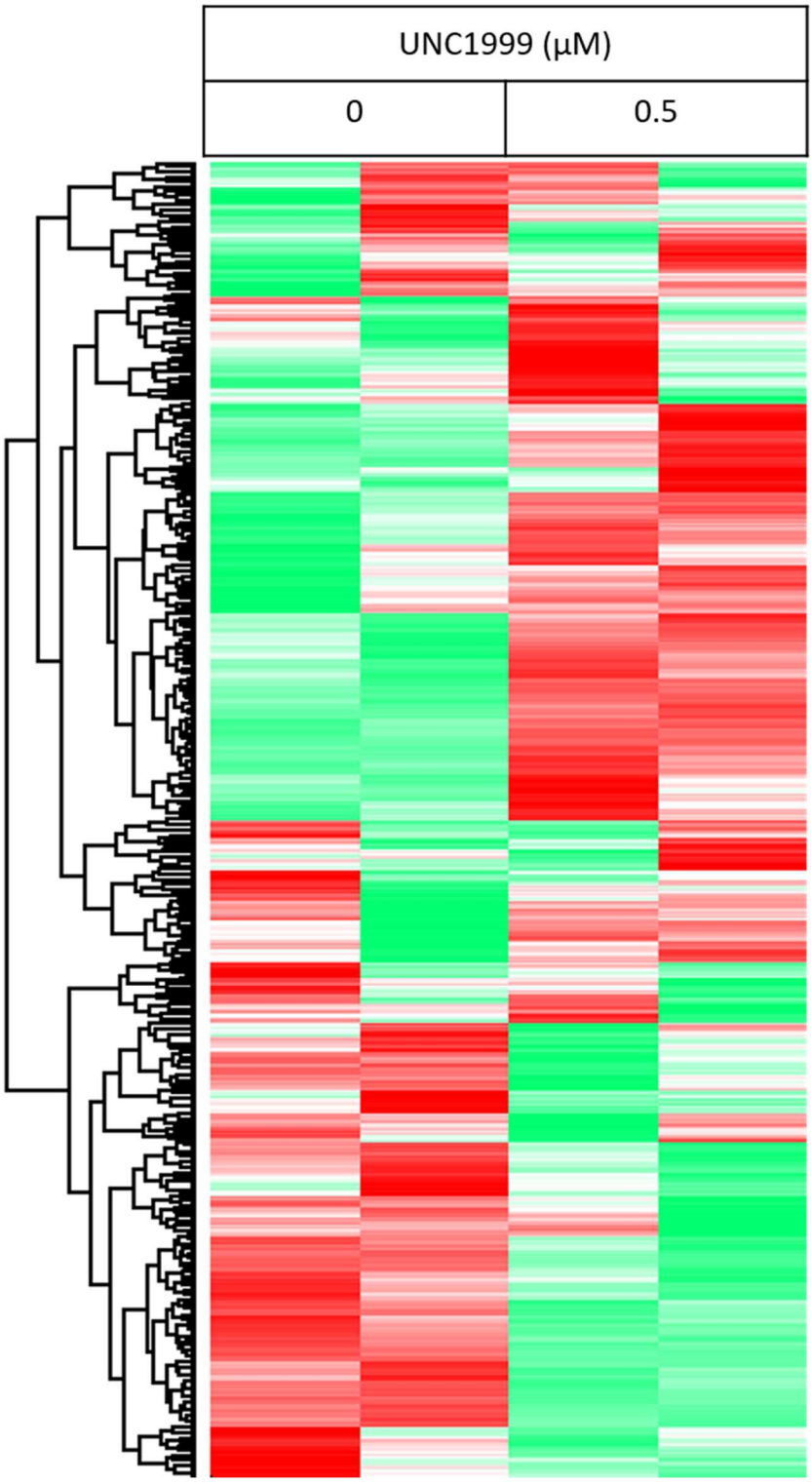


Figure 3.5 - A) Western blot showing decrease in H3K27Me3 indicating effective inhibition of EZH2. B) A heatmap of unsupervised hierarchical clustering analysis of z-score normalized protein secretion MDA231 cells treated for 72 hours with 0.5 μM UNC1999 or DMSO. Each condition is represented by 1 biological replicates and 2 technical replicates. Red indicates higher secretion, green indicates lower secretion, and white indicates mean secretion.

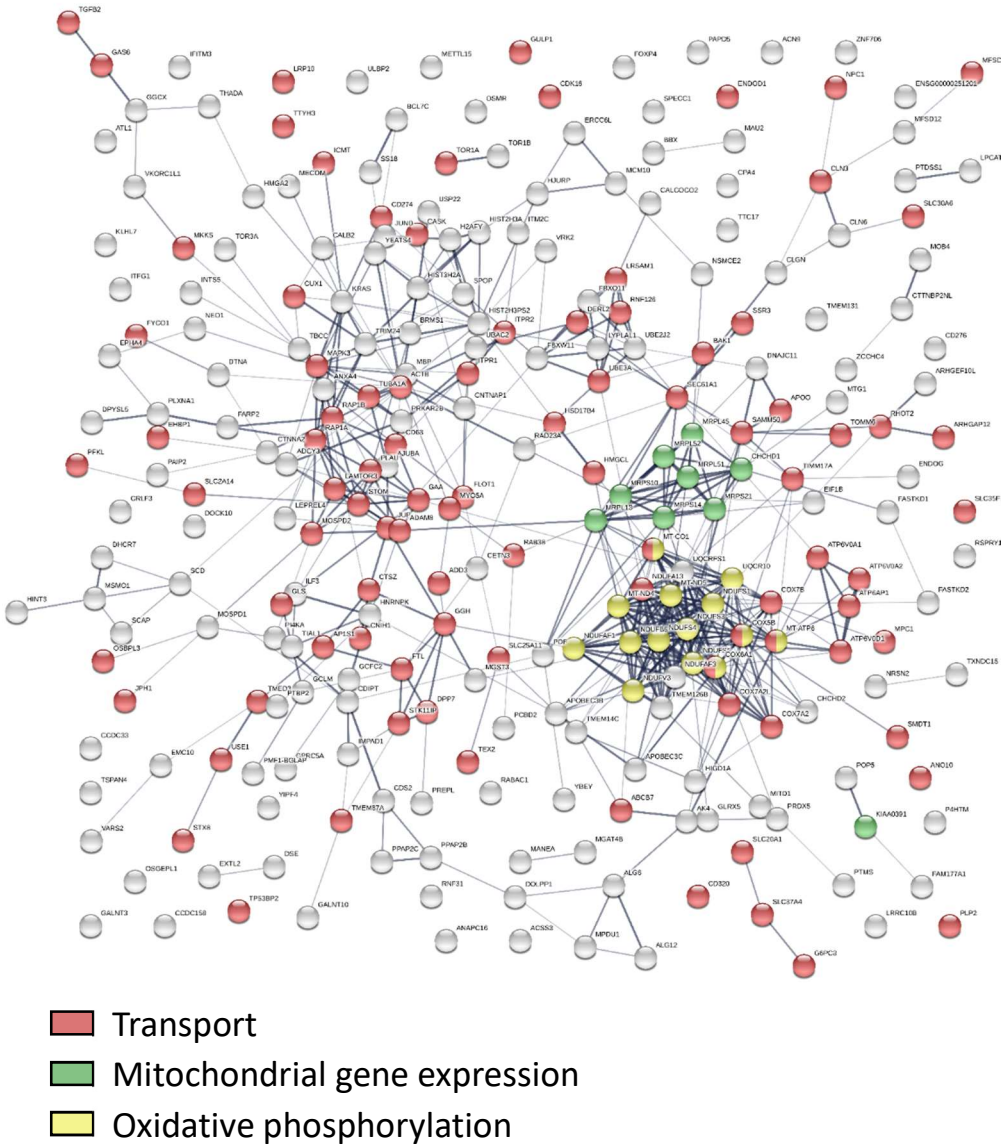


Figure 3.6 – STRING analysis of proteins with increased intracellular abundance after treatment with UNC1999. Select biological processes are highlighted.

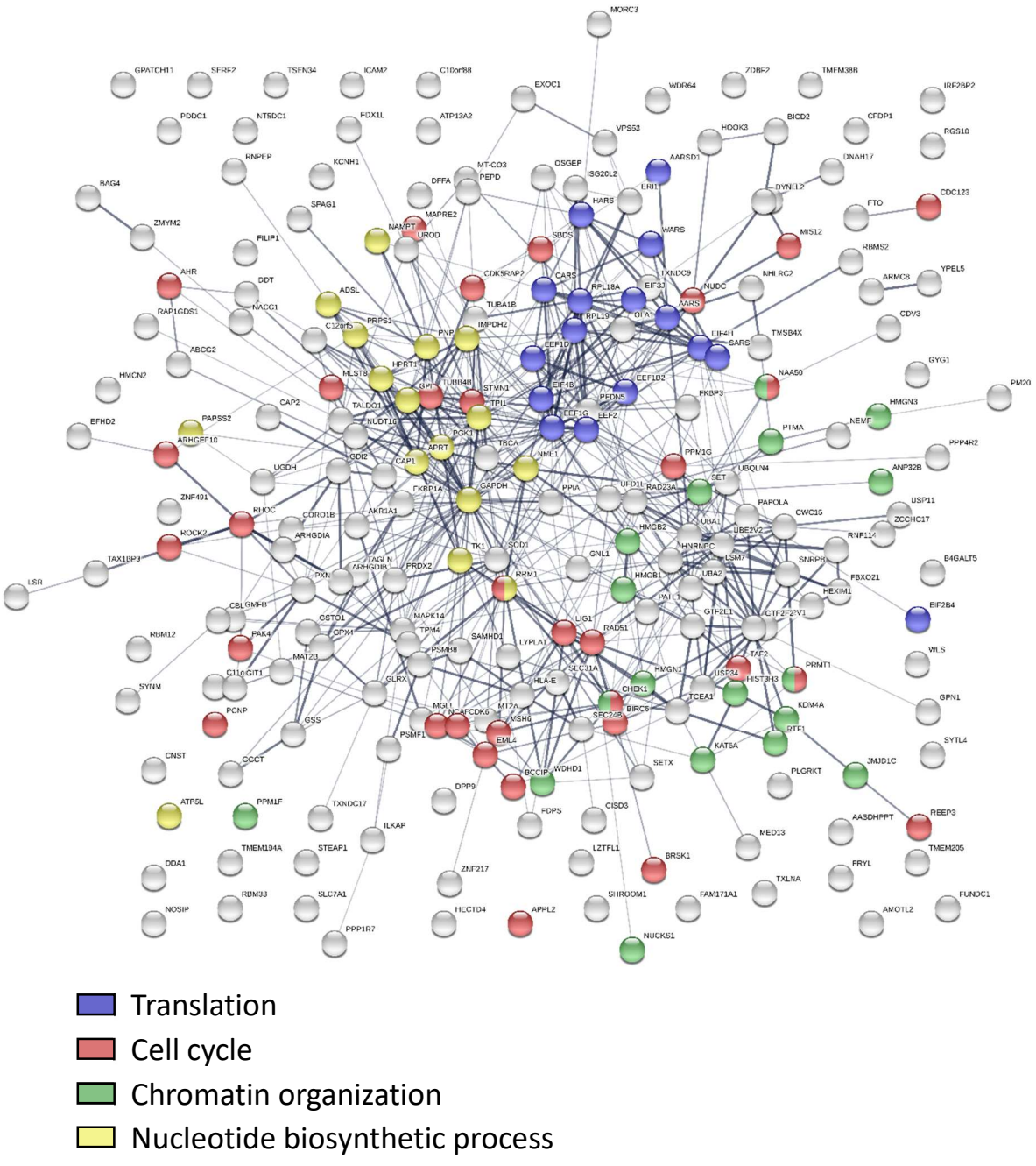


Figure 3.7 – STRING analysis of proteins with decreased intracellular abundance after treatment with UNC1999. Select biological processes are highlighted.

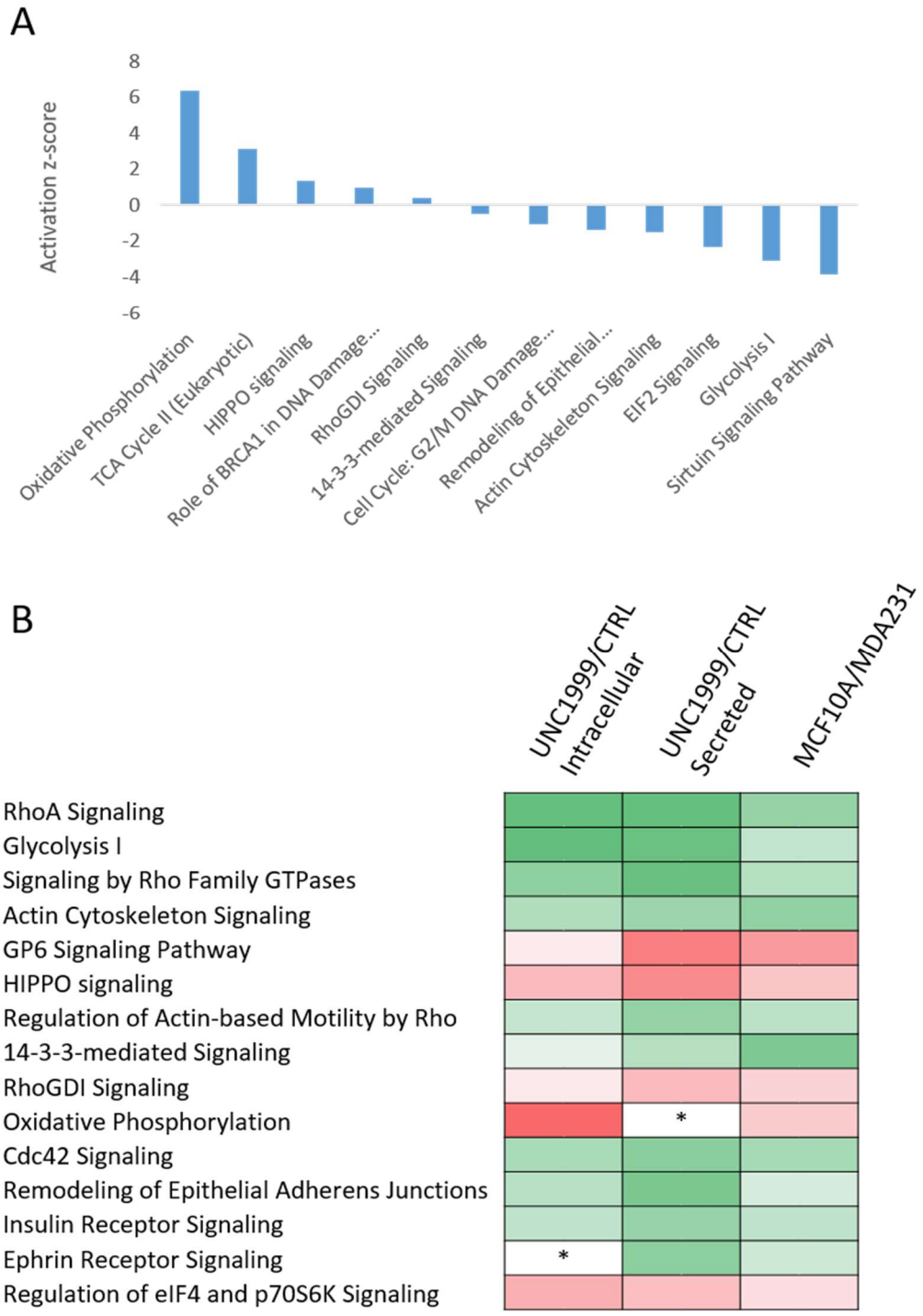


Figure 3.8 - A) Pathway activation analysis of EZH2-dependent signaling. Positive z-scores indicate pathway activation; negative z-scores indicate pathway suppression. B) Comparison of pathway activation based intracellular changes with EZH2 inhibition (left column), secretome changes with EZH2 inhibition (center column), and basal-specific secreted proteins (right column). Red indicates pathway activation and green indicates pathway suppression. Pathways indicated by * are enriched but activation status is uncertain.

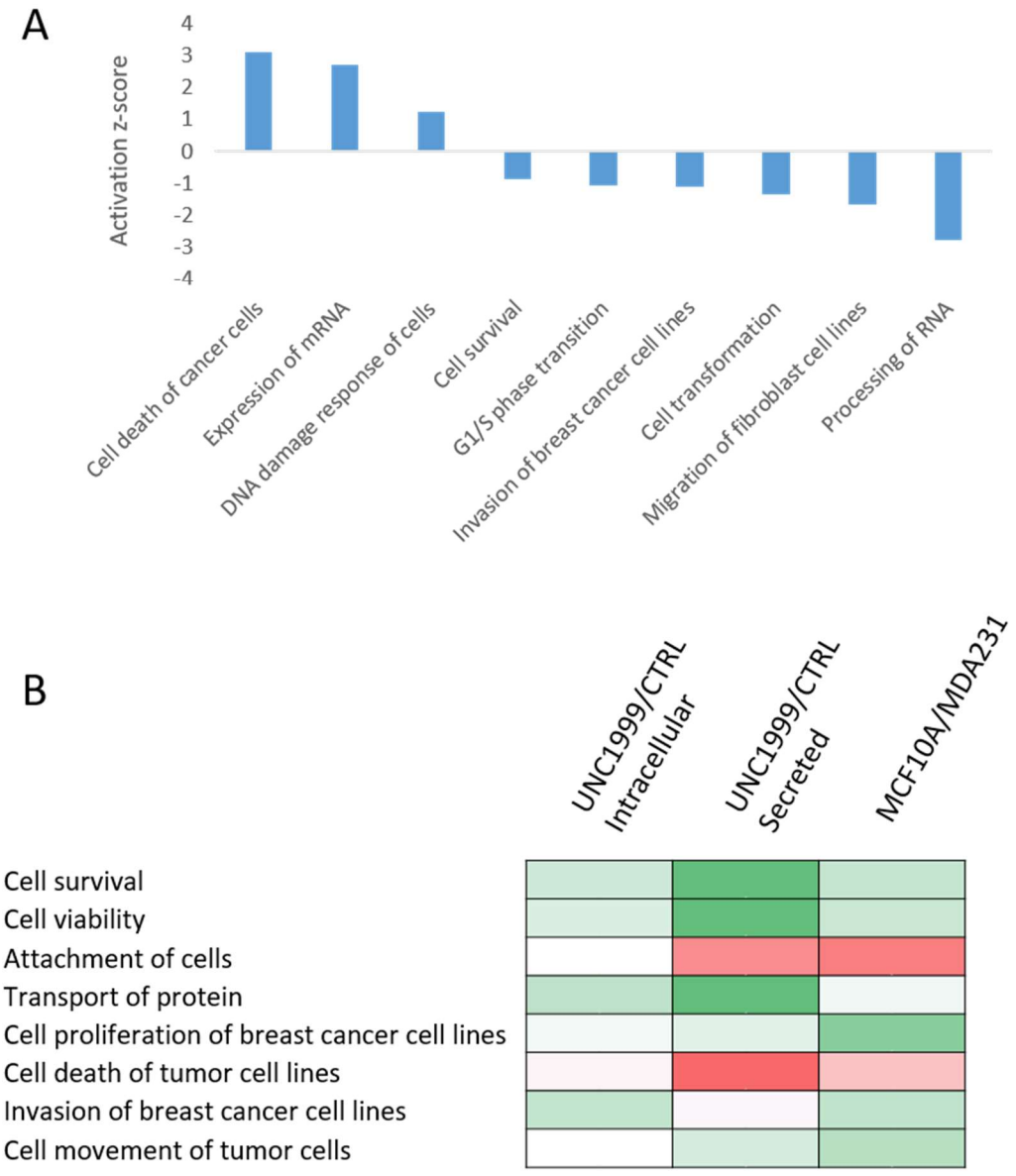


Figure 3.9 - A) Biological function analysis of EZH2-dependent functions. Positive z-scores indicate the function increased after treatment with UNC1999; negative z-scores indicate the function decreased after treatment. B) Comparison of biological function analysis based intracellular changes with EZH2 inhibition (left column), secretome changes with EZH2 inhibition (center column), and basal-specific secreted proteins (right column). Red indicates function increase after UNC1999 treatment and green indicates function decrease.

CHAPTER 4: QUANTITATIVE PROTEOMIC ANALYSIS OF SPLICEOSOME PROTEIN SNRPD1 REVEALS COMPLEX ROLE IN BREAST CANCER CELLS

INTRODUCTION

Most human genes span multiple exons which must be spliced together to form a mature mRNA product. Approximately 90-95% of multi-exon genes undergo alternative splicing, [241, 242] a highly regulated process in which multiple isoforms can be produced by the same gene thereby providing functional diversity to the proteome.[243] Alternative splicing may take the form of differential initiation or termination site usage, intron retention, or exon skipping. (Figure 4.1) However, deregulation of alternative splicing due to e.g. mutations in splicing factors, expression changes in splicing regulators, and mutation of splicing regulator sequences has been observed in many types of cancer, and aberrant splicing patterns are a hallmark of human tumors.[244]

Most splicing is performed by the spliceosome, a multi-megadalton complex having five uridine-rich small nuclear ribonucleoproteins (snRNPs) as the primary components.[245, 246] These snRNPs (U1, U2, U4, U5, and U6) are each composed of a small nuclear RNA (snRNA), seven Sm or SNRP proteins which form a ring structure around the snRNA, and a variable number of complex-specific proteins.[246] Several hundred proteins serve as regulatory factors through interaction with the core machinery.[247] Importantly, SNRP protein overexpression or methylation-induced alterations to SNRP protein activity can cause alternative splicing deregulation.[248-250]

SNRPD1 (also known as SmD1) is one of several SNRP proteins overexpressed in breast cancer.[248] Previous work in our lab demonstrated that SNRPD1 overexpression is

BLBC-specific and is often due to copy number gain.[203] Knockdown of SNRPD1 by siRNA [248] or CRISPR/Cas9 [203] reduced proliferation, survival, and invasion in breast cancer cell lines. Further, SNRPD1 is a major promoter of pluripotency acquisition and maintenance in human stem cells.[251] Independent of the role in alternative splicing, SNRPD1 is essential for microRNA biogenesis in Drosophila.[252]

Despite playing roles in several biological processes and being linked to multiple cancer types, little is known about the effect of SNRPD1 expression at a systems level. Therefore, we set out to characterize the intra- and extracellular effects of SNRPD1 depletion. We performed a comparative quantitative proteomic analysis of MDA-231 cells with CRISPR/Cas9 mediated SNRPD1 knockdown (KD) versus empty vector control. We found that SNRPD1 knockdown caused intracellular proteomic changes which are suggestive of a less aggressive phenotype. Our analysis also revealed that signaling pathway activation after SNRPD1 KD was inconsistent with changes observed in biological process enrichment. Finally, we show that SNRPD1 KD caused a significant increase to protein secretion leading to contradictions between the intracellular and extracellular proteomes.

METHODS

Chemicals and reagents. Cell culture media and fetal bovine serum were obtained from Gibco. All other components of cell culture media and protease inhibitor cocktails were purchased from Sigma (St, Louis, MO). Trypsin was purchased from Promega. All chemicals were HPLC-grade unless specifically indicated. MDA-231 cells were purchased from ATCC (Manassas,VA).

CRISPR/Cas9 knockout. Oligonucleotides for the SNRPD1 sgRNA (Forward-CACCGCCGTACCTGTGATTGTTCCA; Reverse-AAACTGGAACAATCACAGGTACGGC) were annealed and cloned in BsmBI-digested lentiCRISPRv2 (Sanjana et al., 2014). The empty vector was used as a negative control. jetPRIME was used to transfect a total of 10 µg of

plasmids, including target plasmid, pMD2.G, and psPAX2 with a ratio of 10:5:9 into 293T cells. At 48 hours after transfection the media containing the virus were collected. MDA-MB-231 cells at 60-80% confluency were incubated with the virus containing media for 24-48 hours, and then subjected to 1.0 μg/mL puromycin selection. After 4-7 days puromycin selection, the stably transfected cells were collected for further analysis.

Cell culture and secreted protein collection. MDA-231 cells were cultured in DMEM containing 10% fetal bovine serum. For intracellular analysis, cells were harvested at ~80% confluence and lysed with 5x cell pellet volume of RIPA buffer (50 mM Tris pH 8.0, 150mM NaCl, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 1 μM EDTA, 1 μM phenylmethylsulfonyl fluoride). The lysates were sonicated 5 x 1 s at 30%. After 10 minutes of centrifugation at maximum speed, the supernatants were transferred to a fresh tube and the pellets were discarded. Cold acetone (4x volume) was added to the lysates, mixed, and stored overnight at -80°. After centrifugation at maximum speed for 10 minutes, the pellets were resuspended in 50 μl buffer (8 M Urea, 50 mM Tris-HCl pH 8.0, 150 mM NaCl), reduced with dithiothreitol (5 mM final) for 30 minutes at room temperature, and alkylated with iodoacetamide (15 mM final) for 45 minutes in the dark at room temperature. Alkylation was quenched with dithiothreitol (10 mM final). Samples were diluted 4-fold with 25mM Tris-HCl pH 8.0, 1mM $CaCl_2$ and digested with 1:200 (wt/wt) trypsin overnight at room temperature. Peptides were desalted on a StageTip containing a 4 × 1 mm C18 extraction disk (3M) and dried.[139]

For secretome analysis, growth media was removed when cells reached approximately 70% confluence, cells were washed twice with PBS, and serum-free media without phenol red was added to the plate. After 24 hours the conditioned media were collected and centrifuged at 500 x g for 5 minutes to remove cellular debris, then the supernatant was syringe-filtered with 0.2 μm 13 mm diameter polytetrafluoroethylene filters (VWR International) and transferred to fresh tubes. Samples were stored at -80 °C until further processed. After thawing, proteins were concentrated by trichloroacetic acid/sodium deoxycholate precipitation. Briefly, 1/10 of the

sample volume of 0.15% sodium deoxycholate was added to each sample, then tubes were incubated on ice for 15 minutes. Next, 1/10 of the original sample volume of cold 72% trichloroacetic acid was added and the tubes were incubated on ice for 15 minutes. Samples were centrifuged for 10 minutes at max speed, 4° C. The pellets were washed in cold acetone and air dried until no residual odor was detected. Next, the pellets were resuspended in 50 µl buffer (8 M Urea, 50 mM Tris-HCl pH 8.0, 150 mM NaCl), and reduction, alkylation, and trypsinization were performed as with intracellular samples.

LC-MS/MS analysis. LC-MS/MS analysis was performed as previously described.[203] Briefly, desalted peptides were dissolved to a concentration of 1 µg/µl 0.1% formic acid (Thermo-Fisher) for intracellular analysis or in 20 µl 0.1% formic acid for secretome analysis. An injection of 2 µl was analyzed by an Easy nanoLC 1200 with a 15 cm C18 reverse phase column (15 cm × 75 µm ID, C18, 2 µm, Acclaim Pepmap RSLC, Thermo-Fisher) coupled to a Q-Exactive HF-X Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Peptides were eluted at a constant flow rate of 300 nl/min with a gradient of 5-30% buffer B (acetonitrile and 0.1% formic acid) for 75 min, 30%-45% buffer B for 15 min, 45%-100% buffer B for 1 min, and 100% B for 9 min for intracellular analysis or 5-30% buffer B (acetonitrile and 0.1% formic acid) for 17 min, 30%-40% buffer B for 3 min, 40%-100% buffer B for 1 min, and 100% B for 9 min for secretome analysis. Experiments were performed using a data-dependent top 20 method in positive-ion mode. Full MS was performed at a resolution of 60,000 and m/z=200. Up to the top 20 most intense ions with charge ≥ 2 from full MS were selected with an isolation window of 1.4 m/z and higher energy collisional dissociation was used to fragment peptides at a normalized collision energy of 27 eV. The maximum ion injection time for full MS was 100 ms with ion target value of 3e6, and maximum ion injection time for MS/MS was 100 ms with ion target value of 1e5. Selected sequenced ions were dynamically excluded for 30 seconds.

Mass spec data and LFQ analysis. Mass spectral processing and peptide identification were performed on the Andromeda search engine in MaxQuant software (Version 1.5.3.17) against a

human UniProt database. Cysteine carbamidomethylation was set as a defined modification, and methionine oxidation and protein amino-terminal acetylation were set as dynamic modifications. Peptide inference was made with a false discovery rate (FDR) of 1% and peptides were assigned to proteins with FDR of 5%. At least 7 amino acids were required with no more than two missed cleavages. The precursor ion mass tolerance was 8 ppm and the fragment ion mass tolerance was 0.5 Da. Experiments were conducted in multiple replicates (three biological replicates each with two technical replicates) using a match between runs option enabled and time window at 0.7 minutes. Data processing and statistical analysis were performed on Perseus (Version 1.5.1.6).[140]

Analysis of functional category and networks of subtype-specific secreted proteins. The biological processes and molecular functions of secretome proteins were categorized by Ingenuity Pathway Analysis (IPA) [141] and STRING [142] similar to previously described.[143]

RESULTS

SNRPD1 knockdown causes significant widespread changes to diverse cell processes and signaling pathways

To examine the role of SNRPD1 from a systems view, label free quantitative (LFQ) mass spectrometry was performed to analyze the differential expression of CRISPR/Cas9 mediated knockdown of SNRPD1 vs empty vector. A total of 4520 proteins were identified in at least one of three biological replicates with two technical replicates performed for each. Of these, 1156 proteins were found in 2 or fewer replicates and were not further analyzed. Figure 4.2 shows a heat map of the LFQ results. Strikingly, statistical analysis by t test found that 2365 of the remaining 3364 proteins (70.3%) showed statistically significant change in expression (fold-change > 1.5; $p < 0.05$) in SNRPD1 KD cells. Reduction in SNRPD1 expression down was confirmed by MS results (1.6-fold decrease; $p = 3 \times 10\text{-}6$) and western blot (Figure 4.3 a, b).

To determine the biological processes altered by SNRPD1 knockdown, we used STRING to analyze the 2365 proteins with significant changes in expression. Among proteins with decreased expression after SNRPD1 knockdown we found a statistically significant reduction in the expression of proteins involved in ribosome biogenesis, RNA catabolic process, and mRNA splicing via spliceosome. (Figure 4.4) Strikingly, 89% of large and small ribosomal subunit proteins we identified showed decreased expression in the KD cells. (Figure 4.3a) Cell cycle proteins were also overrepresented among proteins with decreased expression after SNRPD1 knockdown. In contrast, expression of proteins which promote differentiation, including tumor suppressor BIN1 [253] increased in SNRPD1 knockout cells. (Figure 4.5) Surprisingly, proteins involved in oxidation-reduction processes, such as PRDX2, PRDX5, and BLVRA which are protective against oxidative stress [254, 255] and may promote cell survival also showed upregulated expression. Finally, knockdown of SNRPD1 caused increased expression of transport proteins involved in protein secretion. These results indicate the contrasting roles of SNRPD1 in suppressing some features, such as elevated secretion and cell survival, which are characteristic of cancer, while promoting other cancer-characteristic features like increased protein expression and cell cycle progression.

We next analyzed signaling pathway activation and suppression using Ingenuity Pathway Analysis (IPA). (Figure 4.6a) Surprisingly, despite STRING analysis indications of a generally less aggressive phenotype of SNRPD1 KD cells, IPA analysis found activation of several signaling pathways associated with cancer progression. For example, integrin signaling, which promotes epithelial-mesenchymal transition and tumor invasion, [256] was highly active after SNRPD1 knockdown. Likewise, we observed an increase in RhoA signaling which has been implicated in cell survival, proliferation, and metastasis.[257] Actin cytoskeleton signaling was also activated in SNRPD1 KD cells, and this pathway is critical to tumor cell motility and invasion.[258]

Despite the surprising activation of these signaling pathways, IPA analysis of biological functions affected by SNRPD1 knockdown are consistent with STRING results. (Figure 4.6b) Biological functions found to be increased included cell death of cancer cells and autophagy of cells. SNRPD1 knockdown by siRNA has previously been reported to promote autophagy.[248] Surprisingly, although EIF2 signaling was suppressed we also observed increased translation. On the other hand, we found decreases in RNA processing and cell cycle progression in addition to cell survival and cell viability. The biological functions of the proteins we have identified do not reflect the apparent activation of signaling pathways which promote cell growth and proliferation.

The secretome reflects consistent pathway activation to intracellular proteome but contradictory biological functions

Given the increased secretion we observed in the analysis of intracellular proteins, we next set out to determine the role of SNRPD1-dependent protein secretion. We compared the secreted proteomic profile of MDA-231 cells with CRISPR-Cas9 medicated knockdown of SNRPD1 versus empty vector. A total of 1215 proteins were identified between the two cell lines, and a heatmap indicating protein quantitation is shown in Figure 4.7. Based on comparison to the Vesiclepedia database 978 (80%) of the identified proteins were previously known to be secreted by breast cancer cell lines, and 1147 (94%) were known to be secreted by at least one of the cell lines or tissue types in the database. These results indicated the excellent quality of our secretome identifications. After removing proteins identified in <2/3 of replicates or identified by only one unique peptide, 702 proteins remained for further analysis.

To identify proteins differentially secreted in a statistically significant manner, we performed a t test analysis on the two groups. (Figure 4.8) We found that 378 of 702 (54%) proteins were differentially secreted after SNRPD1 knockdown. Of these, 237 were secreted in greater abundance by SNRPD1 KD cells. Interestingly, analysis of the biological functions of

these proteins by STRING indicated enrichment of RNA catabolic process, ribosome biogenesis, and mRNA splicing via spliceosome. (Figure 4.9) These results strongly differ from our analysis of the intracellular proteome which found the same processes reduced with SNRPD1 knockdown. We likewise observed increased extracellular abundance of cell cycle proteins in contrast to intracellular reduction of those proteins in SNRPD1 KD cells. On the other hand, increased secretion after SNRPD1 KD was consistent between the intracellular and extracellular results.

STRING analysis of the 141 proteins secreted in lower abundance by SNRPD1 KD cells indicated a less aggressive phenotype after knockdown. For example, we found that SNRPD1 knockdown resulted in secretion of fewer proteins related to cell migration, locomotion, and wound healing. (Figure 4.10) Further, we identified angiogenesis and vasculature development as suppressed processes in the SNRPD1 KD secretome. Overrepresentation of growth and cell adhesion proteins among lower secreted proteins also suggests that SNRPD1 knockdown contributed to reduced aggressive characteristics in the cells.

To determine signaling pathway activation we next applied Ingenuity Pathway Analysis to the secretome data. In general, pathway activation indicated by secreted proteins was consistent with intracellular results. (Figure 4.11a) For example, both secretome and intracellular results indicated increases in RhoA signaling, integrin signaling, and regulation of actin based motility. However, there were also notable discrepancies between the secretome and intracellular pathway activation. For example, the secretome of SNRPD1 KD cells indicated high activation of EIF2 signaling while intracellular proteins indicated strong suppression of the pathway. This result reflects the decrease of intracellular proteins and increase of extracellular proteins related to ribosome biogenesis and RNA processing. On the other hand, despite intracellular activation of IL-8 signaling which promotes angiogenesis and tumor growth, secretome results indicated modest suppression of the pathway.

Analysis of the biological functions of the secreted proteins further accentuated the differences between the intracellular and extracellular proteomes. (Figure 4.11b) Proteins promoting cell survival and proliferation were higher in the secretome of SNRPD1 KD versus control with a concordant decrease in proteins involved in cell death. Interestingly, a decrease in proteins related to tumor invasion, angiogenesis, and metastasis was also observed.

DISCUSSION

To probe the systems-level effects of spliceosome protein SNRDP1 dysregulation we employed a label free proteomic approach for both intra- and extracellular analysis. Using various bioinformatics tools, we compared proteomic alterations occurring in MDA-231 breast cancer cells with CRISPR/Cas9-mediated SNRPD1 knockdown vs an empty vector control. Although SNRPD1 is often overexpressed in aggressive breast cancer and contributes to tumor invasion progression and invasion, [203, 248] to our knowledge this is the first study to examine the role of SNRPD1 on a broad scale.

Knockdown of SNRPD1 caused extensive alterations of signaling pathways and biological functions in MDA-231 cells, with over 70% of proteins quantified having a statistically significant change in expression. These pervasive protein abundance alterations resulted in disruption or promotion of wide-ranging processes including cell signaling, protein secretion, oxidative stress response, metabolism, and the cell cycle. The greatest changes occurred in proteins which interact most closely with SNRPD1, specifically other SNRP proteins, auxiliary spliceosome proteins, and ribosomal proteins. EIF2 signaling, which drives mRNA translation initiation, was consequently suppressed.

The overall effect of SNRPD1 knockdown was unclear. Although proteins involved in differentiation increased and cell cycle decreased suggested a less aggressive phenotype in SNRPD1 KD cells, signaling pathway analysis indicated activation of several pathways associated with cancer proliferation, invasion, and metastasis. In previous studies, SNRPD1

knockdown caused decreases to proliferation and invasive potential of MDA-231 cells. These results suggest that the cellular phenotype is changing in spite of pathway activation, likely due to the lack of available translational machinery. Additional work finding a buildup of untranslated transcripts would support this proposal.

Secretome analysis was likewise enigmatic. Although pathway activation was nearly identical to intracellular results, the biological functions associated with the proteins identified were nearly opposite. Three observations lead us to hypothesize that the SNRPD1 KD cells are ejecting excess cellular content into the extracellular space: 1) secretion was increased in the knockdown cell lines; 2) many intracellular proteins including ribosomal proteins were increased in the KD secretome; and 3) the secretome starkly contrasts the intracellular proteome. The presence of intracellular proteins in the extracellular space is likely not due to cell death. Although SNRPD1 depletion using siRNA caused cell death, [248] our knockdown resulted in a more modest 1.6-fold decrease in SNRPD1 abundance. We observed a healthy appearance and continued growth of the SNRPD1 KD cells, albeit at a slower rate than the control cell lines. (data not shown) Further, IPA indicated only modest apoptosis signaling in any experiment. The clearly distinguished results between intracellular and secretome biological functions also suggests mass protein expulsion from the cell, wherein certain functions are underrepresented intracellularly because those proteins were secreted, causing extracellular overrepresentation, and not replenished.

There are some important limitations to this study. First, the role of SNRPD1 is likely due to aberrant alternative splicing. However, due to proteomic coverage limitations of mass spectrometry, alternative isoforms may not be properly identified or quantitated. Importantly, trypsin preferentially cleaves peptides at exon-exon junctions, so junction-spanning peptides are rare. A multi-omic approach integrating transcriptomic experiments could significantly expand these findings. Additionally, our work has focused only on SNRPD1 knockdown.

Overexpression of SNRPD1 in a less aggressive breast cancer cell line, such as MCF7 or T47D, should be performed to support our results.

CONCLUSIONS

In summary, our systems-level analysis of SNRPD1 function by comparative quantitative proteomic analysis revealed complex and far-ranging alterations to both the intracellular and extracellular proteomes. The biological function changes associated with SNRPD1 knockdown suggest SNRPD1 may be a therapeutic target for reducing breast cancer aggressiveness despite signaling pathway activation typically associated with growth and proliferation.

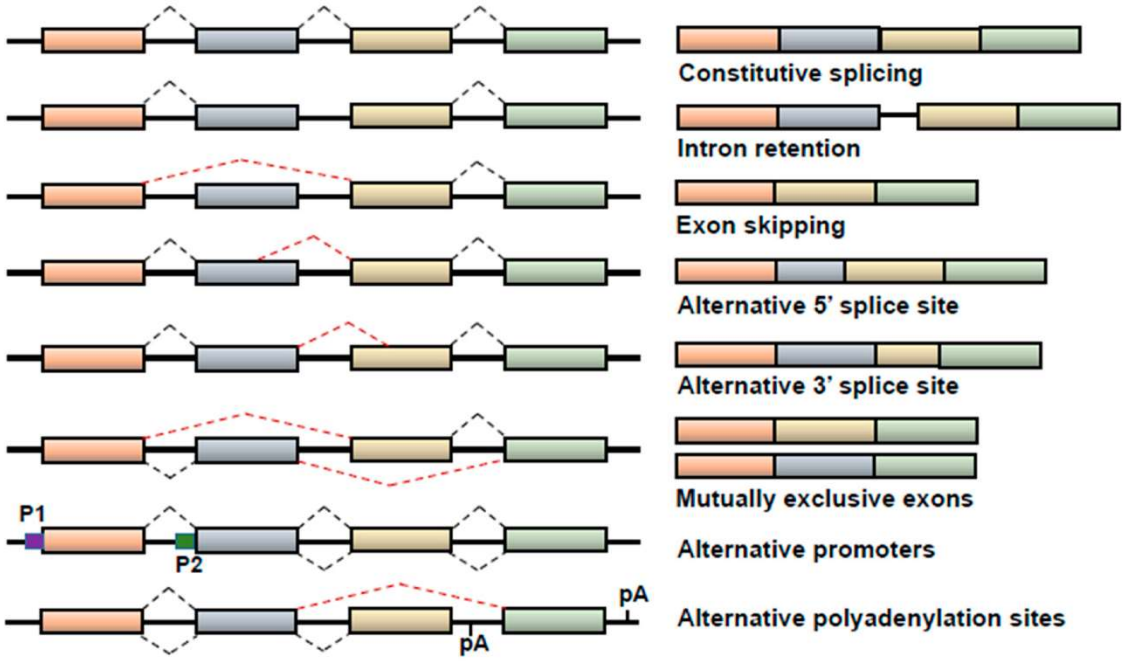


Figure 4.1 - Various forms of alternative splicing are shown. A pre-mRNA is shown on the left and possible mature mRNA variants are shown on the right. Reprinted from Wang and Lee under the Creative Commons Attribution License. (https://creativecommons.org/licenses/by/4.0/)

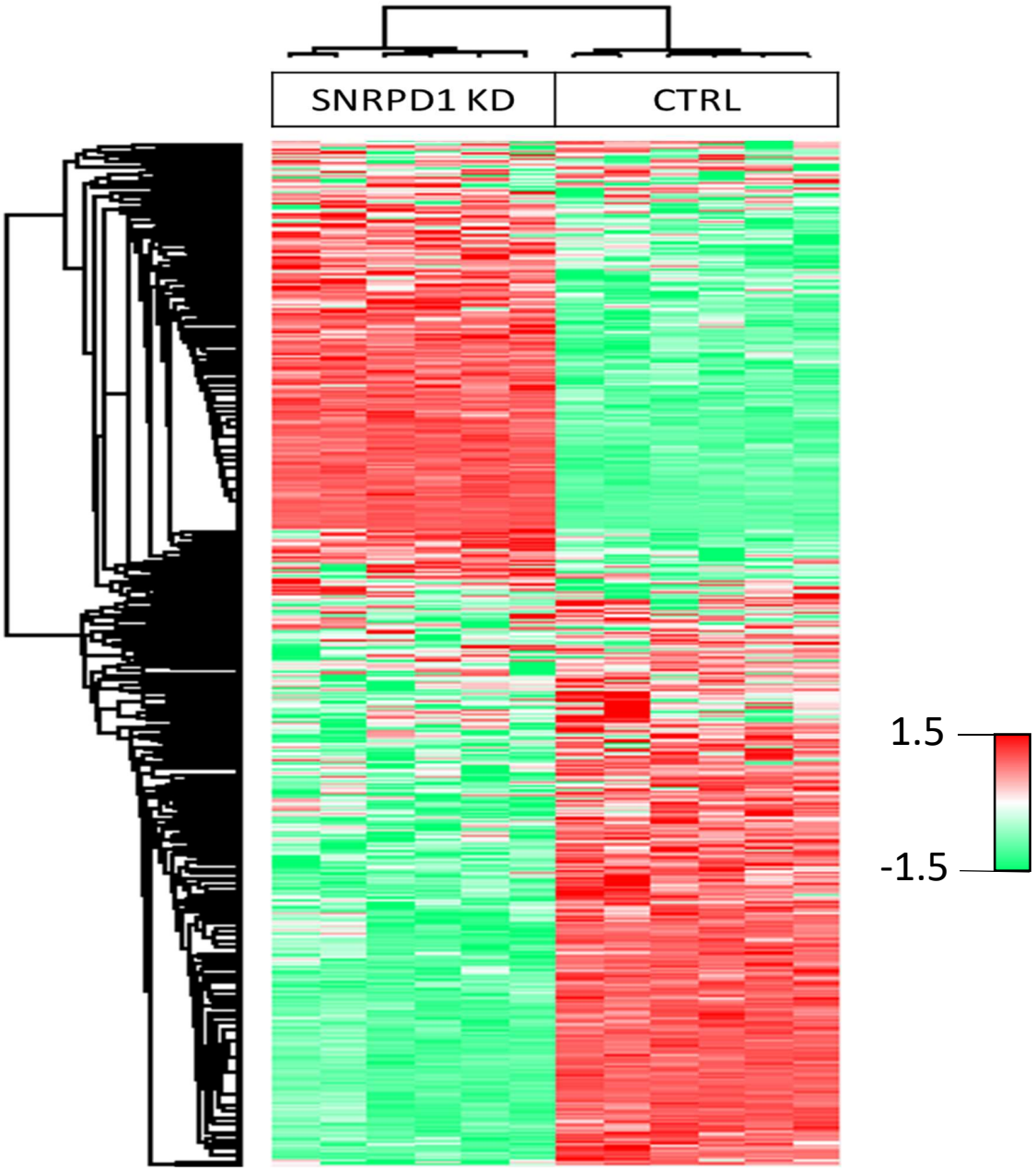


Figure 4.2 - A heatmap of unsupervised hierarchical clustering analysis of z-scored intracellular protein abundance. Each row indicates a protein and each column indicates a replicate measurement. Two technical replicates were performed for each of three biological replicate of SNRPD1 KD or empty vector control (CTRL) MDA-231 breast cancer cells. Red indicates higher abundance, green indicates lower abundance, and white indicates mean abundance.

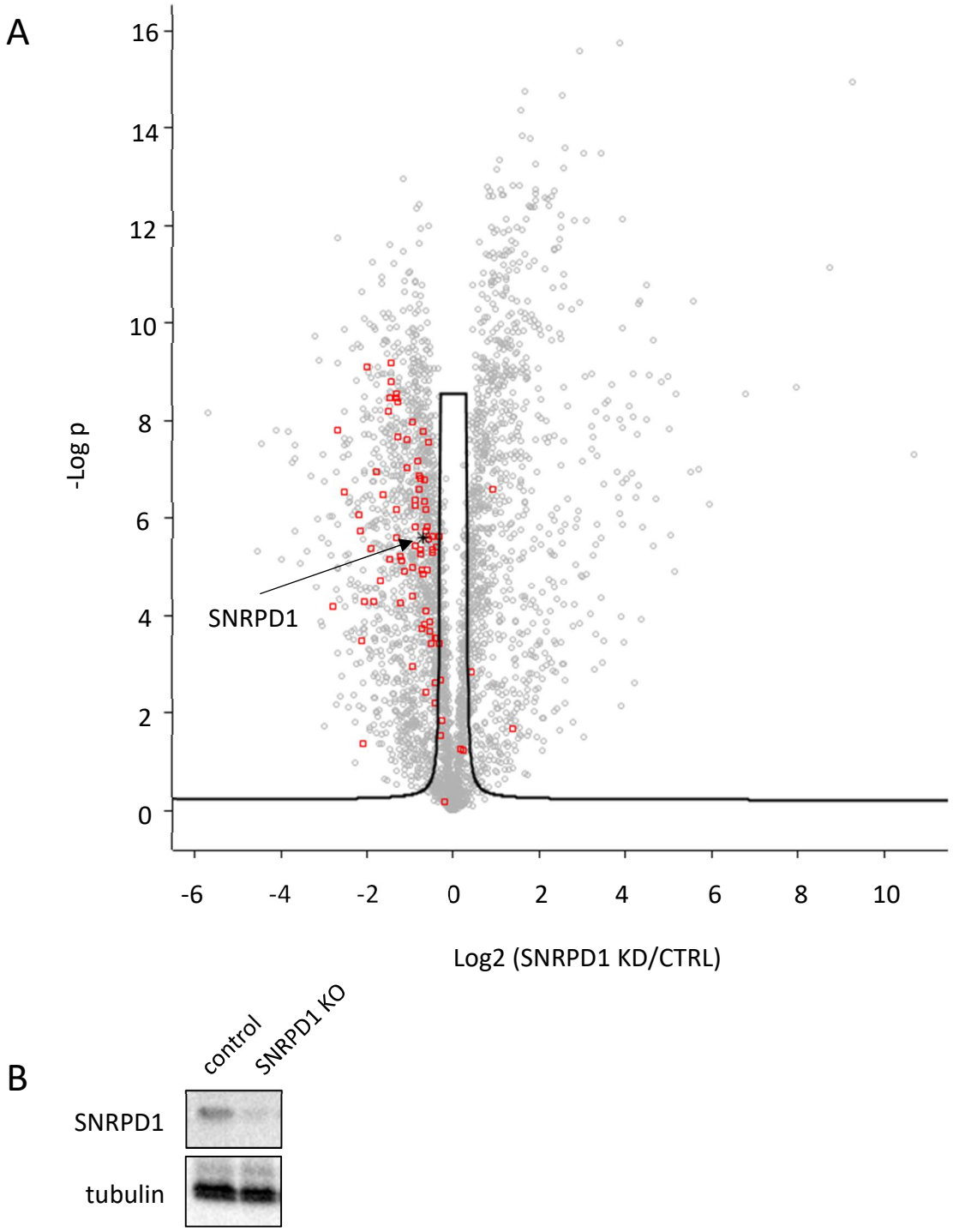


Figure 4.3 – A) Volcano plot of intracellular proteins identified in SNRPD1 KD and empty vector control (CTRL) MDA-231 breast cancer cells. Positive values indicate higher expression in SNRPD1 KD cells and negative values indicate higher expression in CTRL cells. Ribosomal proteins with statistically significant altered exression and fold change >1.5 are indicated in red. SNRPD1 is indicated by a black star. B) Western blot of SNRPD1 knockdown.

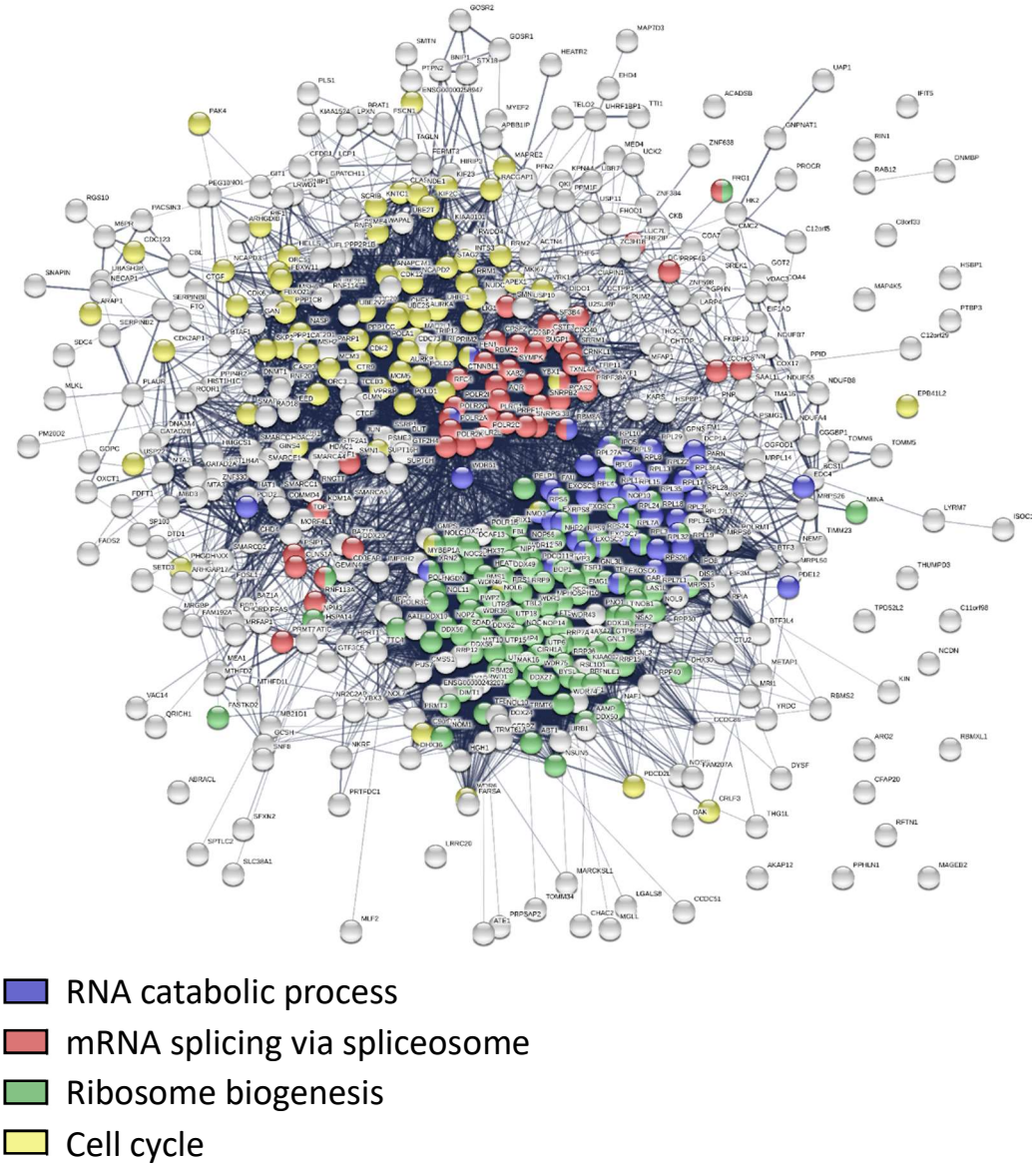


Figure 4.4 – STRING interaction network of downregulated intracellular proteins in SNRPD1 KD cells. Select biological processes have been highlighted.

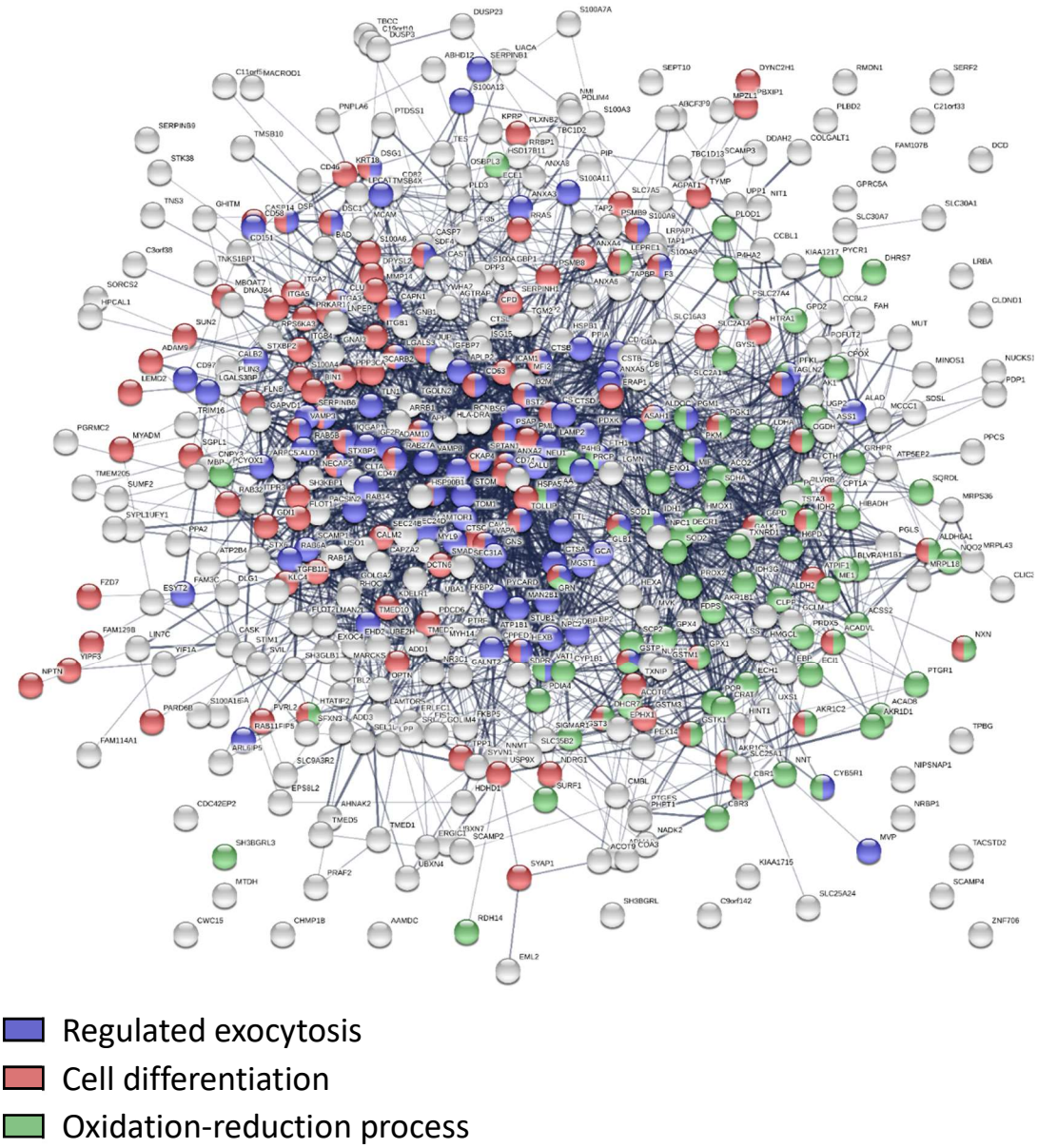


Figure 4.5 – STRING interaction network of upregulated intracellular proteins in SNRPD1 KD cells. Select biological processes have been highlighted.

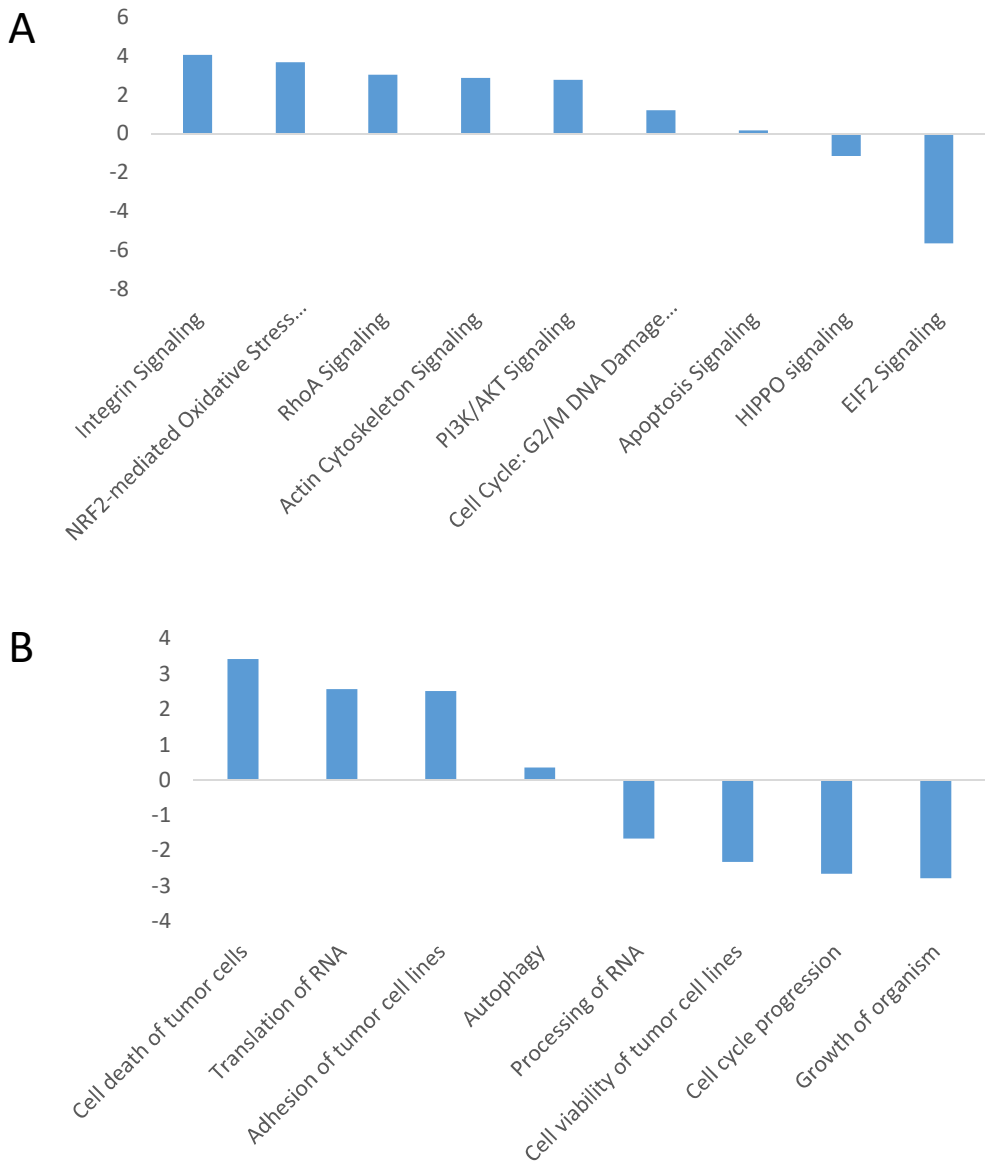


Figure 4.6 – IPA analysis of changes in intracellular proteins in SNRPD1 KD cells. A) IPA biological function analysis. Positive z-scores indicate biological function increase in SNRPD1 KD cells; negative z-scores indicate biological function decrease in SNRPD1 KD cells. B) IPA pathway activation analysis. Positive z-scores indicate pathway activation; negative z-scores indicate pathway suppression.

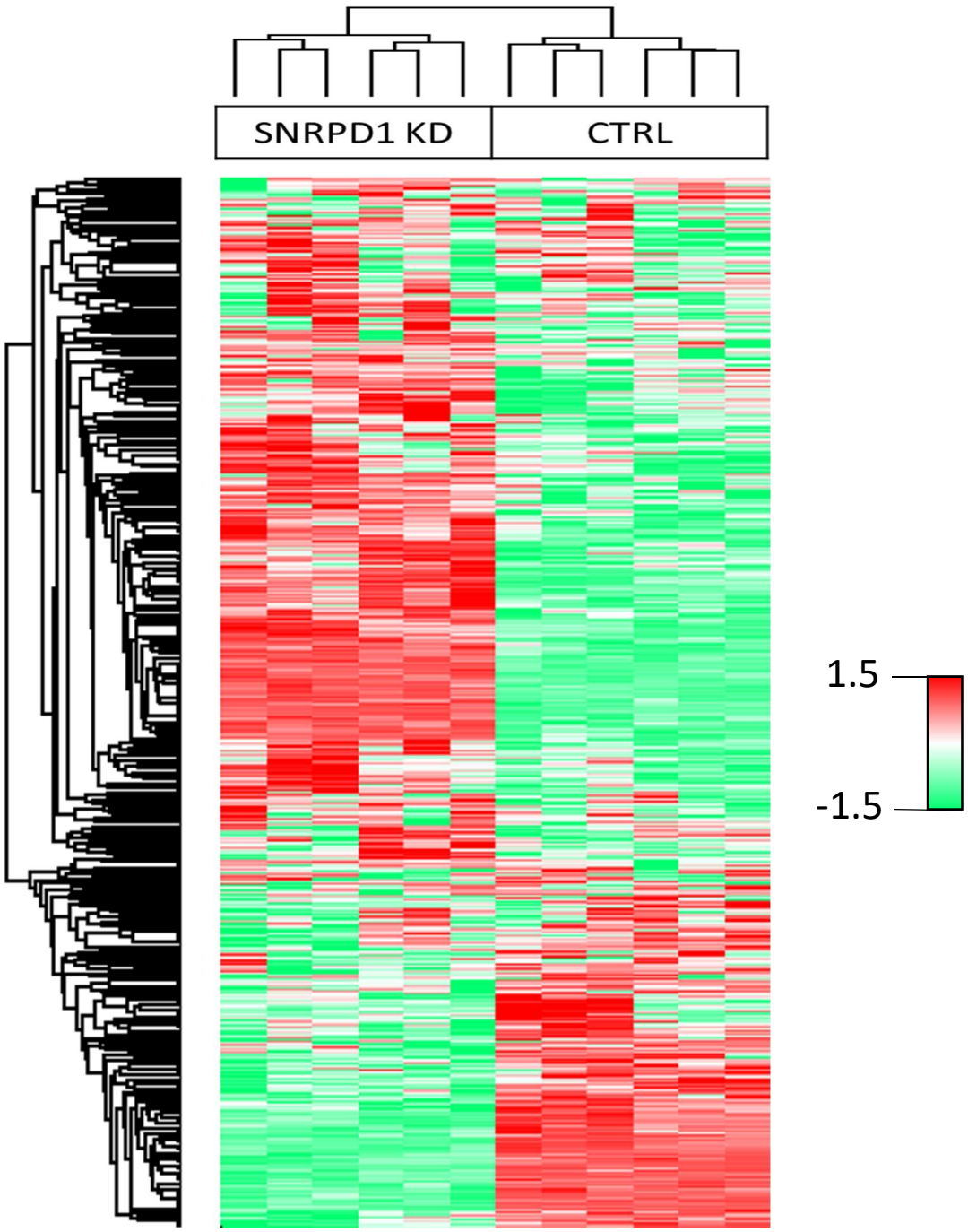


Figure 4.7 - A heatmap of unsupervised hierarchical clustering analysis of z-scored extracellular protein abundance. Each row indicates a protein and each column indicates a replicate measurement. Two technical replicates were performed for each of three biological replicate of SNRPD1 KD or empty vector control (CTRL) MDA-231 breast cancer cells. Red indicates higher secretion, green indicates lower secretion, and white indicates mean secretion.

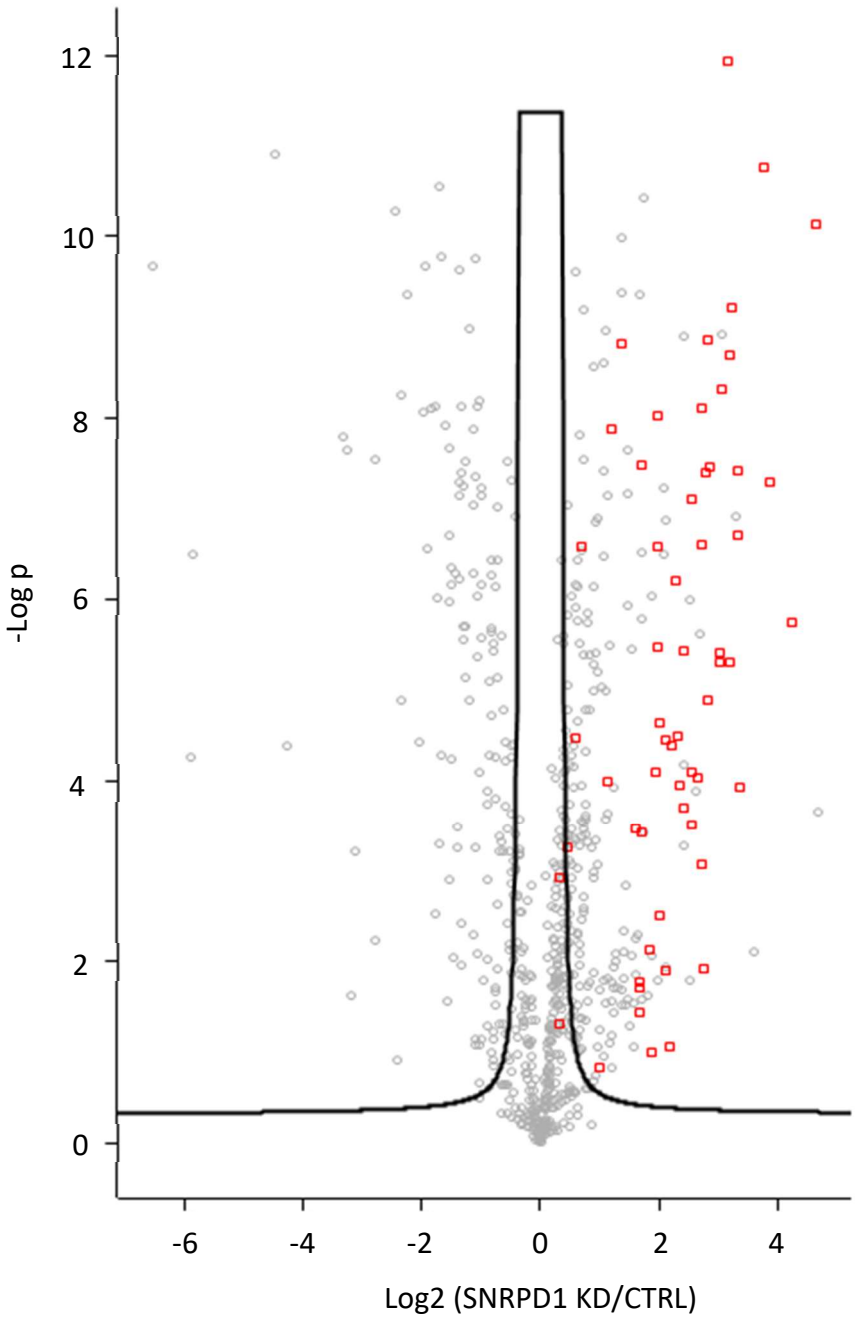


Figure 4.8 – Volcano plot of secreted proteins identified in SNRPD1 KD and empty vector control (CTRL) MDA-231 breast cancer cells. Positive values indicate higher secretion by SNRPD1 KD cells and negative values indicate higher secretion by CTRL cells. Ribosomal proteins with statistically significant altered expression and fold change >1.5 are indicated in red.

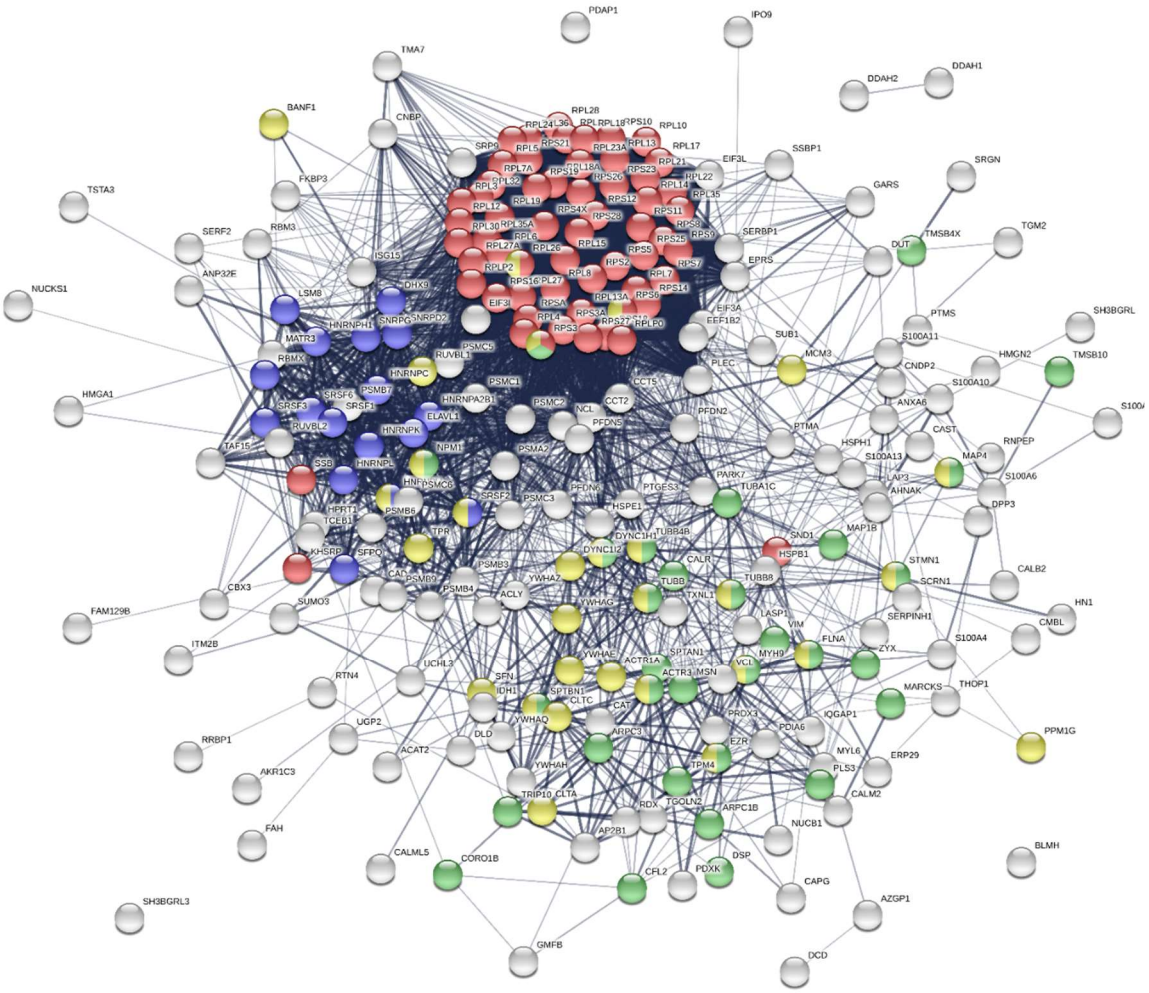
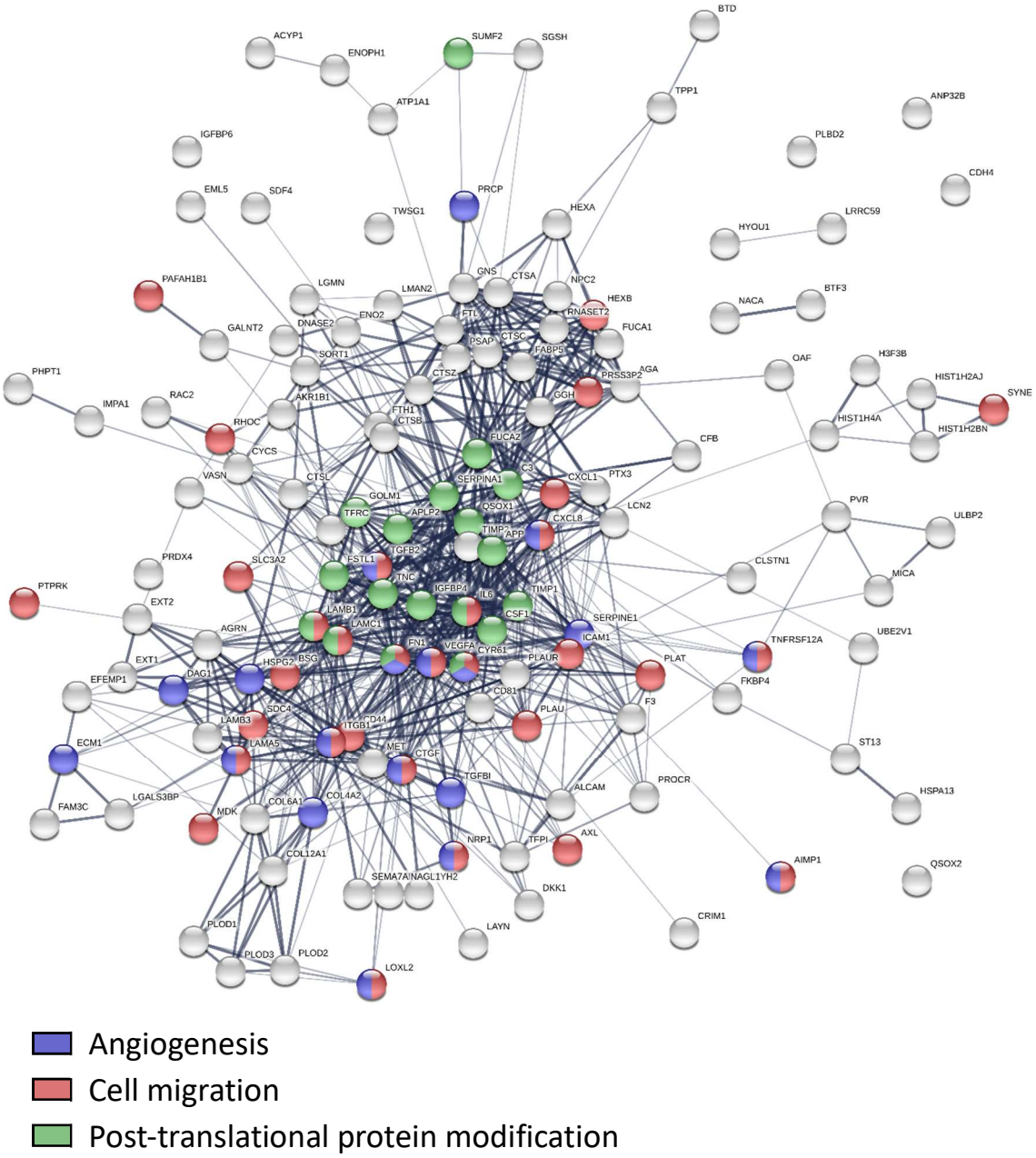


Figure 4.9 – STRING interaction network of higher secreted proteins in SNRPD1 KD cells. Select biological processes have been highlighted.



Angiogenesis

Cell migration

Post-translational protein modification

Figure 4.10 – STRING interaction network of higher secreted proteins in SNRPD1 KD cells. Select biological processes have been highlighted.

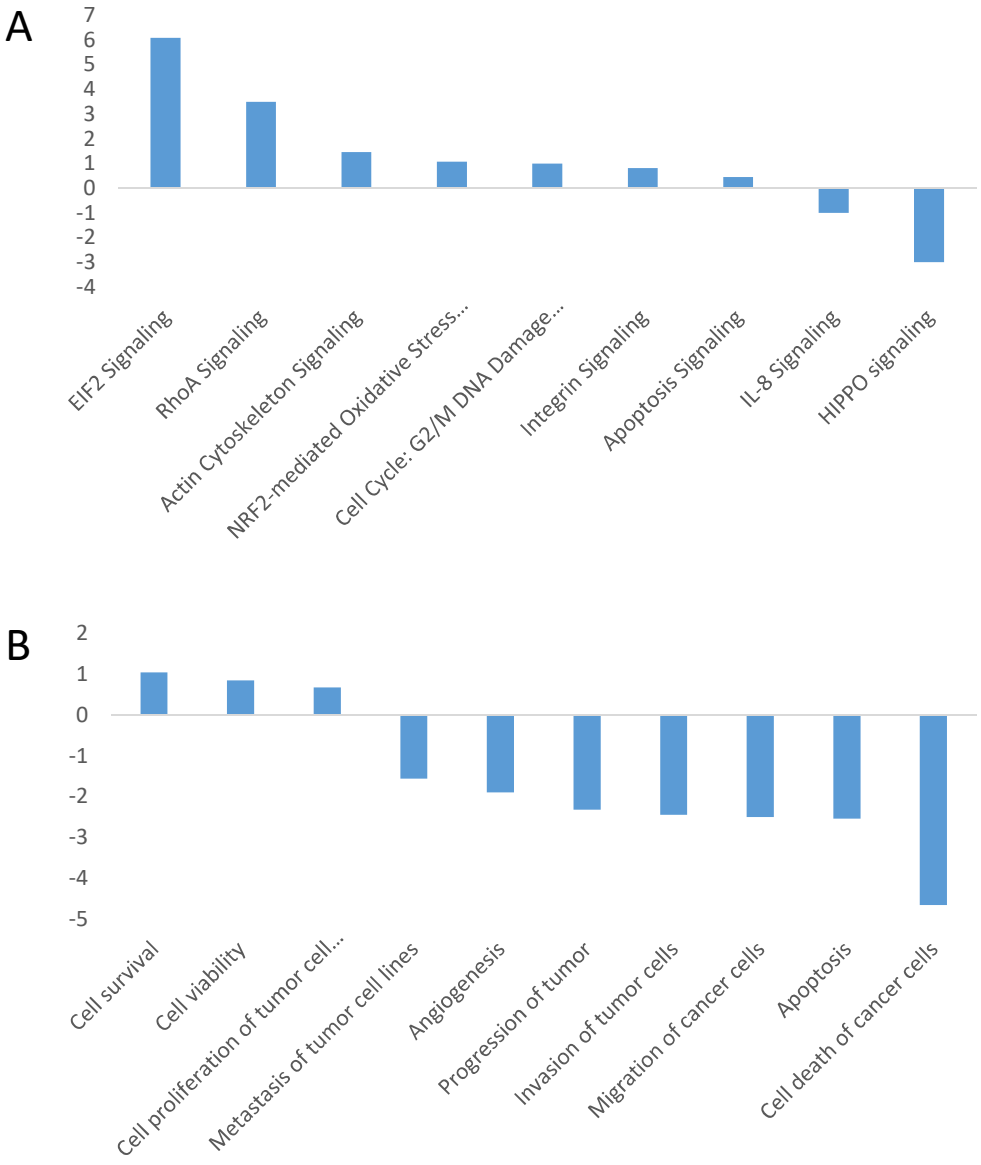


Figure 4.11 – IPA analysis of changes in secreted proteins in SNRPD1 KD cells. A) IPA biological function analysis. Positive z-scores indicate biological function increase in secretome of SNRPD1 KD cells; negative z-scores indicate biological function decrease in secretome of SNRPD1 KD cells. B) IPA pathway activation analysis. Positive z-scores indicate pathway activation; negative z-scores indicate pathway suppression.

CHAPTER 5: PROTEOMIC DISSECTION OF SYSTEMS-LEVEL EFFECTS OF IDH1 MUTATION IN DIFFUSE GLIOMA

INTRODUCTION

Gliomas are the most common type of primary brain tumor and are typically classified as grade I to IV based on histopathological and clinical criteria.[259] Grade I gliomas are typically benign, rarely transform to invasive cancer, and are often curable by surgery alone.[260] Grade IV gliomas, or glioblastoma multiforme (GBM), are highly aggressive tumors, offering patients very poor prognosis and among the lowest survival rates of any cancers.[260] About 20% of GBM are secondary tumors arising from the progression of a grade II or III tumor.[261] Among grade II and III tumors, approximately 80% have a mutation in the isocitrate dehydrogenase 1 (IDH1) enzyme which is typically R132H although other variants have been reported.[262] This mutation is an early, if not initiating, step in oncogenesis.[263] Despite the role of IDH1 mutation in oncogenic transformation, patients with these mutations have longer median overall survival rates than patients with the wild type enzyme.[264]

The R132H mutation of IDH1 results in a number of contributory downstream pathogenic effects. The primary consequence of mutation is the inhibition of IDH1 enzymatic function in conversion of isocitrate to α-ketoglutarate (α–KG) resulting in disruption of the TCA cycle, increased abundance of isocitrate, and decreased availability of α-KG.[260] In a further upset to the metabolic balance, R132H mutation imparts gain-of-function activity to IDH1 whereby the enzyme promotes the conversion of α-KG to R-2-hydroxyglutarate (R-2HG).[265] R-2HG is an oncometabolite which disrupts normal cellular function through diverse mechanisms, the full extent of which are currently poorly understood. Due to structural similarities, R-2HG can act as

an inhibitor to α-KG dependent enzymes including less sensitive species due to the potential millimolar concentrations of R-2HG in the cytoplasm of IDH1-mutated tumors.[265] One class impacted enzymes is α-KG-dependent dioxygenases which have been associated with tumorigenesis and cancer progression.[266] Importantly, α-KG dependent dioxygenases include the Jumonji family of histone demethylases and the ten eleven translocation (TET) family of 5-methylcytosine hydroxyases.[267] The inhibition of these enzymes leads to increased histone methylation and a CpG island methylator phenotype respectively in IDH1 mutated gliomas.[268] This hypermethylation results in suppression of important tumor suppressor genes and chromatin structure remodeling.[268]

In addition to epigenetic effects, aberrant α-KG-dependent dioxygenase activity also causes aberrant collagen maturation and response to hypoxia. Proline and lysine residues on collagen IV, a major component of the basement membrane, are subject to modification by α-KG-dependent hydroxylation which is critical for stability of the triple helix structure of mature collagen.[269] Disruption of this process by R-2HG leads to a fragile basement membrane, increased type IV collagen solubility, and accumulation of misfolded proteins in the endoplasmic reticulum.[270] Finally, despite inconsistency in previous reports, most recent studies show IDH1 mutation suppresses HIF1-α expression even in hypoxic conditions.[271-274] Notably, reports finding IDH1 causes upregulation of HIF1-α often utilized the U87 glioma cell line with overexpressed R132H IDH1 [275] or an acute myeloid leukemia mouse model with R132H mutation in myeloid progenitor cells.[270, 276] On the other hand, those testing endogenous mutant IDH1 in patient-derived glioma cells found HIF1-α suppressed.[271-274] The likely mechanism of HIF1-α suppression is via R-2HG activation of EGLN1 which promotes HIF1-α affinity for VHL ubiquitin ligase, thereby promoting HIF1-α proteosomal degradation.[274] Because of the significant role in glioma biology, R-2HG is an important marker which can be monitored non-invasively by MRI to track therapeutic progress.[277, 278]

Despite recent advances in our understanding of IDH1 mutation effects, the full impact of the resulting metabolic disruption is not known. To investigate additional biological processes and pathways which may be affected by IDH1 mutation, we employed a label-free quantitative (LFQ) proteomic approach to comparatively analyze intracellular proteins in patient-derived IDH1 mutant vs wild type glioma cells. Our data were validated by corroboration of known effects of IDH1 mutations, and we report several new pathways and processes which may contribute to IDH1 mutant glioma pathogenesis or progression.

METHODS

Chemicals and reagents. Cell culture media and fetal bovine serum were obtained from Gibco. All other components of cell culture media and protease inhibitor cocktails were purchased from Sigma (St, Louis, MO). Trypsin was purchased from Promega. All chemicals were HPLC-grade unless specifically indicated. MDA-231 cells were purchased from ATCC (Manassas,VA).

Cell culture. Patient-derived glioma cells were cultured in NeuroCult™ NS-A Basal Medium (Human) supplemented with NeuroCult™ NS-A Proliferation Kit (Human), 2 µg/ml heparin, 10 ng/ml human recombinant bFGF, and 20 ng/ml human recombinant EGF. Cells were harvested at ~80% confluence and lysed with 5x cell pellet volume of RIPA buffer (50 mM Tris pH 8.0, 150mM NaCl, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 1 µM EDTA, 1 µM phenylmethylsulfonyl fluoride). The lysates were sonicated 5 x 1 s at 30%. After 10 minutes of centrifugation at maximum speed, the supernatants were transferred to a fresh tube and the pellets were discarded. Cold acetone (4x volume) was added to the lysates, mixed, and stored overnight at -80°. After centrifugation at maximum speed for 10 minutes, the pellets were resuspended in 50 µl buffer (8 M Urea, 50 mM Tris-HCl pH 8.0, 150 mM NaCl), reduced with dithiothreitol (5 mM final) for 30 minutes at room temperature, and alkylated with iodoacetamide (15 mM final) for 45 minutes in the dark at room temperature. Alkylation was

quenched with dithiothreitol (10 mM final). Samples were diluted 4-fold with 25mM Tris-HCl pH 8.0, 1mM $CaCl_2$ and digested with 1:200 (wt:wt) trypsin overnight at room temperature. Peptides were desalted on a StageTip containing a 4 × 1 mm C18 extraction disk (3M), separated into 100 µg aliquots, and dried.[139] Peptides were resuspended in 22 µl 10 mM ammonium formate (pH 10) and fractionated by reverse phase fractionation. The fractions were combined in a non-continuous manner as described.[279] Each fraction was desalted and dried as before.

LC-MS/MS analysis. LC-MS/MS analysis was performed as previously described.[203] Briefly, desalted peptides were dissolved to a concentration of 1 µg/µl 0.1% formic acid (Thermo-Fisher). An injection of 4 µl was analyzed by an Easy nanoLC 1000 with a 15 cm C18 reverse phase column (15 cm × 75 µm ID, C18, 2 µm, Acclaim Pepmap RSLC, Thermo-Fisher) coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Peptides were eluted at a constant flow rate of 300 nl/min with a gradient of 2-30% buffer B (acetonitrile and 0.1% formic acid) for 150 min, 30%-80% buffer B for 15 min, and 80% B for 15 min. Experiments were performed using a data-dependent top 20 method in positive-ion mode. Full MS was performed at a resolution of 70,000 and m/z=200. Up to the top 20 most intense ions with charge ≥ 2 from full MS were selected with an isolation window of 2.0 m/z and higher energy collisional dissociation was used to fragment peptides at a normalized collision energy of 27 eV. The maximum ion injection time for full MS was 250 ms with ion target value of 1e6, and maximum ion injection time for MS/MS was 120 ms with ion target value of 2e5. Selected sequenced ions were dynamically excluded for 30 seconds.

Mass spec data and LFQ analysis. Mass spectral processing and peptide identification were performed on the Andromeda search engine in MaxQuant software (Version 1.5.3.17) against a human UniProt database. Cysteine carbamidomethylation was set as a defined modification, and methionine oxidation and protein amino-terminal acetylation were set as dynamic modifications. Peptide inference was made with a false discovery rate (FDR) of 1% and peptides were assigned to proteins with FDR of 5%. At least 7 amino acids were required with

no more than two missed cleavages. The precursor ion mass tolerance was 8 ppm and the fragment ion mass tolerance was 0.5 Da. Experiments were conducted in multiple replicates (three biological replicates each with two technical replicates) using a match between runs option enabled and time window at 0.7 minutes. Data processing and statistical analysis were performed on Perseus (Version 1.5.1.6).[140]

Analysis of functional category and networks of subtype-specific secreted proteins. The biological processes and molecular functions of secretome proteins were categorized by Ingenuity Pathway Analysis (IPA) [141] and STRING [142] similar to previously described. [143]

RESULTS

To further understand the role of IDH1 mutation in glioma, we compared the differential protein expression of IDH1 mutant and wild type glioma cells using LFQ analysis. One biological replicate of each condition was analyzed with two technical replicates each. With the aim of maximizing the number of proteins identified in order to gain a more complete view of the systems involved, we performed high pH fractionation of each sample and combined the fractions in a concatenated fashion such that non-consecutive fractions were combined at equal intervals. By performing an orthogonal separation prior to the inline low pH separation employed during LC-MS/MS, we identified 10288 proteins in at least one replicate. After removing proteins that were identified in only one technical replicate or by only one unique peptide, 8161 proteins remained for analysis. A heatmap showing protein quantitation (Figure 5.1) clearly demonstrates the significant alterations to protein expression with IDH1 mutation. Statistical analysis by t test found that 4985 proteins (61%) were differentially expressed between the samples. (Figure 5.2) Of these, 2416 were upregulated and 2569 were downregulated. The upregulated proteins included PDGFRA and PGK1 which are known to be increased due to IGH1 mutation.[273, 280] Downregulated proteins included CTNNB1 which was previously found lower in IDH1 mutant cells.[281]

To identify biological processes dysregulated by IDH1 mutation, we used STRING to determine enrichment of Gene Ontology Biological Process (GOBP) terms. Strikingly, over half (57%) of proteins with increased expression were involved in metabolism, and 24 of the 25 most enriched GOBP terms were related to metabolic processes. (Figure 5.3 a,b) These included oxoacid metabolic process, nucleotide metabolic process, and cellular amino acid metabolic process, indicating the wide ranging impact of IDH1 mutation on cellular metabolism. Surprisingly, considering the less aggressive nature of IDH1 mutant gliomas, we also found that proteins participating in neurogenesis, neuron projection development, and regulation of neurotransmitters were also increased versus IDH1 wild type cells. Additionally, we observed an increased abundance of oxidative phosphorylation proteins which is consistent with reports that cells with IDH1 mutation are more reliant on oxidative phosphorylation for energy production.[282]

On the other hand, analysis of lower abundance proteins in IDH1 mutant cells found little enrichment of GOBP terms relating to metabolism. Indeed, many of the enriched terms were indicative of the less aggressive character of gliomas with IDH1 mutation versus wild type. (Figure 5.4 a,b) For example, cell morphogenesis and cell motility, which are associated with poorer progression-free survival [283] and tumor invasiveness [284] respectively, were both enriched in the downregulated group. Proteins involved in response to hypoxia, blood vessel development, and angiogenesis were also lower in abundance in IDH1 mutant cells, consistent with reports that R-2HG suppresses hypoxic response by inhibiting HIF1-α signaling and reduced transcription of the proangiogenic factor VEGF.[273, 285] Finally, we found a lower abundance of proteins related cell signaling, particularly in cell surface receptor mediated signaling and signal transduction. While IDH1 mutation can cause signaling dysregulation, including mTOR signaling, [286] defects in paracrine or autocrine signaling have not been reported.

To further investigate IDH1-mediated signaling pathway dysregulation, we used IPA to analyze pathway activation and suppression. (Figure 5.5) This analysis confirmed suppression of the mTOR signaling pathway as expected based on GOBP enrichment. Additionally, IPA analysis further emphasized the role of metabolic dysfunction in IDH1 mutated gliomas. Activation of oxidative phosphorylation pathways as suggested by STRING was validated, while sirtuin signaling was suppressed and cholesterol biosynthesis was activated through multiple pathways. Sirtuins are a family of deacylases and ADP-ribosyltransferases that catalyze myriad nicotinamide adenine dinucleotide (NAD+)-dependent reactions with diverse impacts on cellular metabolism, including upregulation of glycolysis, glutamine anaplerosis, and TCA cycle reversal.[287] We also found that IDH1 mutation led to reduced cell cycle regulation at both the G1/S checkpoint and G2/M DNA damage checkpoint, as well as suppression of apoptosis signaling. Surprisingly, although mTOR signaling was suppressed, the upstream activator PI3K/AKT signaling pathway was activated in conjunction with suppression of PTEN signaling. The PI3K/AKT pathway promotes cell growth and proliferation and PTEN is an important regulator of the pathway.[288] The disconnect between the upstream PI3K/AKT and downstream mTOR pathways suggest a disrupting influence which merits further investigation. Interestingly, while most of our results indicate the relative indolence of IDH1 mutant glioma cells, we found activation of ephrin receptor signaling which is characteristic of aggressive invasion in GBM.[289] Ephrin receptor signaling acts through crosstalk with Akt, [290] however the typical aggressive phenotype associated with ephrin receptor signaling activation may be eroded by the disruption between the PI3K/AKT and mTOR pathways.

DISCUSSION

Gliomas harboring IDH1 mutations were first reported in 2008.[261] Despite a decade of investigation, the specific mechanisms by which this early event in oncogenesis promotes malignant transformation is still unclear. IDH1 mutation leads to alterations in cellular

metabolism, redox potential, gene expression via epigenetic effects, DNA damage repair, and other potentially important routes.[262, 291] One or more of these factors could influence oncogenic transformation, however there may also be other processes involved which are not yet known. Further, therapies targeting the initiating step of gliomagenesis by directly inhibiting mutant IDH1 activity may be insufficient considering the subsequent cellular alterations driving transformation and disease progression. The goal of our comparative proteomic study was therefore to identify previously unknown biological processes and pathways which may elucidate IDH1 mutant glioma biology and potential therapeutic targets.

From a technical view, the identification of over 10000 proteins using a strategy of concatenated combination of high pH chromatography fractions is a notable achievement. Maximizing the number of proteins identified and quantified is important to illuminating unknown systems perturbed by IDH1. Many of our results confirmed previous reports on IDH1 mutated glioma, and served as validation of our sample preparation and handling. Importantly, we found significant overrepresentation of metabolic proteins and neuronal growth in the upregulated proteome. Enrichment of proteins in angiogenic and cell invasion processes in the downregulated proteome was likewise found as expected.

Beyond the expected effects of IDH1 mutation, we report several new biological processes and pathways perturbed in IDH1 mutant cells. First, we observed a decrease in proteins related to protein secretion and vesicle mediated transport. Although the secretome of GBM has been extensively studied, [292] the secretome of IDH1 mutant gliomas has been overlooked. In GBM and many other cancers, cell secretion is upregulated and plays an important role in pathogenesis.[292] Reduced secretion by IDH1 gliomas may contribute to slower growth relative to their wild type counterparts. This finding additionally suggests that inhibition of secretory pathways may provide a novel therapeutic approach in the treatment of aggressive GBM.

We also found lower abundance of proteins related to signal transduction and cell surface receptor mediated signaling in IDH1 mutant glioma cells than in wild type. This observation is probably due to increased activation of these pathways in GBM rather than defects due to IDH1 mutation. For example, the receptor EGFR is frequently overexpressed in GBM, leading to increased cell growth and proliferation.[293] Numerous therapies have been developed to target EGFR, however none have successfully altered disease progression, potentially due to crosstalk between pathways.[294] Paracrine and autocrine signaling in GBM are active areas of investigation, but targeted comparative analysis of extracellular signaling in wild type and IDH1 mutant cells is needed due to the clear differences in activation of these pathways between groups.

Insulin receptor signaling was one signaling pathway we found more activated in IDH1 mutant glioma than in wild type cells. Insulin receptor signaling has been implicated as a contributory factor in a number of cancers including breast, prostate, thyroid, and glioblastoma, [295-298] but not in lower grade gliomas. Our results indicate that insulin receptor signaling is more active in IDH1 mutated glioma than wild type. Because insulin receptor signaling promotes cell growth and proliferation, [299] activation of this pathway could be important to glioma progression. It is important to note the possibility that insulin signaling is not truly activated as indicated by IPA analysis of our data, but is an artifact of the pervasive metabolic dysfunction in IDH1 mutant glioma. If confirmed, however, this pathway is a putative target for therapeutic intervention. Finally, we found mild activation of 14-3-3 signaling in the IDH1 mutant cells. Because 14-3-3 signaling can be either tumor suppressive or tumor promotive, additional research is required to determine the specific role of this pathway and any potential for therapeutic exploitation in glioma.

A major limiting factor of this work is that only one biological replicate from the mutant and wild type groups was analyzed. The biological processes and pathways that we identify here may be specific to the individual tumors from which the cells were derived rather than

being more broadly applicable to gliomas in general. The lack of sufficient biological replicates is a likely reason for the unusually large number of proteins showing statistically significant abundance change. Analysis of additional replicates will reduce the number of proteins with significant abundance alterations and clarify which observations reported here are cell line specific.

CONCLUSIONS

In sum, we have used comparative proteomic profiling to identify new processes and pathways which are disrupted in IDH1 mutant glioma cells. The quality of our data was validated by the confirmation of several well-known effects of IDH1 mutation. Although these findings must be clarified through analysis of additional biological replicates, we have identified several new areas for potential future investigation.

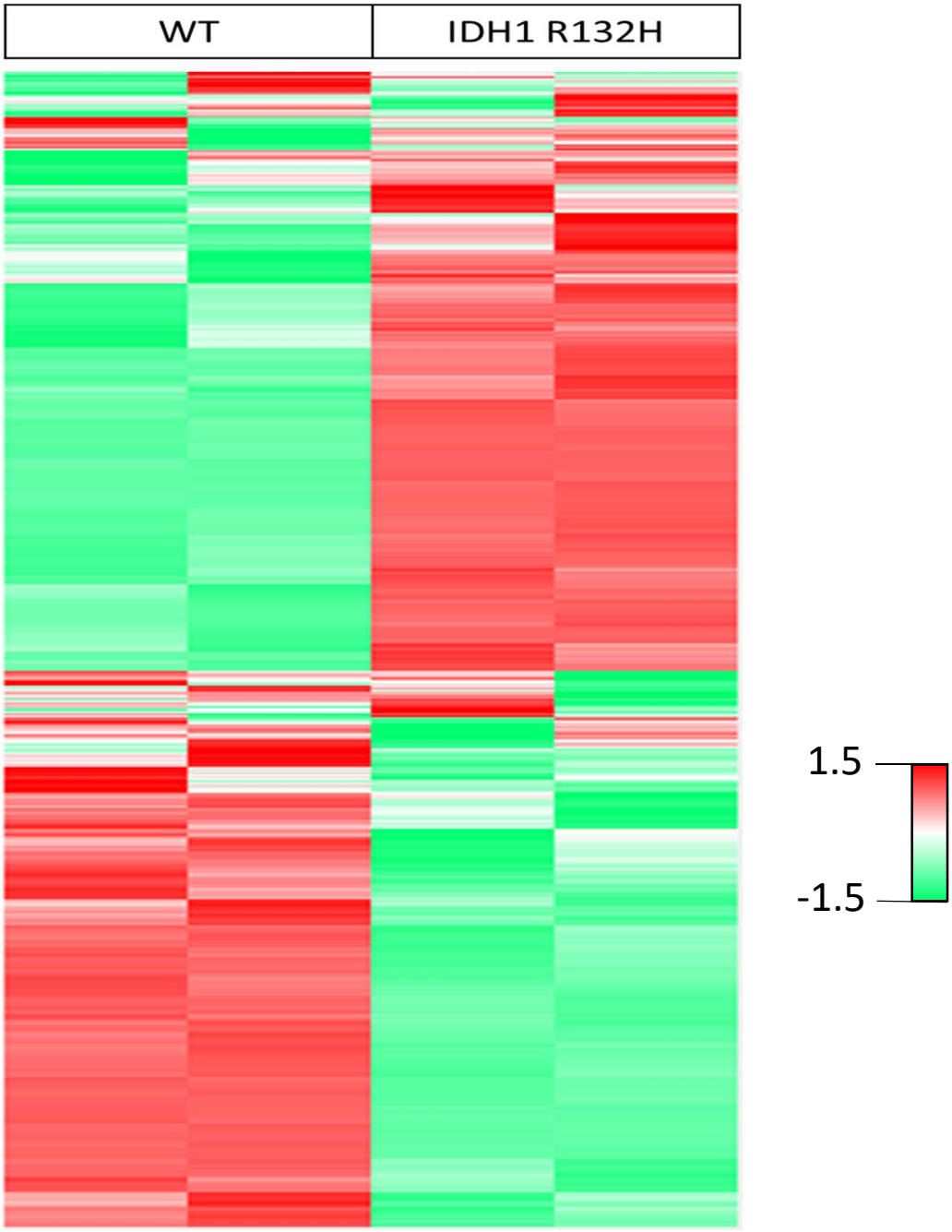


Figure 5.1 - A heatmap of unsupervised hierarchical clustering analysis of z-scored protein abundance. Each row indicates a protein and each column indicates a replicate measurement. Two technical replicates were performed for one biological replicate of IDH1 mutant or wild type glioma cells. Red indicates higher abundance, green indicates lower abundance, and white indicates mean abundance.

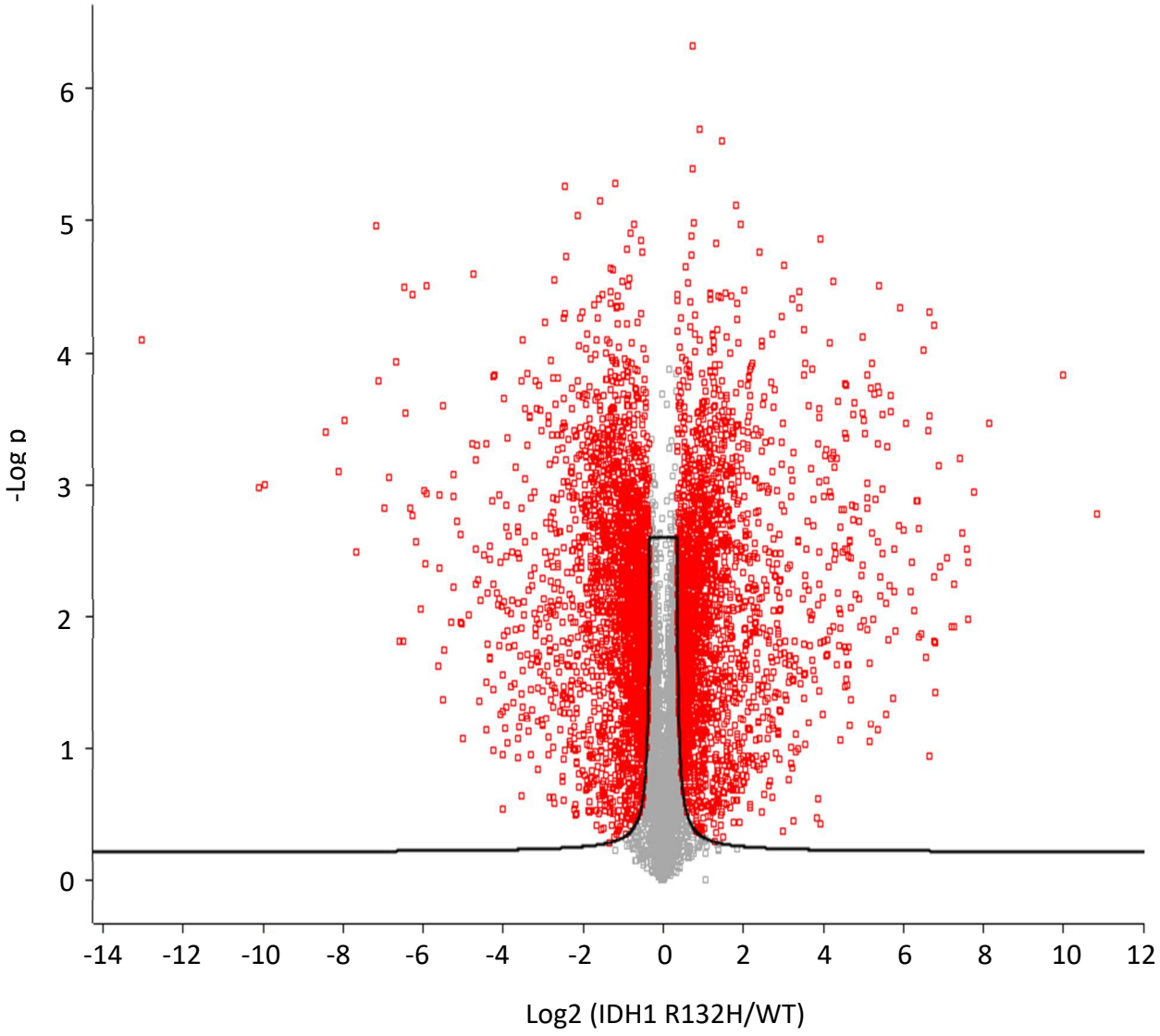
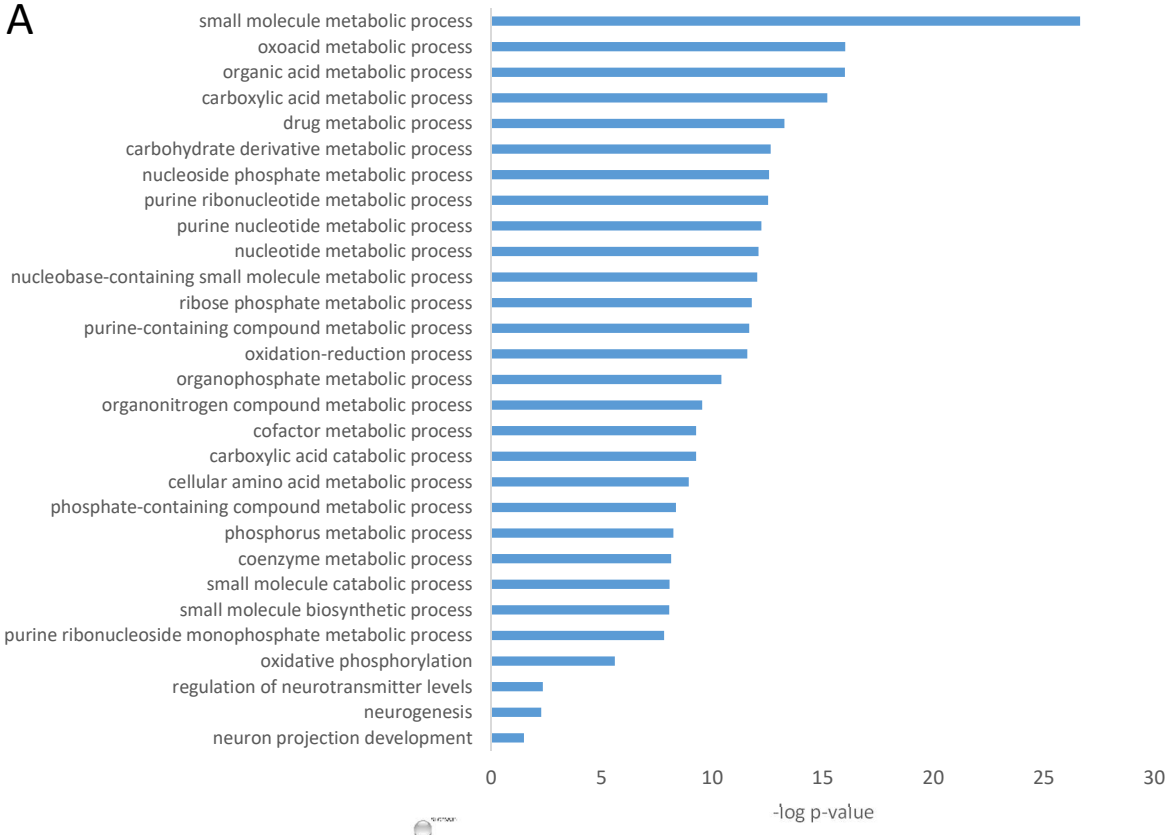


Figure 5.2 – Volcano plot of proteins identified in IDH1 mutant and wild type glioma cells. Positive values indicate higher expression in IDH1 mutant cells and negative values indicate higher expression in wild type cells. Proteins with statistically significant altered expression and fold change >1.5 are indicated in red.

A


small molecule metabolic process
oxoacid metabolic process
organic acid metabolic process
carboxylic acid metabolic process
drug metabolic process
carbohydrate derivative metabolic process
nucleoside phosphate metabolic process
purine ribonucleotide metabolic process
purine nucleotide metabolic process
nucleotide metabolic process
nucleobase-containing small molecule metabolic process
ribose phosphate metabolic process
purine-containing compound metabolic process
oxidation-reduction process
organophosphate metabolic process
organonitrogen compound metabolic process
cofactor metabolic process
carboxylic acid catabolic process
cellular amino acid metabolic process
phosphate-containing compound metabolic process
phosphorus metabolic process
coenzyme metabolic process
small molecule catabolic process
small molecule biosynthetic process
purine ribonucleoside monophosphate metabolic process
oxidative phosphorylation
regulation of neurotransmitter levels
neurogenesis
neuron projection development
0
5
10
15
20
25
30
-log p-value

B

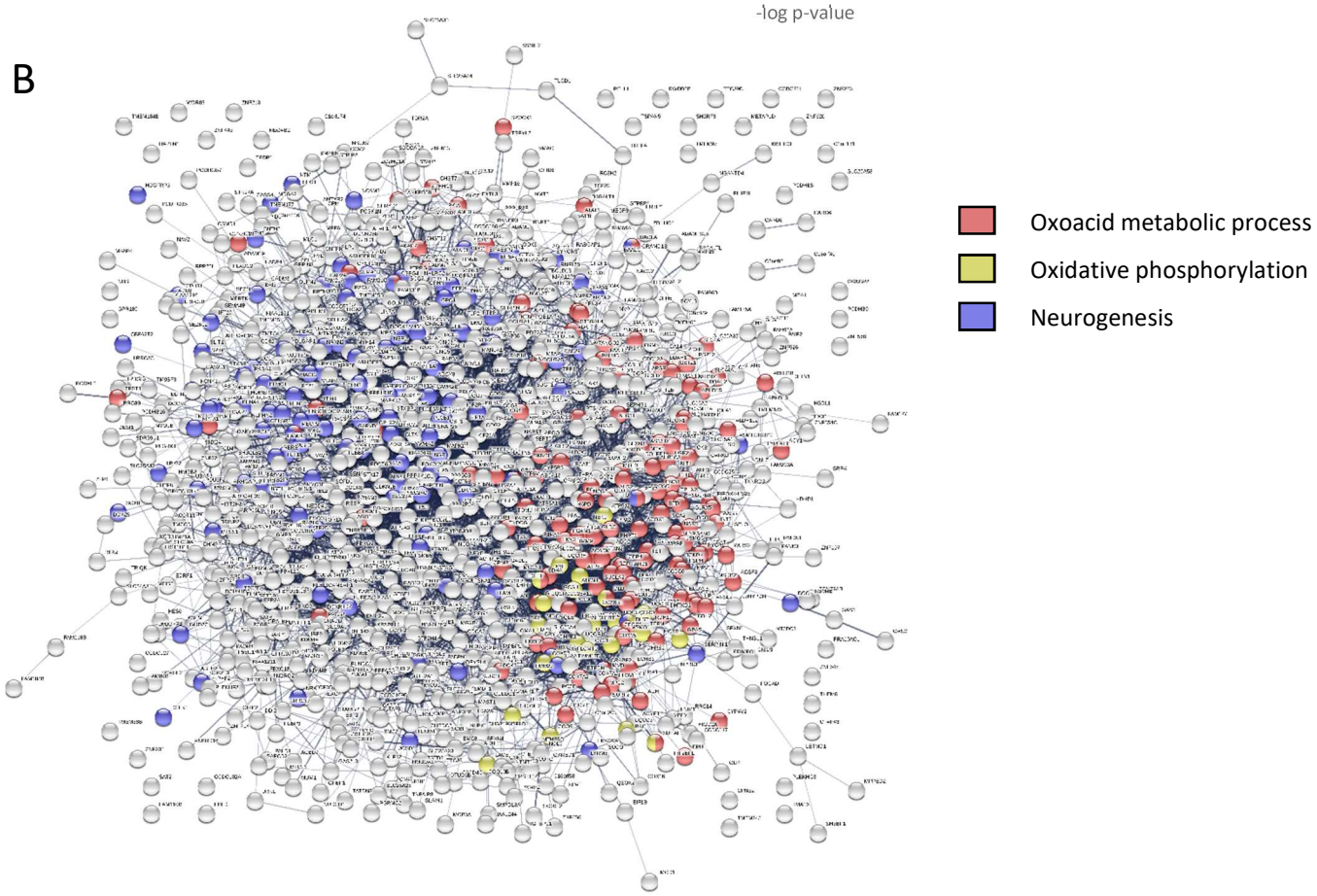
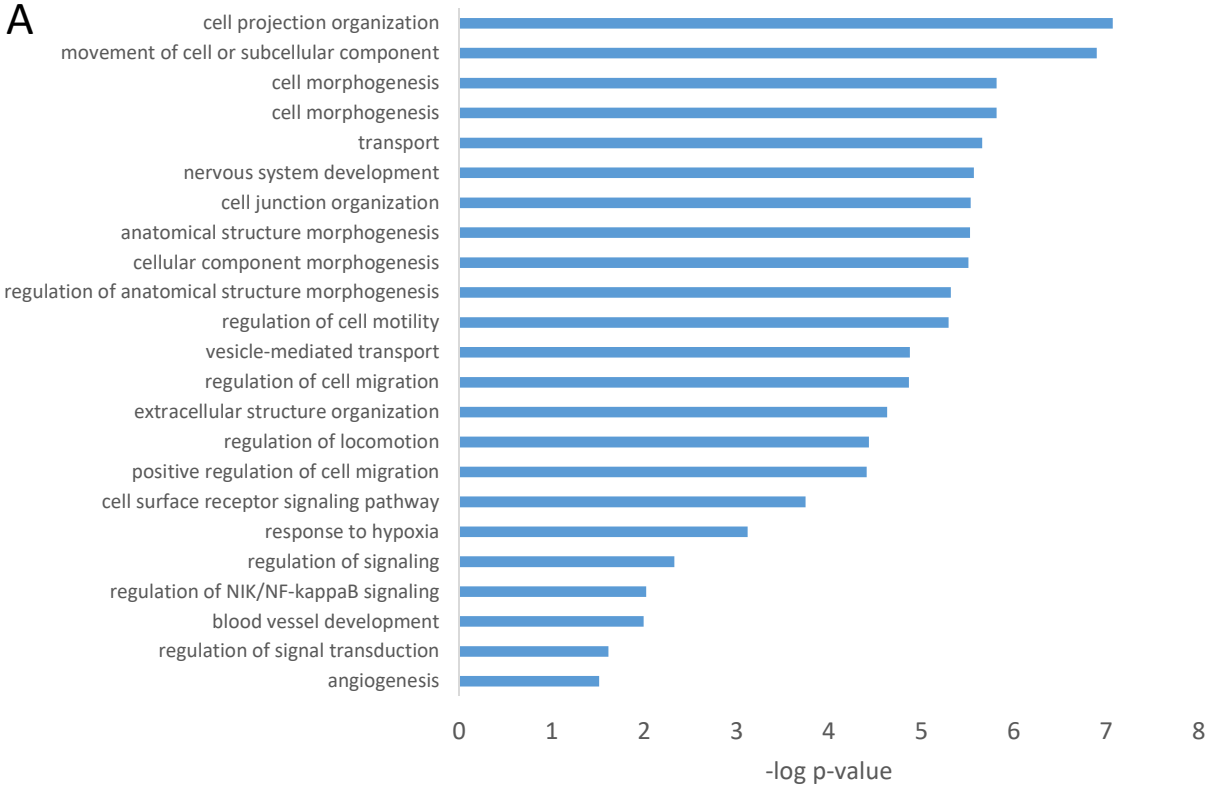

Oxoacid metabolic process
Oxidative phosphorylation
Neurogenesis

Figure 5.3 – GOBP and STRING analysis of upregulated proteins in IDH1 mutant glioma cells. A) Biological processes over-represented in proteins with higher abundance in IDH1 mutant glioma cells. B) Interaction network of proteins with higher abundance in IDH1 mutant glioma cells. Select biological processes have been highlighted.

A


cell projection organization
movement of cell or subcellular component
cell morphogenesis
cell morphogenesis
transport
nervous system development
cell junction organization
anatomical structure morphogenesis
cellular component morphogenesis
regulation of anatomical structure morphogenesis
regulation of cell motility
vesicle-mediated transport
regulation of cell migration
extracellular structure organization
regulation of locomotion
positive regulation of cell migration
cell surface receptor signaling pathway
response to hypoxia
regulation of signaling
regulation of NIK/NF-kappaB signaling
blood vessel development
regulation of signal transduction
angiogenesis
0
1
2
3
4
5
6
7
8
-log p-value

B

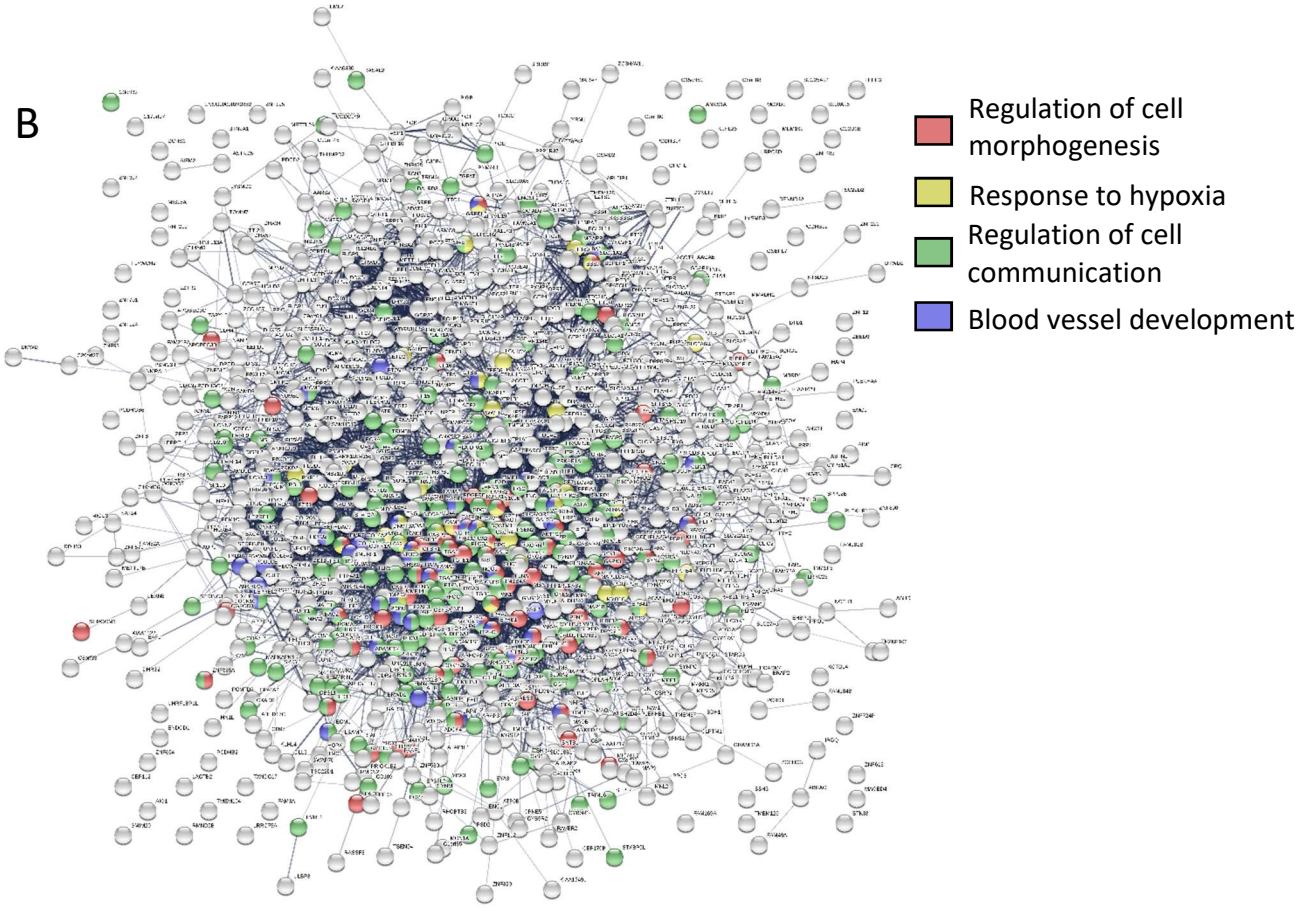

Regulation of cell morphogenesis
Response to hypoxia
Regulation of cell communication
Blood vessel development

Figure 5.4 – GOBP and STRING analysis of downregulated proteins in IDH1 mutant glioma cells. A) Biological processes over-represented in proteins with lower abundance in IDH1 mutant glioma cells. B) Interaction network of proteins with lower abundance in IDH1 mutant glioma cells. Select biological processes have been highlighted.

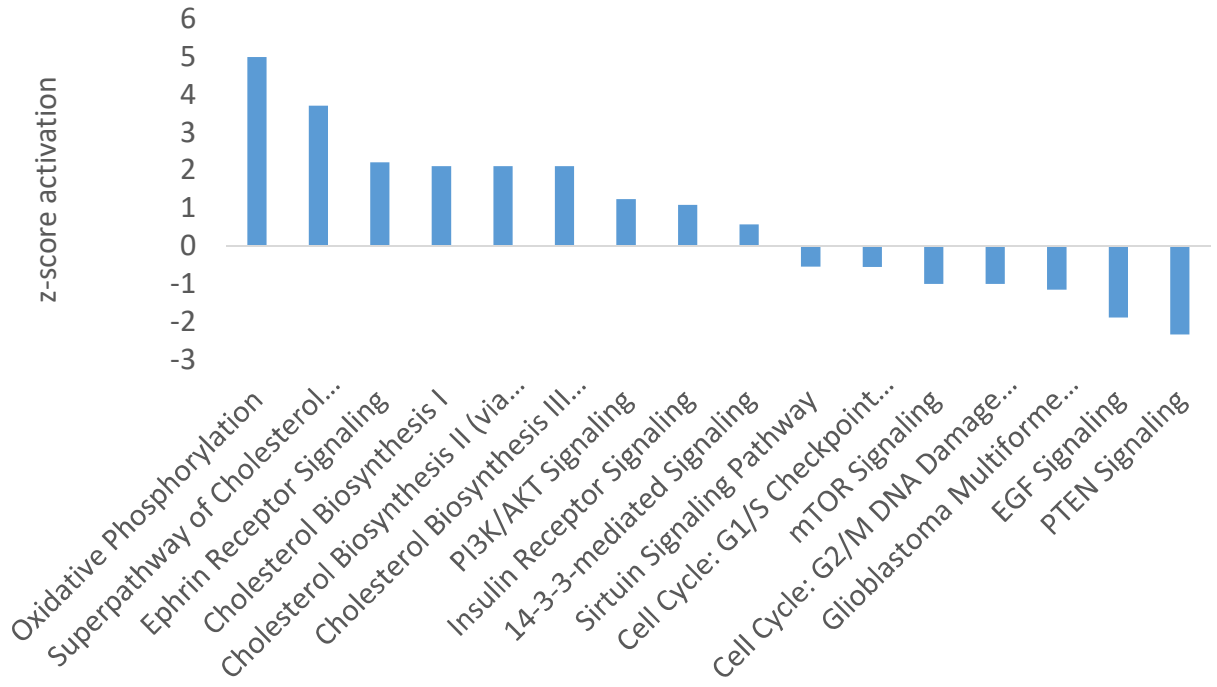


Figure 5.5 – IPA pathway activation analysis. Positive z-scores indicate pathway activation; negative z-scores indicate pathway suppression.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 2018, 68(6):394-424.

2. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A: Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: a cancer journal for clinicians* 2017, 67(6):439-448.

3. Cronin KA, Lake AJ, Scott S, Sherman RL, Noone AM, Howlader N, Henley SJ, Anderson RN, Firth AU, Ma J *et al*: Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics. *Cancer* 2018, 124(13):2785-2800.

4. Bleyer A, Baines C, Miller AB: Impact of screening mammography on breast cancer mortality. *International journal of cancer* 2016, 138(8):2003-2012.

5. Welch HG, Prorok PC, O'Malley AJ, Kramer BS: Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *The New England journal of medicine* 2016, 375(15):1438-1447.

6. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A: Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians* 2016, 66(1):31-42.

7. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2016. *CA: a cancer journal for clinicians* 2016, 66(1):7-30.

8. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M: Genetic susceptibility to breast cancer. *Molecular oncology* 2010, 4(3):174-191.

9. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M *et al*: Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute* 2015, 107(5).

10. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA: Family history and the risk of breast cancer: a systematic review and meta-analysis. *International journal of cancer* 1997, 71(5):800-809.

11. Collaborative Group on Hormonal Factors in Breast Cancer: Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 2001, 358(9291):1389-1399.

12. Sud A, Kinnersley B, Houlston RS: Genome-wide association studies of cancer: current insights and future perspectives. *Nature reviews Cancer* 2017, 17(11):692-704.

13. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, Jervis S, van Leeuwen FE, Milne RL, Andrieu N *et al*: Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *Jama* 2017, 317(23):2402-2416.

14. Rudolph A, Chang-Claude J, Schmidt MK: Gene-environment interaction and risk of breast cancer. *British journal of cancer* 2016, 114(2):125-133.

15. Reis-Filho JS, Lakhani SR: Breast cancer special types: why bother? *The Journal of pathology* 2008, 216(4):394-398.

16. Tang P, Wang J, Bourne P: Molecular classifications of breast carcinoma with similar terminology and different definitions: are they the same? *Human pathology* 2008, 39(4):506-513.

17. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, Shi B: Breast cancer intrinsic subtype classification, clinical use and future trends. *American journal of cancer research* 2015, 5(10):2929-2943.

18. Polyak K: Heterogeneity in breast cancer. *The Journal of clinical investigation* 2011, 121(10):3786-3788.

19. Turashvili G, Brogi E: Tumor Heterogeneity in Breast Cancer. *Frontiers in medicine* 2017, 4:227.

20. Zardavas D, Irrthum A, Swanton C, Piccart M: Clinical management of breast cancer heterogeneity. *Nature reviews Clinical oncology* 2015, 12(7):381-394.

21. Elston CW, Ellis IO: Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991, 19(5):403-410.

22. Bloom HJ, Richardson WW: Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *British journal of cancer* 1957, 11(3):359-377.

23. Schottenfeld D, Nash AG, Robbins GF, Beattie EJ, Jr.: Ten-year results of the treatment of primary operable breast carcinoma: A summary of 304 patients evaluated by the TNM system. *Cancer* 1976, 38(2):1001-1007.

24. Giuliano AE, Edge SB, Hortobagyi GN: Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. *Annals of surgical oncology* 2018, 25(7):1783-1785.

25. Weigelt B, Reis-Filho JS: Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nature reviews Clinical oncology* 2009, 6(12):718-730.

26. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, Fox SB, Ichihara S, Jacquemier J, Lakhani SR *et al*: Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast cancer research : BCR* 2010, 12(4):207.

27. Rakha EA, El-Sayed ME, Menon S, Green AR, Lee AH, Ellis IO: Histologic grading is an independent prognostic factor in invasive lobular carcinoma of the breast. *Breast cancer research and treatment* 2008, 111(1):121-127.

28. Ellis IO, Galea M, Broughton N, Locker A, Blamey RW, Elston CW: Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology* 1992, 20(6):479-489.

29. Weigelt B, Geyer FC, Reis-Filho JS: Histological types of breast cancer: how special are they? *Molecular oncology* 2010, 4(3):192-208.

30. Li CI, Uribe DJ, Daling JR: Clinical characteristics of different histologic types of breast cancer. *British journal of cancer* 2005, 93(9):1046-1052.

31. Beatson GT: On the Treatment of Inoperable Cases of Carcinoma of the Mamma: Suggestions for a New Method of Treatment, with Illustrative Cases. *Transactions Medico-Chirurgical Society of Edinburgh* 1896, 15:153-179.

32. Folca PJ, Glascock RF, Irvine WT: Studies with tritium-labelled hexoestrol in advanced breast cancer. Comparison of tissue accumulation of hexoestrol with response to bilateral adrenalectomy and oophorectomy. *Lancet* 1961, 2(7206):796-798.

33. Glascock RF, Hoekstra WG: Selective accumulation of tritium-labelled hexoestrol by the reproductive organs of immature female goats and sheep. *The Biochemical journal* 1959, 72:673-682.

34. Gorski J, Toft D, Shyamala G, Smith D, Notides A: Hormone receptors: studies on the interaction of estrogen with the uterus. *Recent progress in hormone research* 1968, 24:45-80.

35. Toft D, Gorski J: A receptor molecule for estrogens: isolation from the rat uterus and preliminary characterization. *Proceedings of the National Academy of Sciences of the United States of America* 1966, 55(6):1574-1581.

36. Toft D, Shyamala G, Gorski J: A receptor molecule for estrogens: studies using a cell-free system. *Proceedings of the National Academy of Sciences of the United States of America* 1967, 57(6):1740-1743.

37. McGuire WL: Current status of estrogen receptors in human breast cancer. *Cancer* 1975, 36(2):638-644.

38. McGuire WL: Estrogen receptors in human breast cancer. *The Journal of clinical investigation* 1973, 52(1):73-77.

39. Knight WA, Livingston RB, Gregory EJ, McGuire WL: Estrogen receptor as an independent prognostic factor for early recurrence in breast cancer. *Cancer research* 1977, 37(12):4669-4671.

40. Osborne CK, Yochmowitz MG, Knight WA, 3rd, McGuire WL: The value of estrogen and progesterone receptors in the treatment of breast cancer. *Cancer* 1980, 46(12 Suppl):2884-2888.

41. Horwitz KB, McGuire WL: Predicting response to endocrine therapy in human breast cancer: a hypothesis. *Science* 1975, 189(4204):726-727.

42. Bardou VJ, Arpino G, Elledge RM, Osborne CK, Clark GM: Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2003, 21(10):1973-1979.

43. Arpino G, Weiss H, Lee AV, Schiff R, De Placido S, Osborne CK, Elledge RM: Estrogen receptor-positive, progesterone receptor-negative breast cancer: association with growth factor receptor expression and tamoxifen resistance. *Journal of the National Cancer Institute* 2005, 97(17):1254-1261.

44. Hayes DF: Markers of endocrine sensitivity. *Breast cancer research : BCR* 2008, 10 Suppl 4:S18.

45. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M *et al*: American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Archives of pathology & laboratory medicine* 2010, 134(7):e48-72.

46. Schechter AL, Stern DF, Vaidyanathan L, Decker SJ, Drebin JA, Greene MI, Weinberg RA: The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. *Nature* 1984, 312(5994):513-516.

47. King CR, Kraus MH, Aaronson SA: Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science* 1985, 229(4717):974-976.

48. Coussens L, Yang-Feng TL, Liao YC, Chen E, Gray A, McGrath J, Seeburg PH, Libermann TA, Schlessinger J, Francke U *et al*: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. *Science* 1985, 230(4730):1132-1139.

49. Akiyama T, Sudo C, Ogawara H, Toyoshima K, Yamamoto T: The product of the human c-erbB-2 gene: a 185-kilodalton glycoprotein with tyrosine kinase activity. *Science* 1986, 232(4758):1644-1646.

50. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL: Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987, 235(4785):177-182.

51. Revillion F, Bonneterre J, Peyrat JP: ERBB2 oncogene in human breast cancer and its clinical significance. *Eur J Cancer* 1998, 34(6):791-808.

52. Press MF, Bernstein L, Thomas PA, Meisner LF, Zhou JY, Ma Y, Hung G, Robinson RA, Harris C, El-Naggar A *et al*: HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1997, 15(8):2894-2904.

53. Onitilo AA, Engel JM, Greenlee RT, Mukesh BN: Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clinical medicine & research* 2009, 7(1-2):4-13.

54. Loibl S, Gianni L: HER2-positive breast cancer. *Lancet* 2017, 389(10087):2415-2429.

55. Howlader N, Altekruse SF, Li CI, Chen VW, Clarke CA, Ries LA, Cronin KA: US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *Journal of the National Cancer Institute* 2014, 106(5).

56. Burstein HJ: The distinctive nature of HER2-positive breast cancers. *The New England journal of medicine* 2005, 353(16):1652-1654.

57. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M *et al*: Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *The New England journal of medicine* 2001, 344(11):783-792.

58. Massarweh S, Schiff R: Unraveling the mechanisms of endocrine resistance in breast cancer: new therapeutic opportunities. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2007, 13(7):1950-1954.

59. Tao Z, Shi A, Lu C, Song T, Zhang Z, Zhao J: Breast Cancer: Epidemiology and Etiology. *Cell biochemistry and biophysics* 2015, 72(2):333-338.

60. Collignon J, Lousberg L, Schroeder H, Jerusalem G: Triple-negative breast cancer: treatment challenges and solutions. *Breast Cancer (Dove Med Press)* 2016, 8:93-107.

61. Foulkes WD, Smith IE, Reis-Filho JS: Triple-negative breast cancer. *The New England journal of medicine* 2010, 363(20):1938-1948.

62. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L: Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nature reviews Clinical oncology* 2016, 13(11):674-690.

63. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: Molecular portraits of human breast tumours. *Nature* 2000, 406(6797):747-752.

64. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al*: Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2009, 27(8):1160-1167.

65. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* 2001, 98(19):10869-10874.

66. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L *et al*: Immunohistochemical and clinical characterization

of the basal-like subtype of invasive breast carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2004, 10(16):5367-5374.

67. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT *et al*: Epidemiology of basal-like breast cancer. *Breast cancer research and treatment* 2008, 109(1):123-139.

68. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, Perou CM, Nielsen TO: Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008, 14(5):1368-1376.

69. Nguyen PL, Taghian AG, Katz MS, Niemierko A, Abi Raad RF, Boon WL, Bellon JR, Wong JS, Smith BL, Harris JR: Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008, 26(14):2373-2378.

70. Prat A, Perou CM: Deconstructing the molecular portraits of breast cancer. *Molecular oncology* 2011, 5(1):5-23.

71. Musgrove EA, Sutherland RL: Biological determinants of endocrine resistance in breast cancer. *Nature reviews Cancer* 2009, 9(9):631-643.

72. Biomarkers Definitions Working Group: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology and therapeutics* 2001, 69(3):89-95.

73. Mabert K, Cojoc M, Peitzsch C, Kurth I, Souchelnytskyi S, Dubrovska A: Cancer biomarker discovery: current status and future perspectives. *International journal of radiation biology* 2014, 90(8):659-677.

74. Heneghan HM, Miller N, Lowery AJ, Sweeney KJ, Newell J, Kerin MJ: Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. *Annals of surgery* 2010, 251(3):499-505.

75. Zugazagoitia J, Guedes C, Ponce S, Ferrer I, Molina-Pinelo S, Paz-Ares L: Current Challenges in Cancer Treatment. *Clinical therapeutics* 2016, 38(7):1551-1566.

76. Bennett MR, Devarajan P: Chapter One - Characteristics of an Ideal Biomarker of Kidney Diseases. In: *Biomarkers of Kidney Disease (Second Edition).* edn. Edited by Edelstein CL: Academic Press; 2017: 1-20.

77. Pavlou MP, Diamandis EP: The cancer cell secretome: a good source for discovering biomarkers? *J Proteomics* 2010, 73(10):1896-1906.

78. Chen IH, Xue L, Hsu CC, Paez JS, Pan L, Andaluz H, Wendt MK, Iliuk AB, Zhu JK, Tao WA: Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. *Proceedings of the National Academy of Sciences of the United States of America* 2017, 114(12):3175-3180.

79. Schwarzenbach H, Nishida N, Calin GA, Pantel K: Clinical relevance of circulating cell-free microRNAs in cancer. *Nature reviews Clinical oncology* 2014, 11(3):145-156.

80. Heitzer E, Ulz P, Geigl JB: Circulating tumor DNA as a liquid biopsy for cancer. *Clinical chemistry* 2015, 61(1):112-123.

81. Roos L, van Dongen J, Bell CG, Burri A, Deloukas P, Boomsma DI, Spector TD, Bell JT: Integrative DNA methylome analysis of pan-cancer biomarkers in cancer discordant monozygotic twin-pairs. *Clinical epigenetics* 2016, 8:7.

82. Makawita S, Diamandis EP: The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry-based approaches: current strategies for candidate verification. *Clinical chemistry* 2010, 56(2):212-222.

83. Kulasingam V, Diamandis EP: Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. *Molecular & cellular proteomics : MCP* 2007, 6(11):1997-2011.

84. Kulasingam V, Diamandis EP: Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature clinical practice Oncology* 2008, 5(10):588-599.

85. Xue H, Lu B, Lai M: The cancer secretome: a reservoir of biomarkers. *Journal of translational medicine* 2008, 6:52.

86. Jones VS, Huang RY, Chen LP, Chen ZS, Fu L, Huang RP: Cytokines in cancer drug resistance: Cues to new therapeutic strategies. *Biochimica et biophysica acta* 2016, 1865(2):255-265.

87. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M: Plasma Proteome Profiling to Assess Human Health and Disease. *Cell systems* 2016, 2(3):185-195.

88. Anderson NL, Anderson NG: The human plasma proteome: history, character, and diagnostic prospects. *Molecular & cellular proteomics : MCP* 2002, 1(11):845-867.

89. Hortin GL, Sviridov D, Anderson NL: High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clinical chemistry* 2008, 54(10):1608-1616.

90. Travis J, Bowen J, Tewksbury D, Johnson D, Pannell R: Isolation of albumin from whole human plasma and fractionation of albumin-depleted plasma. *The Biochemical journal* 1976, 157(2):301-306.

91. Echan LA, Tang HY, Ali-Khan N, Lee K, Speicher DW: Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* 2005, 5(13):3292-3303.

92. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJ, Liebler DC: Depletion of abundant plasma proteins and limitations of plasma proteomics. *Journal of proteome research* 2010, 9(10):4982-4991.

93. Qian WJ, Kaleta DT, Petritis BO, Jiang H, Liu T, Zhang X, Mottaz HM, Varnum SM, Camp DG, 2nd, Huang L *et al*: Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. *Molecular & cellular proteomics : MCP* 2008, 7(10):1963-1973.

94. Yocum AK, Yu K, Oe T, Blair IA: Effect of immunoaffinity depletion of human serum during proteomic investigations. *Journal of proteome research* 2005, 4(5):1722-1731.

95. Gromov P, Gromova I, Olsen CJ, Timmermans-Wielenga V, Talman ML, Serizawa RR, Moreira JM: Tumor interstitial fluid - a treasure trove of cancer biomarkers. *Biochimica et biophysica acta* 2013, 1834(11):2259-2270.

96. Alexander H, Stegner AL, Wagner-Mann C, Du Bois GC, Alexander S, Sauter ER: Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2004, 10(22):7500-7510.

97. Yu CJ, Wang CL, Wang CI, Chen CD, Dan YM, Wu CC, Wu YC, Lee IN, Tsai YH, Chang YS *et al*: Comprehensive proteome analysis of malignant pleural effusion for lung cancer biomarker discovery by using multidimensional protein identification technology. *Journal of proteome research* 2011, 10(10):4671-4682.

98. Chen YL, Cheng WF, Chang MC, Lin HW, Huang CT, Chien CL, Chen CA: Interferon-gamma in ascites could be a predictive biomarker of outcome in ovarian carcinoma. *Gynecologic oncology* 2013, 131(1):63-68.

99. Mbeunkui F, Metge BJ, Shevde LA, Pannell LK: Identification of differentially secreted biomarkers using LC-MS/MS in isogenic cell lines representing a progression of breast cancer. *Journal of proteome research* 2007, 6(8):2993-3002.

100. Ahn Y, Kang UB, Kim J, Lee C: Mining of serum glycoproteins by an indirect approach using cell line secretome. *Molecules and cells* 2010, 29(2):123-130.

101. Hanash SM, Pitteri SJ, Faca VM: Mining the plasma proteome for cancer biomarkers. *Nature* 2008, 452(7187):571-579.

102. Haslene-Hox H, Tenstad O, Wiig H: Interstitial fluid-a reflection of the tumor cell microenvironment and secretome. *Biochimica et biophysica acta* 2013, 1834(11):2336-2346.

103. Wu CC, Hsu CW, Chen CD, Yu CJ, Chang KP, Tai DI, Liu HP, Su WH, Chang YS, Yu JS: Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Molecular & cellular proteomics : MCP* 2010, 9(6):1100-1117.

104. Polisetty RV, Gupta MK, Nair SC, Ramamoorthy K, Tiwary S, Shiras A, Chandak GR, Sirdeshmukh R: Glioblastoma cell secretome: analysis of three glioblastoma cell lines reveal 148 non-redundant proteins. *Journal of proteomics* 2011, 74(10):1918-1925.

105. Schiarea S, Solinas G, Allavena P, Scigliuolo GM, Bagnati R, Fanelli R, Chiabrando C: Secretome analysis of multiple pancreatic cancer cell lines reveals perturbations of key functional networks. *Journal of proteome research* 2010, 9(9):4376-4392.

106. Bosse K, Haneder S, Arlt C, Ihling CH, Seufferlein T, Sinz A: Mass spectrometry-based secretome analysis of non-small cell lung cancer cell lines. *Proteomics* 2016, 16(21):2801-2814.

107. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG: Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & cellular proteomics : MCP* 2005, 4(10):1487-1502.

108. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B: Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* 2007, 389(4):1017-1031.

109. Rappsilber J, Ryder U, Lamond AI, Mann M: Large-scale proteomic analysis of the human spliceosome. *Genome research* 2002, 12(8):1231-1245.

110. Sanders SL, Jennings J, Canutescu A, Link AJ, Weil PA: Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. *Molecular and cellular biology* 2002, 22(13):4723-4738.

111. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M: Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & cellular proteomics : MCP* 2005, 4(9):1265-1272.

112. Dowle AA, Wilson J, Thomas JR: Comparing the Diagnostic Classification Accuracy of iTRAQ, Peak-Area, Spectral-Counting, and emPAI Methods for Relative Quantification in Expression Proteomics. *Journal of proteome research* 2016, 15(10):3550-3562.

113. Liu H, Sadygov RG, Yates JR, 3rd: A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical chemistry* 2004, 76(14):4193-4201.

114. Zhang Y, Wen Z, Washburn MP, Florens L: Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Analytical chemistry* 2009, 81(15):6317-6326.

115. Lu P, Vogel C, Wang R, Yao X, Marcotte EM: Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature biotechnology* 2007, 25(1):117-124.

116. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP: Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *Journal of proteome research* 2006, 5(9):2339-2347.

117. Powell DW, Weaver CM, Jennings JL, McAfee KJ, He Y, Weil PA, Link AJ: Cluster analysis of mass spectrometry data reveals a novel component of SAGA. *Molecular and cellular biology* 2004, 24(16):7249-7259.

118. Zhou JY, Schepmoes AA, Zhang X, Moore RJ, Monroe ME, Lee JH, Camp DG, Smith RD, Qian WJ: Improved LC-MS/MS spectral counting statistics by recovering low-scoring spectra matched to confidently identified peptide sequences. *Journal of proteome research* 2010, 9(11):5698-5704.

119. Zhang Y, Wen Z, Washburn MP, Florens L: Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Analytical chemistry* 2010, 82(6):2272-2281.

120. Chelius D, Bondarenko PV: Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of proteome research* 2002, 1(4):317-323.

121. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M: Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* 2014, 13(9):2513-2526.

122. Callister SJ, Barry RC, Adkins JN, Johnson ET, Qian WJ, Webb-Robertson BJ, Smith RD, Lipton MS: Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *Journal of proteome research* 2006, 5(2):277-286.

123. Nahnsen S, Bielow C, Reinert K, Kohlbacher O: Tools for label-free peptide quantification. *Molecular & cellular proteomics : MCP* 2013, 12(3):549-556.

124. Blein-Nicolas M, Zivy M: Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochimica et biophysica acta* 2016, 1864(8):883-895.

125. Bantscheff M, Lemeer S, Savitski MM, Kuster B: Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry* 2012, 404(4):939-965.

126. Ahrne E, Molzahn L, Glatter T, Schmidt A: Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* 2013, 13(17):2567-2578.

127. Trudgian DC, Ridlova G, Fischer R, Mackeen MM, Ternette N, Acuto O, Kessler BM, Thomas B: Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline. *Proteomics* 2011, 11(14):2790-2797.

128. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F *et al*: Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 2016, 534(7605):55-62.

129. Ruggles KV, Tang Z, Wang X, Grover H, Askenazi M, Teubl J, Cao S, McLellan MD, Clauser KR, Tabb DL *et al*: An Analysis of the Sensitivity of Proteogenomic Mapping of

Somatic Mutations and Novel Splicing Events in Cancer. *Molecular & cellular proteomics : MCP* 2016, 15(3):1060-1071.

130. United States Cancer Statistics: 1999–2014 Incidence and Mortality Web-based Report [https://nccd.cdc.gov/uscs/.]

131. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, Aas T, Alexandrov LB, Larsimont D, Davies H *et al*: Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nature medicine* 2015, 21(7):751-759.

132. Perou CM, Borresen-Dale AL: Systems biology and genomics of breast cancer. *Cold Spring Harbor perspectives in biology* 2011, 3(2).

133. Koren S, Bentires-Alj M: Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy. *Mol Cell* 2015, 60(4):537-546.

134. Chin L, Andersen JN, Futreal PA: Cancer genomics: from discovery science to personalized medicine. *Nature medicine* 2011, 17(3):297-303.

135. Simon R, Roychowdhury S: Implementing personalized cancer genomics in clinical trials. *Nature reviews Drug discovery* 2013, 12(5):358-369.

136. Whelan SA, He J, Lu M, Souda P, Saxton RE, Faull KF, Whitelegge JP, Chang HR: Mass spectrometry (LC-MS/MS) identified proteomic biosignatures of breast cancer in proximal fluid. *J Proteome Res* 2012, 11(10):5034-5045.

137. Kulasingam V, Diamandis EP: Tissue culture-based breast cancer biomarker discovery platform. *Int J Cancer* 2008, 123(9):2007-2012.

138. Tagliabracci VS, Wiley SE, Guo X, Kinch LN, Durrant E, Wen J, Xiao J, Cui J, Nguyen KB, Engel JL *et al*: A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell* 2015, 161(7):1619-1632.

139. Rappsilber J, Mann M, Ishihama Y: Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* 2007, 2(8):1896-1906.

140. Cox J, Mann M: Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry* 2011, 80:273-299.

141. Kramer A, Green J, Pollard J, Jr., Tugendreich S: Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2014, 30(4):523-530.

142. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* 2017, 45(D1):D362-D368.

143. Liu C, Yu Y, Liu F, Wei X, Wrobel JA, Gunawardena HP, Zhou L, Jin J, Chen X: A chromatin activity-based chemoproteomic approach reveals a transcriptional repressome for gene-specific silencing. *Nat Commun* 2014, 5:5733.

144. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E *et al*: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* 2012, 2(5):401-404.

145. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E *et al*: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013, 6(269):pl1.

146. cgdsr: R-Based API for Accessing the MSKCC Cancer Genomics Data Server (CGDS). R package version 1.2.5 [http://CRAN.R-project.org/package=cgdsr]

147. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C *et al*: Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 2015, 163(2):506-519.

148. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012, 486(7403):346-352.

149. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut SJ *et al*: The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* 2016, 7:11479.

150. Gu Z, Eils R, Schlesner M: Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics (Oxford, England)* 2016.

151. Therneau TM, Grambsch PM: Modeling survival data : extending the Cox model. New York: Springer; 2000.

152. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J: The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature methods* 2016, 13(9):731-740.

153. Liu C, Yu Y, Liu F, Wei X, Wrobel JA, Gunawardena HP, Zhou L, Jin J, Chen X: A chromatin activity-based chemoproteomic approach reveals a transcriptional repressome for gene-specific silencing. *Nature communications* 2014, 5:5733.

154. Erdogan O, Xie L, Wang L, Wu B, Kong Q, Wan Y, Chen X: Proteomic dissection of LPS-inducible, PHF8-dependent secretome reveals novel roles of PHF8 in TLR4-induced acute inflammation and T cell proliferation. *Scientific reports* 2016, 6:24833.

155. Meinken J, Walker G, Cooper CR, Min XJ: MetazSecKB: the human and animal secretome and subcellular proteome knowledgebase. *Database : the journal of biological databases and curation* 2015, 2015.

156. Boersema PJ, Geiger T, Wisniewski JR, Mann M: Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Molecular & cellular proteomics : MCP* 2013, 12(1):158-171.

157. Ma F, Li H, Wang H, Shi X, Fan Y, Ding X, Lin C, Zhan Q, Qian H, Xu B: Enriched CD44(+)/CD24(-) population drives the aggressive phenotypes presented in triple-negative breast cancer (TNBC). *Cancer letters* 2014, 353(2):153-159.

158. Chen HA, Chang YW, Tseng CF, Chiu CF, Hong CC, Wang W, Wang MY, Hsiao M, Ma JT, Chen CH *et al*: E1A-mediated inhibition of HSPA5 suppresses cell migration and invasion in triple-negative breast cancer. *Annals of surgical oncology* 2015, 22(3):889-898.

159. Cheng Q, Chang JT, Geradts J, Neckers LM, Haystead T, Spector NL, Lyerly HK: Amplification and high-level expression of heat shock protein 90 marks aggressive phenotypes of human epidermal growth factor receptor 2 negative breast cancer. *Breast cancer research : BCR* 2012, 14(2):R62.

160. Mahler-Araujo B, Savage K, Parry S, Reis-Filho JS: Reduction of E-cadherin expression is associated with non-lobular breast carcinomas of basal-like and triple negative phenotype. *Journal of clinical pathology* 2008, 61(5):615-620.

161. Hill JJ, Tremblay TL, Fauteux F, Li J, Wang E, Aguilar-Mahecha A, Basik M, O'Connor-McCourt M: Glycoproteomic comparison of clinical triple-negative and luminal breast tumors. *Journal of proteome research* 2015, 14(3):1376-1388.

162. Tan GJ, Peng ZK, Lu JP, Tang FQ: Cathepsins mediate tumor metastasis. *World journal of biological chemistry* 2013, 4(4):91-101.

163. Al-Mulla F, Marafie M, Zea Tan T, Paul Thiery J: Raf kinase inhibitory protein role in the molecular subtyping of breast cancer. *Journal of cellular biochemistry* 2014, 115(3):488-497.

164. Jiang P, Enomoto A, Takahashi M: Cell biology of the movement of breast cancer cells: intracellular signalling and the actin cytoskeleton. *Cancer letters* 2009, 284(2):122-130.

165. Shi P, Liu W, Tala, Wang H, Li F, Zhang H, Wu Y, Kong Y, Zhou Z, Wang C *et al*: Metformin suppresses triple-negative breast cancer stem cells by targeting KLF5 for degradation. *Cell discovery* 2017, 3:17010.

166. Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adelaide J, Cervera N, Fekairi S, Xerri L, Jacquemier J, Birnbaum D *et al*: Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* 2006, 25(15):2273-2284.

167. Zhao J, Meyerkord CL, Du Y, Khuri FR, Fu H: 14-3-3 proteins as potential therapeutic targets. *Seminars in cell & developmental biology* 2011, 22(7):705-712.

168. Shankar J, Nabi IR: Actin cytoskeleton regulation of epithelial mesenchymal transition in metastatic cancer cells. *PloS one* 2015, 10(3):e0119954.

169. Wang Y, Liu J, Ying X, Lin PC, Zhou BP: Twist-mediated Epithelial-mesenchymal Transition Promotes Breast Tumor Cell Invasion via Inhibition of Hippo Pathway. *Scientific reports* 2016, 6:24606.

170. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al*: Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009, 27(8):1160-1167.

171. Takahashi RU, Takeshita F, Honma K, Ono M, Kato K, Ochiya T: Ribophorin II regulates breast tumor initiation and metastasis through the functional suppression of GSK3beta. *Scientific reports* 2013, 3:2474.

172. Neal CL, Yao J, Yang W, Zhou X, Nguyen NT, Lu J, Danes CG, Guo H, Lan KH, Ensor J *et al*: 14-3-3zeta overexpression defines high risk for breast cancer recurrence and promotes cancer cell survival. *Cancer research* 2009, 69(8):3425-3432.

173. Martinez A, Vos M, Guedez L, Kaur G, Chen Z, Garayoa M, Pio R, Moody T, Stetler-Stevenson WG, Kleinman HK *et al*: The effects of adrenomedullin overexpression in breast tumor cells. *Journal of the National Cancer Institute* 2002, 94(16):1226-1237.

174. Siclari VA, Mohammad KS, Tompkins DR, Davis H, McKenna CR, Peng X, Wessner LL, Niewolna M, Guise TA, Suvannasankha A *et al*: Tumor-expressed adrenomedullin accelerates breast cancer bone metastasis. *Breast cancer research : BCR* 2014, 16(6):458.

175. Hata K, Takebayashi Y, Akiba S, Fujiwaki R, Iida K, Nakayama K, Nakayama S, Fukumoto M, Miyazaki K: Expression of the adrenomedullin gene in epithelial ovarian cancer. *Molecular human reproduction* 2000, 6(10):867-872.

176. Simpson PT, Gale T, Reis-Filho JS, Jones C, Parry S, Steele D, Cossu A, Budroni M, Palmieri G, Lakhani SR: Distribution and significance of 14-3-3sigma, a novel myoepithelial marker, in normal, benign, and malignant breast tissue. *The Journal of pathology* 2004, 202(3):274-285.

177. Boudreau A, Tanner K, Wang D, Geyer FC, Reis-Filho JS, Bissell MJ: 14-3-3sigma stabilizes a complex of soluble actin and intermediate filament to enable breast tumor invasion. *Proceedings of the National Academy of Sciences of the United States of America* 2013, 110(41):E3937-3944.

178. Switzer CH, Cheng RY, Vitek TM, Christensen DJ, Wink DA, Vitek MP: Targeting SET/I(2)PP2A oncoprotein functions as a multi-pathway strategy for cancer therapy. *Oncogene* 2011, 30(22):2504-2513.

179. Tsunezumi J, Higashi S, Miyazaki K: Matrilysin (MMP-7) cleaves C-type lectin domain family 3 member A (CLEC3A) on tumor cell surface and modulates its cell adhesion activity. *Journal of cellular biochemistry* 2009, 106(4):693-702.

180. Lau D, Elezagic D, Hermes G, Morgelin M, Wohl AP, Koch M, Hartmann U, Hollriegl S, Wagener R, Paulsson M *et al*: The cartilage-specific lectin C-type lectin domain family 3 member A (CLEC3A) enhances tissue plasminogen activator-mediated plasminogen activation. *The Journal of biological chemistry* 2018, 293(1):203-214.

181. Mader CC, Oser M, Magalhaes MA, Bravo-Cordero JJ, Condeelis J, Koleske AJ, Gil-Henn H: An EGFR-Src-Arg-cortactin pathway mediates functional maturation of invadopodia and breast cancer cell invasion. *Cancer research* 2011, 71(5):1730-1741.

182. Kuhajda FP: Fatty acid synthase and cancer: new application of an old pathway. *Cancer research* 2006, 66(12):5977-5980.

183. Akkiprik M, Feng Y, Wang H, Chen K, Hu L, Sahin A, Krishnamurthy S, Ozer A, Hao X, Zhang W: Multifunctional roles of insulin-like growth factor binding protein 5 in breast cancer. *Breast cancer research : BCR* 2008, 10(4):212.

184. Yamaga R, Ikeda K, Boele J, Horie-Inoue K, Takayama K, Urano T, Kaida K, Carninci P, Kawai J, Hayashizaki Y *et al*: Systemic identification of estrogen-regulated genes in breast cancer cells through cap analysis of gene expression mapping. *Biochemical and biophysical research communications* 2014, 447(3):531-536.

185. Naderi A: Prolactin-induced protein in breast cancer. *Advances in experimental medicine and biology* 2015, 846:189-200.

186. Arrigo AP, Gibert B: HspB1, HspB5 and HspB4 in Human Cancers: Potent Oncogenic Role of Some of Their Client Proteins. *Cancers* 2014, 6(1):333-365.

187. Yang Y, Zhang Y, Wu Q, Cui X, Lin Z, Liu S, Chen L: Clinical implications of high NQO1 expression in breast cancers. *Journal of experimental & clinical cancer research : CR* 2014, 33:14.

188. Jing H, Song J, Zheng J: Discoidin domain receptor 1: New star in cancer-targeted therapy and its complex role in breast carcinoma. *Oncology letters* 2018, 15(3):3403-3408.

189. Stuhlmiller TJ, Miller SM, Zawistowski JS, Nakamura K, Beltran AS, Duncan JS, Angus SP, Collins KA, Granger DA, Reuther RA *et al*: Inhibition of Lapatinib-Induced Kinome Reprogramming in ERBB2-Positive Breast Cancer by Targeting BET Family Bromodomains. *Cell reports* 2015, 11(3):390-404.

190. Heng B, Lim CK, Lovejoy DB, Bessede A, Gluch L, Guillemin GJ: Understanding the role of the kynurenine pathway in human breast cancer immunobiology. *Oncotarget* 2016, 7(6):6506-6520.

191. Albo D, Berger DH, Wang TN, Hu X, Rothman V, Tuszynski GP: Thrombospondin-1 and transforming growth factor-beta I promote breast tumor cell invasion through up-regulation of the plasminogen/plasmin system. *Surgery* 1997, 122(2):493-499; discussion 499-500.

192. Fontana A, Filleur S, Guglielmi J, Frappart L, Bruno-Bossio G, Boissier S, Cabon F, Clezardin P: Human breast tumors override the antiangiogenic effect of stromal thrombospondin-1 in vivo. *International journal of cancer* 2005, 116(5):686-691.

193. Wu ZS, Wu Q, Yang JH, Wang HQ, Ding XD, Yang F, Xu XC: Prognostic significance of MMP-9 and TIMP-1 serum and tissue expression in breast cancer. *International journal of cancer* 2008, 122(9):2050-2056.

194. Kaiser BK, Yim D, Chow IT, Gonzalez S, Dai Z, Mann HH, Strong RK, Groh V, Spies T: Disulphide-isomerase-enabled shedding of tumour-associated NKG2D ligands. *Nature* 2007, 447(7143):482-486.

195. Gao H, Sun B, Fu H, Chi X, Wang F, Qi X, Hu J, Shao S: PDIA6 promotes the proliferation of HeLa cells through activating the Wnt/beta-catenin signaling pathway. *Oncotarget* 2016, 7(33):53289-53298.

196. Iwasaki M, Homma S, Hishiya A, Dolezal SJ, Reed JC, Takayama S: BAG3 regulates motility and adhesion of epithelial cancer cells. *Cancer research* 2007, 67(21):10252-10259.

197. Nusrat A, Giry M, Turner JR, Colgan SP, Parkos CA, Carnes D, Lemichez E, Boquet P, Madara JL: Rho protein regulates tight junctions and perijunctional actin organization in polarized epithelia. *Proceedings of the National Academy of Sciences of the United States of America* 1995, 92(23):10629-10633.

198. Kleer CG, van Golen KL, Zhang Y, Wu ZF, Rubin MA, Merajver SD: Characterization of RhoC expression in benign and malignant breast disease: a potential new marker for small breast carcinomas with metastatic ability. *The American journal of pathology* 2002, 160(2):579-584.

199. Jia T, Liu YE, Liu J, Shi YE: Stimulation of breast cancer invasion and metastasis by synuclein gamma. *Cancer research* 1999, 59(3):742-747.

200. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, Vidaurre T, Holmes F, Souchon E, Wang H *et al*: A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *Jama* 2011, 305(18):1873-1881.

201. Chuthapisith S, Bean BE, Cowley G, Eremin JM, Samphao S, Layfield R, Kerr ID, Wiseman J, El-Sheemy M, Sreenivasan T *et al*: Annexins in human breast cancer: Possible predictors of pathological response to neoadjuvant chemotherapy. *Eur J Cancer* 2009, 45(7):1274-1281.

202. Florczyk U, Golda S, Zieba A, Cisowski J, Jozkowicz A, Dulak J: Overexpression of biliverdin reductase enhances resistance to chemotherapeutics. *Cancer letters* 2011, 300(1):40-47.

203. Wang L, Wrobel JA, Xie L, Li D, Zurlo G, Shen H, Yang P, Wang Z, Peng Y, Gunawardena HP *et al*: Novel RNA-Affinity Proteogenomics Dissects Tumor Heterogeneity for Revealing Personalized Markers in Precision Prognosis of Cancer. *Cell Chem Biol* 2018, 25(5):619-633 e615.

204. Pavlou MP, Diamandis EP, Blasutig IM: The long journey of cancer biomarkers from the bench to the clinic. *Clinical chemistry* 2013, 59(1):147-157.

205. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, Douville C, Javed AA, Wong F, Mattox A *et al*: Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018, 359(6378):926-930.

206. Gatza ML, Silva GO, Parker JS, Fan C, Perou CM: An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nature genetics* 2014, 46(10):1051-1059.

207. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL *et al*: Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* 2010, 464(7291):999-1005.

208. The Cancer Genome Atlas Network: Comprehensive molecular portraits of human breast tumours. *Nature* 2012, 490(7418):61-70.

209. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET: Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences of the United States of America* 2003, 100(18):10393-10398.

210. De Laurentiis M, Cianniello D, Caputo R, Stanzione B, Arpino G, Cinieri S, Lorusso V, De Placido S: Treatment of triple negative breast cancer (TNBC): current options and future perspectives. *Cancer treatment reviews* 2010, 36 Suppl 3:S80-86.

211. Berrada N, Delaloge S, Andre F: Treatment of triple-negative metastatic breast cancer: toward individualized targeted treatments or chemosensitization? *Annals of oncology : official journal of the European Society for Medical Oncology* 2010, 21 Suppl 7:vii30-35.

212. Collett K, Eide GE, Arnes J, Stefansson IM, Eide J, Braaten A, Aas T, Otte AP, Akslen LA: Expression of enhancer of zeste homologue 2 is significantly associated with increased tumor cell proliferation and is a marker of aggressive breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2006, 12(4):1168-1174.

213. Hu S, Yu L, Li Z, Shen Y, Wang J, Cai J, Xiao L, Wang Z: Overexpression of EZH2 contributes to acquired cisplatin resistance in ovarian cancer cells in vitro and in vivo. *Cancer biology & therapy* 2010, 10(8):788-795.

214. Zhang Y, Liu G, Lin C, Liao G, Tang B: Silencing the EZH2 gene by RNA interference reverses the drug resistance of human hepatic multidrug-resistant cancer cells to 5-Fu. *Life sciences* 2013, 92(17-19):896-902.

215. Kim KH, Roberts CW: Targeting EZH2 in cancer. *Nature medicine* 2016, 22(2):128-134.

216. Easwaran H, Tsai HC, Baylin SB: Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Molecular cell* 2014, 54(5):716-727.

217. Gan L, Yang Y, Li Q, Feng Y, Liu T, Guo W: Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. *Biomarker research* 2018, 6:10.

218. Ezhkova E, Pasolli HA, Parker JS, Stokes N, Su IH, Hannon G, Tarakhovsky A, Fuchs E: Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* 2009, 136(6):1122-1135.

219. Tan JZ, Yan Y, Wang XX, Jiang Y, Xu HE: EZH2: biology, disease, and structure-based drug discovery. *Acta pharmacologica Sinica* 2014, 35(2):161-174.

220. Lee LY, Thaysen-Andersen M, Baker MS, Packer NH, Hancock WS, Fanayan S: Comprehensive N-glycome profiling of cultured human epithelial breast cells identifies unique secretome N-glycosylation signatures enabling tumorigenic subtype classification. *Journal of proteome research* 2014, 13(11):4783-4795.

221. Kirkegaard T, Jaattela M: Lysosomal involvement in cell death and cancer. *Biochimica et biophysica acta* 2009, 1793(4):746-754.

222. Tardy C, Codogno P, Autefage H, Levade T, Andrieu-Abadie N: Lysosomes and lysosomal proteins in cancer cell death (new players of an old struggle). *Biochimica et biophysica acta* 2006, 1765(2):101-125.

223. Silvera D, Formenti SC, Schneider RJ: Translational control in cancer. *Nature reviews Cancer* 2010, 10(4):254-266.

224. Petrocca F, Altschuler G, Tan SM, Mendillo ML, Yan H, Jerry DJ, Kung AL, Hide W, Ince TA, Lieberman J: A genome-wide siRNA screen identifies proteasome addiction as a vulnerability of basal-like triple-negative breast cancer cells. *Cancer cell* 2013, 24(2):182-196.

225. Pan D: The hippo signaling pathway in development and cancer. *Developmental cell* 2010, 19(4):491-505.

226. Bozza WP, Zhang Y, Hallett K, Rivera Rosado LA, Zhang B: RhoGDI deficiency induces constitutive activation of Rho GTPases and COX-2 pathways in association with breast cancer progression. *Oncotarget* 2015, 6(32):32723-32736.

227. Radisky ES, Raeeszadeh-Sarmazdeh M, Radisky DC: Therapeutic Potential of Matrix Metalloproteinase Inhibition in Breast Cancer. *Journal of cellular biochemistry* 2017, 118(11):3531-3548.

228. Friedl P, Gilmour D: Collective cell migration in morphogenesis, regeneration and cancer. *Nature reviews Molecular cell biology* 2009, 10(7):445-457.

229. Kaenel P, Mosimann M, Andres AC: The multifaceted roles of Eph/ephrin signaling in breast cancer. *Cell adhesion & migration* 2012, 6(2):138-147.

230. Zhang ZZ, Shen ZY, Shen YY, Zhao EH, Wang M, Wang CJ, Cao H, Xu J: HOTAIR Long Noncoding RNA Promotes Gastric Cancer Metastasis through Suppression of Poly r(C)-Binding Protein (PCBP) 1. *Molecular cancer therapeutics* 2015, 14(5):1162-1170.

231. Chen S, Sheng C, Liu D, Yao C, Gao S, Song L, Jiang W, Li J, Huang W: Enhancer of zeste homolog 2 is a negative regulator of mitochondria-mediated innate immune responses. *J Immunol* 2013, 191(5):2614-2623.

232. Zhang R, Wang R, Chang H, Wu F, Liu C, Deng D, Fan W: Downregulation of Ezh2 expression by RNA interference induces cell cycle arrest in the G0/G1 phase and apoptosis in U87 human glioma cells. *Oncology reports* 2012, 28(6):2278-2284.

233. Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D: Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Molecular cell* 2008, 32(4):503-518.

234. Lu J, Tan M, Cai Q: The Warburg effect in tumor progression: mitochondrial oxidative metabolism as an anti-metastasis mechanism. *Cancer letters* 2015, 356(2 Pt A):156-164.

235. Pang B, Zheng XR, Tian JX, Gao TH, Gu GY, Zhang R, Fu YB, Pang Q, Li XG, Liu Q: EZH2 promotes metabolic reprogramming in glioblastomas through epigenetic repression of EAF2-HIF1alpha signaling. *Oncotarget* 2016, 7(29):45134-45143.

236. Tao T, Chen M, Jiang R, Guan H, Huang Y, Su H, Hu Q, Han X, Xiao J: Involvement of EZH2 in aerobic glycolysis of prostate cancer through miR-181b/HK2 axis. *Oncology reports* 2017, 37(3):1430-1436.

237. Yoon JH, Yea K, Kim J, Choi YS, Park S, Lee H, Lee CS, Suh PG, Ryu SH: Comparative proteomic analysis of the insulin-induced L6 myotube secretome. *Proteomics* 2009, 9(1):51-60.

238. Choi SS, Lee HJ, Lim I, Satoh J, Kim SU: Human astrocytes: secretome profiles of cytokines and chemokines. *PloS one* 2014, 9(4):e92325.

239. Matta A, Karim MZ, Isenman DE, Erwin WM: Molecular Therapy for Degenerative Disc Disease: Clues from Secretome Analysis of the Notochordal Cell-Rich Nucleus Pulposus. *Scientific reports* 2017, 7:45623.

240. Brandi J, Dalla Pozza E, Dando I, Biondani G, Robotti E, Jenkins R, Elliott V, Park K, Marengo E, Costello E *et al*: Secretome protein signature of human pancreatic cancer stem-like cells. *Journal of proteomics* 2016, 136:1-12.

241. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* 2008, 40(12):1413-1415.

242. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008, 456(7221):470-476.

243. Lee SC, Abdel-Wahab O: Therapeutic targeting of splicing in cancer. *Nature medicine* 2016, 22(9):976-986.

244. Anczukow O, Krainer AR: The spliceosome, a potential Achilles heel of MYC-driven tumors. *Genome medicine* 2015, 7:107.

245. Wahl MC, Will CL, Luhrmann R: The spliceosome: design principles of a dynamic RNP machine. *Cell* 2009, 136(4):701-718.

246. Will CL, Luhrmann R: Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* 2011, 3(7).

247. Jurica MS, Moore MJ: Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* 2003, 12(1):5-14.

248. Quidville V, Alsafadi S, Goubar A, Commo F, Scott V, Pioche-Durieu C, Girault I, Baconnais S, Le Cam E, Lazar V *et al*: Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. *Cancer research* 2013, 73(7):2247-2258.

249. Saltzman AL, Pan Q, Blencowe BJ: Regulation of alternative splicing by the core spliceosomal machinery. *Genes & development* 2011, 25(4):373-384.

250. Bezzi M, Teo SX, Muller J, Mok WC, Sahu SK, Vardy LA, Bonday ZQ, Guccione E: Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery. *Genes & development* 2013, 27(17):1903-1916.

251. Kim YD, Lee J, Kim HS, Lee MO, Son MY, Yoo CH, Choi JK, Lee SC, Cho YS: The unique spliceosome signature of human pluripotent stem cells is mediated by SNRPA1, SNRPD1, and PNN. *Stem cell research* 2017, 22:43-53.

252. Xiong XP, Vogler G, Kurthkoti K, Samsonova A, Zhou R: SmD1 Modulates the miRNA Pathway Independently of Its Pre-mRNA Splicing Function. *PLoS genetics* 2015, 11(8):e1005475.

253. Pan K, Liang XT, Zhang HK, Zhao JJ, Wang DD, Li JJ, Lian Q, Chang AE, Li Q, Xia JC: Characterization of bridging integrator 1 (BIN1) as a potential tumor suppressor and prognostic marker in hepatocellular carcinoma. *Mol Med* 2012, 18:507-518.

254. Nicolussi A, D'Inzeo S, Capalbo C, Giannini G, Coppa A: The role of peroxiredoxins in cancer. *Molecular and clinical oncology* 2017, 6(2):139-153.

255. Kim SY, Kang HT, Choi HR, Park SC: Biliverdin reductase A in the prevention of cellular senescence against oxidative stress. *Experimental & molecular medicine* 2011, 43(1):15-23.

256. Cooper J, Giancotti FG: Integrin Signaling in Cancer: Mechanotransduction, Stemness, Epithelial Plasticity, and Therapeutic Resistance. *Cancer cell* 2019, 35(3):347-367.

257. Vega FM, Ridley AJ: Rho GTPases in cancer cell biology. *FEBS letters* 2008, 582(14):2093-2101.

258. Yamaguchi H, Condeelis J: Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Biochimica et biophysica acta* 2007, 1773(5):642-652.

259. Bai H, Harmanci AS, Erson-Omay EZ, Li J, Coskun S, Simon M, Krischek B, Ozduman K, Omay SB, Sorensen EA *et al*: Integrated genomic characterization of IDH1-mutant glioma malignant progression. *Nature genetics* 2016, 48(1):59-66.

260. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, Kos I, Batinic-Haberle I, Jones S, Riggins GJ *et al*: IDH1 and IDH2 mutations in gliomas. *The New England journal of medicine* 2009, 360(8):765-773.

261. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL *et al*: An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008, 321(5897):1807-1812.

262. Cohen AL, Holmen SL, Colman H: IDH1 and IDH2 mutations in gliomas. *Current neurology and neuroscience reports* 2013, 13(5):345.

263. Lai A, Kharbanda S, Pope WB, Tran A, Solis OE, Peale F, Forrest WF, Pujara K, Carrillo JA, Pandita A *et al*: Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011, 29(34):4482-4490.

264. Sanson M, Marie Y, Paris S, Idbaih A, Laffaire J, Ducray F, El Hallani S, Boisselier B, Mokhtari K, Hoang-Xuan K *et al*: Isocitrate dehydrogenase 1 codon 132 mutation is an important prognostic biomarker in gliomas. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2009, 27(25):4150-4154.

265. Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC *et al*: Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 2009, 462(7274):739-744.

266. Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, Ito S, Yang C, Xiao MT, Liu LX *et al*: Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer cell* 2011, 19(1):17-30.

267. Rohle D, Popovici-Muller J, Palaskas N, Turcan S, Grommes C, Campos C, Tsoi J, Clark O, Oldrini B, Komisopoulou E *et al*: An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* 2013, 340(6132):626-630.

268. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, Campos C, Fabius AW, Lu C, Ward PS *et al*: IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 2012, 483(7390):479-483.

269. Myllyharju J, Kivirikko KI: Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in genetics : TIG* 2004, 20(1):33-43.

270. Sasaki M, Knobbe CB, Itsumi M, Elia AJ, Harris IS, Chio, II, Cairns RA, McCracken S, Wakeham A, Haight J *et al*: D-2-hydroxyglutarate produced by mutant IDH1 perturbs collagen maturation and basement membrane function. *Genes & development* 2012, 26(18):2038-2049.

271. Chesnelong C, Chaumeil MM, Blough MD, Al-Najjar M, Stechishin OD, Chan JA, Pieper RO, Ronen SM, Weiss S, Luchman HA *et al*: Lactate dehydrogenase A silencing in IDH mutant gliomas. *Neuro-oncology* 2014, 16(5):686-695.

272. Miroshnikova YA, Mouw JK, Barnes JM, Pickup MW, Lakins JN, Kim Y, Lobo K, Persson AI, Reis GF, McKnight TR *et al*: Tissue mechanics promote IDH1-dependent HIF1alpha-tenascin C feedback to regulate glioblastoma aggression. *Nature cell biology* 2016, 18(12):1336-1345.

273. Kickingereder P, Sahm F, Radbruch A, Wick W, Heiland S, Deimling A, Bendszus M, Wiestler B: IDH mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma. *Scientific reports* 2015, 5:16238.

274. Koivunen P, Lee S, Duncan CG, Lopez G, Lu G, Ramkissoon S, Losman JA, Joensuu P, Bergmann U, Gross S *et al*: Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* 2012, 483(7390):484-488.

275. Zhao S, Lin Y, Xu W, Jiang W, Zha Z, Wang P, Yu W, Li Z, Gong L, Peng Y *et al*: Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha. *Science* 2009, 324(5924):261-265.

276. Wang G, Sai K, Gong F, Yang Q, Chen F, Lin J: Mutation of isocitrate dehydrogenase 1 induces glioma cell proliferation via nuclear factor-kappaB activation in a hypoxia-inducible factor 1-alpha dependent manner. *Molecular medicine reports* 2014, 9(5):1799-1805.

277. Andronesi OC, Kim GS, Gerstner E, Batchelor T, Tzika AA, Fantin VR, Vander Heiden MG, Sorensen AG: Detection of 2-hydroxyglutarate in IDH-mutated glioma patients by in vivo spectral-editing and 2D correlation magnetic resonance spectroscopy. *Science translational medicine* 2012, 4(116):116ra114.

278. Elkhaled A, Jalbert LE, Phillips JJ, Yoshihara HAI, Parvataneni R, Srinivasan R, Bourne G, Berger MS, Chang SM, Cha S *et al*: Magnetic resonance of 2-hydroxyglutarate in IDH1-mutated low-grade gliomas. *Science translational medicine* 2012, 4(116):116ra115.

279. Yang F, Shen Y, Camp DG, 2nd, Smith RD: High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert review of proteomics* 2012, 9(2):129-134.

280. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suva ML, Bernstein BE: Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 2016, 529(7584):110-114.

281. Yao Q, Cai G, Yu Q, Shen J, Gu Z, Chen J, Shi W, Shi J: IDH1 mutation diminishes aggressive phenotype in glioma stem cells. *International journal of oncology* 2018, 52(1):270-278.

282. Grassian AR, Parker SJ, Davidson SM, Divakaruni AS, Green CR, Zhang X, Slocum KL, Pu M, Lin F, Vickers C *et al*: IDH1 mutations alter citric acid cycle metabolism and increase dependence on oxidative mitochondrial metabolism. *Cancer research* 2014, 74(12):3317-3331.

283. Serao NV, Delfino KR, Southey BR, Beever JE, Rodriguez-Zas SL: Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. *BMC medical genomics* 2011, 4:49.

284. Sabit H, Nakada M, Furuta T, Watanabe T, Hayashi Y, Sato H, Kato Y, Hamada J: Characterizing invading glioma cells based on IDH1-R132H and Ki-67 immunofluorescence. *Brain tumor pathology* 2014, 31(4):242-246.

285. Polivka J, Jr., Pesta M, Pitule P, Hes O, Holubec L, Polivka J, Kubikova T, Tonar Z: IDH1 mutation is associated with lower expression of VEGF but not microvessel formation in glioblastoma multiforme. *Oncotarget* 2018, 9(23):16462-16476.

286. Fu X, Chin RM, Vergnes L, Hwang H, Deng G, Xing Y, Pai MY, Li S, Ta L, Fazlollahi F *et al*: 2-Hydroxyglutarate Inhibits ATP Synthase and mTOR Signaling. *Cell metabolism* 2015, 22(3):508-515.

287. German NJ, Haigis MC: Sirtuins and the Metabolic Hurdles in Cancer. *Current biology : CB* 2015, 25(13):R569-583.

288. Birner P, Pusch S, Christov C, Mihaylova S, Toumangelova-Uzeir K, Natchev S, Schoppmann SF, Tchorbanov A, Streubel B, Tuettenberg J *et al*: Mutant IDH1 inhibits PI3K/Akt signaling in human glioma. *Cancer* 2014, 120(16):2440-2447.

289. Day BW, Stringer BW, Boyd AW: Eph receptors as therapeutic targets in glioblastoma. *British journal of cancer* 2014, 111(7):1255-1261.

290. Miao H, Gale NW, Guo H, Qian J, Petty A, Kaspar J, Murphy AJ, Valenzuela DM, Yancopoulos G, Hambardzumyan D *et al*: EphA2 promotes infiltrative invasion of glioma stem cells in vivo through cross-talk with Akt and regulates stem cell properties. *Oncogene* 2015, 34(5):558-567.

291. Sulkowski PL, Corso CD, Robinson ND, Scanlon SE, Purshouse KR, Bai H, Liu Y, Sundaram RK, Hegan DC, Fons NR *et al*: 2-Hydroxyglutarate produced by neomorphic IDH mutations suppresses homologous recombination and induces PARP inhibitor sensitivity. *Science translational medicine* 2017, 9(375).

292. Almiron Bonnin DA, Havrda MC, Israel MA: Glioma Cell Secretion: A Driver of Tumor Progression and a Potential Therapeutic Target. *Cancer research* 2018, 78(21):6031-6039.

293. Nakada M, Kita D, Watanabe T, Hayashi Y, Teng L, Pyko IV, Hamada J: Aberrant signaling pathways in glioma. *Cancers* 2011, 3(3):3242-3278.

294. Westphal M, Maire CL, Lamszus K: EGFR as a Target for Glioblastoma Treatment: An Unfulfilled Promise. *CNS drugs* 2017, 31(9):723-735.

295. Pandini G, Vigneri R, Costantino A, Frasca F, Ippolito A, Fujita-Yamaguchi Y, Siddle K, Goldfine ID, Belfiore A: Insulin and insulin-like growth factor-I (IGF-I) receptor overexpression in breast cancers leads to insulin/IGF-I hybrid receptor overexpression: evidence for a second mechanism of IGF-I signaling. *Clinical cancer research : an official journal of the American Association for Cancer Research* 1999, 5(7):1935-1944.

296. Cox ME, Gleave ME, Zakikhani M, Bell RH, Piura E, Vickers E, Cunningham M, Larsson O, Fazli L, Pollak M: Insulin receptor expression by human prostate cancers. *The Prostate* 2009, 69(1):33-40.

297. Vella V, Sciacca L, Pandini G, Mineo R, Squatrito S, Vigneri R, Belfiore A: The IGF system in thyroid cancer: new concepts. *Molecular pathology : MP* 2001, 54(3):121-124.

298. Sara VR, Prisell P, Sjogren B, Persson L, Boethius J, Enberg G: Enhancement of insulin-like growth factor 2 receptors in glioblastoma. *Cancer letters* 1986, 32(3):229-234.

299. Singh P, Alex JM, Bast F: Insulin receptor (IR) and insulin-like growth factor receptor 1 (IGF-1R) signaling systems: novel treatment strategies for cancer. *Med Oncol* 2014, 31(1):805.