Proteomic Based Approaches for Differentiating Tumor Subtypes

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Linan Wang

Graduate Program in Integrated Biomedical Science Program

The Ohio State University

2017

Dissertation Committee:

Michael Alan Freitas, PhD Advisor

Charles Lawrence Hitchcock, MD/PhD

Kun Huang, PhD

Mark Robert Parthun, PhD
Copyright by
Linan Wang
2017
Abstract

In medicine, successful patient treatment relies on early and accurate diagnosis. Following diagnosis disease specific and effective treatments are necessary, targeting affected cells while sparing normal tissue. While past studies have focused on genomics, the importance of transcriptomics and proteomics is increasingly understood. Proteomics, the study of proteins, will be the focus of this dissertation. Proteomics provide insight in the post transcriptional and translational regulation of proteins, information not available through the study of DNA and RNA alone. These effects play an important role in protein quantity and physiological function. It is well established that changes in protein homeostasis are associated with disease conditions, hence providing the grounds for biomarker discovery. It has been shown that if homeostasis can be restored, disease conditions can be reversed, further emphasizing the role of proteomics in therapeutic target discovery.

Chapter 1 highlights the importance of proteomics in the field of biomedical research with an emphasis on clinical translational sciences in moving discoveries from bench to bedside. Chapters 2 of this dissertation describe the development of methodology for the study of archived clinical biopsy samples. Following biopsy, patient tissue is preserved with formalin fixation and paraffin embedding (FFPE) and archived. Such tissue is stable for research for decades. This document will describe a method to prepare this
tissue in for proteomic studies using LC-MS/MS with a novel on-slide digestion technique to be used with manual microdissection. Then using this technique, possibility to distinguish the different thymoma subtypes with a proteomics approach was investigated to provide an objective diagnosis tool to complement the current histological diagnosis. In Chapter 3, desmoyokin, a protein found to be unique to the medulla of the thymus and not present in the cortex will be described. This protein holds great potential in the separation of type A and B thymomas without confounding of potentially overlapping histological features. Clinically, this allows for objective distinguishing of thymoma subtypes A and B3, which have drastically different clinical outcomes.

Chapter 4 of this dissertation discuss histone epigenetic regulation through post-translational modification (PTM) and the associated histone crosstalk. Histones are proteins that combine with DNA to form chromatin. The effects of these PTMs on global protein abundance was assessed using stable isotope labeling in cell culture (SILAC) and tandem mass spectrometry (LC-MS/MS) providing a comparison of protein relative abundance between wild type cells and histone mutants. Experimental data showed there was a discrepancy between protein expression and mRNA levels, which is currently a part of ongoing investigation. This series of experiments emphasizes the importance of proteomics in modern day biological science and medical research.
Dedication

This document is dedicated to all individuals who have supported me during my pursuit of this PhD. I also would like to dedicate this document to those who pursue a career in science and/or medicine seeking to build a healthier world.
Acknowledgments

In addition to my advisor and committee mentioned above, I wish to acknowledge the Center for Clinical Translational Sciences (CCTS) for the 2 years of financial support through the CCTS TL1 training grant. Furthermore, I would like to acknowledge Dr. Philip Binkley and his exceptional knowledge and mentoring guided me through the Clinical Translational Sciences area of emphasis. Special thanks to Dr. Virginia Sanders for her assistance in the development of grant writing skills and providing feedback for every fellowship grant I sent in her direction. I would like to thank former and present members of The Ohio State University College of Medicine for their support during my training period: Dr. Linda Stone, Dr. Richard Fertel and Patti Fertel. Finally, a special acknowledgement to William Hankey, classmate and colleague, for his ever helpful research suggestions derived from his extensive experience.
Vita


August 2009 to present ........................................ Graduate Research Associate/Fellow,

Department of Molecular Virology,

Immunology and Medical Genetics, The Ohio State University.

Publications


Fields of Study

Major Field: Integrated Biomedical Science Program
# Table of Contents

Abstract ................................................................................................................................. ii

Dedication ............................................................................................................................... iv

Acknowledgments ................................................................................................................ v

Vita ........................................................................................................................................ vi

List of Tables ............................................................................................................................ x

List of Figures .......................................................................................................................... xi

Chapter 1: Proteomics: An Effector Based Approach to Understanding Cancer and Improving Clinical Diagnosis and Treatment ................................................................. 1

Chapter 2: Anionic Detergent Mediated On-Slide Digestion: A Novel Approach to Process Archived Tissue for Bottom-Up Proteomics ................................................................. 23

Chapter 3: Proteomic Signatures of Thymomas ................................................................... 52

Chapter 4: Measuring the Impact of Chromatin Modifications on Global Protein Expression ................................................................................................................................. 89

Summary ................................................................................................................................ 118

References .............................................................................................................................. 120

viii
List of Tables

Table 1. Histological Features of the Thymomas. ................................................................. 54
Table 2. Patient Details for LC-MS/MS Experiment.......................................................... 57
Table 3. Patient age and gender for tissue specimens used for tumor microarray studies. .................................................................................................................................................. 63
Table 4. Collagen Proteins Differentially Expressed in Thymomas from Normal Thymus. .................................................................................................................................................. 67
Table 5. Anti-apoptotic Proteins were Differentially Expressed in Thymomas A and B3. .................................................................................................................................................. 73
Table 6. Proteins Exhibiting Significant Change in SILAC Ratios in H3K4A. ............ 101
Table 7. Proteins Exhibiting Significant Change in SILAC Ratios in H4K16Q. ........... 102
Table 8. Proteins Exhibiting Significant Change in SILAC Ratios in H4K91Q. ........... 103
Table 9. Proteins Exhibiting Significant Change in SILAC Ratios in H4K16R. ........... 104
Table 10. Gene Ontology for Yeast Histone Mutant Proteome Changes. ................. 105
Table 11. DAVID Pathway Analysis. ..................................................................................... 106
Table 12. Gene Ontology Terms for H4K16R Protein Expression Changes........... 110
List of Figures

Figure 1. Tissue Collection Methods Experimental Schematic.............................. 29

Figure 2. Tissue Staining of Reactive Lymph Nodes ........................................... 34

Figure 3. Hematoxylin Staining of Reactive Lymph with 20%, 15%, 10% and 5%
Dilutions.............................................................................................................. 35

Figure 4. Protein Identification and Spectral Count Quantitation in Reactive Lymph
Nodes are Independent of Staining Method......................................................... 37

Figure 5. Spectral Count Correlation of Lymphatic Tissue Components and Tissue
Staining............................................................................................................... 38

Figure 6. Biplot of Principle Components 1 and 2 of Tissue Staining................. 40

Figure 7. Protein Identification and Spectral Count Quantitation in Reactive Lymph
Nodes are Independent of Tissue Collection Method........................................ 43

Figure 8. Impact of Material Input on On-Slide Digestion................................... 46

Figure 9. Tissue Collection from SuperFrost Plus Slides is Mediated by Anionic
Detergents............................................................................................................ 48

Figure 10. Histone 4 Western Blot of Collected Patient Tissue Samples............... 50

Figure 11. Clustering and Venn Diagram of Significant Proteins (q-value <0.20)..... 65

Figure 12. Principal Component Analysis of Thymoma Subtypes....................... 66

Figure 13. Proteomic Profile Contribution of T Lymphocytes.............................. 70
Figure 14. Intratumoral Lymphocytes Throughout Thymoma Subtypes. .................. 71
Figure 15. Changes in Stathmin Expression in Different Thymoma Subtypes. .......... 72
Figure 16. Desmyoyokin is Differentially Expressed in and A and B Thymomas........ 76
Figure 17. Examples of Human Disease Associated with Histone Post Translational
Modifications. ................................................................................................................. 91
Figure 18. Overall Experimental Process. ................................................................. 92
Figure 19. Raw LC-MS/MS data from LTQ Orbitrap. ............................................... 96
Figure 20. Density Plots of SILAC Mutant to Wild Type Ratios.............................. 100
Figure 21. Panther Assessment of Molecular Function ............................................. 106
Figure 22. Protein Expression Levels vs. mRNA Levels in H4K16R ....................... 108
Figure 23. Comparison of Mass Spectrometry, Microarray, and RT PCR Data. ....... 109
Figure 24. Polyribosome Profiling of H4K16............................................................ 112
Figure 25. Polyribosome Profiling of H3K56R......................................................... 113
Chapter 1: Proteomics: An Effector Based Approach to Understanding Cancer and Improving Clinical Diagnosis and Treatment

Introduction

There are few people whose lives have not been influenced by cancer, whether patients themselves or knowing a loved one affected by this deadly disease. Cancer is currently the second leading cause of death in the United States, behind heart disease. It is estimated by the National Cancer Institute (NCI) that in 2016 there will be 1.69 million cases of newly diagnosed cancer and 595,690 people will die from cancer in the United States alone. Almost 40% of people will get cancer at some point in their lifetime. Cancer is a major burden on society economically costing $125 billion in 2010 and is estimated to rise this decade. The overall rates of cancer in the United States has dropped in the last 25 years and the number of cancer survivors are increasing, indicating some success in the battle against cancer. This success has been attributed to improved diagnosis techniques and treatment. Early, more accurate diagnosis techniques have allowed cancer treatment to occur prior to its spread throughout the body, simplifying treatment and reducing complications. New combination therapies targeting multiple aspects of the cancer disease have achieved cancer free states with delayed or no recurrence. Researchers must continue to investigate innovative approaches in order to continue this trend of successful cancer treatment.
Cancer is a broad term describing a large group of disease characterized by the uncontrolled proliferation of abnormal cells. These cells no longer receive extracellular signals from the environment, ignoring growth inhibition and apoptosis signaling. Metastasis occurs when these cells spreads to distant sites of the body through the blood or lymphatic systems. Cancer can occur in virtually any part of the body including breast, lung, colon and liver. Cancer in each of these body areas are further divided into subtypes, which are based on the observed histology of the atypical cells. The staging system is used to characterize the extent of tumor spread and provide a general prognosis. There are several different mechanisms, both genetic and environmental, occurring in combination that leads to disruption of homeostasis. Scientific elucidation of the mechanism leading to cancer is useful in predicting individuals with a predisposition and formation of a prevention plan. Identification of specific abnormalities caused by the loss of homeostasis is useful in serving as diagnostic markers. Finally, extensive effort has been placed in correcting these abnormally function pathways either at the source or downstream in therapeutic development. Alternatively, the differences between normal and abnormal cells provide a means of selectively administering a treatment to destroy cancer cells while sparing normal cells. Though better scientific understanding of cancer, improvements are made to create specific care plans to effectively treat each cancer individually.

Finding the solution to the cancer problem faces three major challenges. First, cancer is an extremely heterogeneous disease knowledge of one type may not apply to other types, requiring individualized procedures and treatments. Secondly, early diagnosis is critical for successful cancer treatment, hence the need for new biomarkers that are of
early and specific diagnostic value. Finally, there is a need for new treatments aimed at restoring cellular regulation or selectively targeting cancer cells while sparing normal cells. Therefore, the fight against cancer is ongoing and must be approached from multiple angles. Among the various novel therapeutic treatments being investigated, epigenetics, specifically histone post-translational modifications on the amino acid residues have been the subject of many studies to identify novel disease biomarkers and therapeutic targets.

What is Proteomics?

The -Omics

Proteomics is a term used to describe the study entire cellular content of proteins and their functions. Proteomics is to proteins as genomics is to DNA and transcriptomics is to RNA transcripts. Proteins are highly attractive targets for study since they are the effectors of cellular processes and are subject to regulatory effects that cannot be detected through study of genomics or transcriptomics. While genomics is useful in the study for DNA point mutations, those mutations may not necessarily have an effect on the final protein product due to amino acid codon redundancy or mutation occurring in a non-coding region. Furthermore, while the gene coding for a protein may be present it may only be active during specific cellular events or external circumstances. Measuring mRNA transcripts allows a quantitative measurement of the transcription of a given gene, however, this number is only a rough estimate of the final protein level due to regulation of protein translation and protein degradation. Proteomics is unique in that it takes into account post-translational modifications (PTMs). PTMs regulate protein activities through reversible
enzymatic modifications to a completed translated protein. Known post-translational modifications include acetylation, methylation, phosphorylation and sumoylation. Proteomics also accounts for protein degradation when quantifying specific proteins for their activities. Finally, proteomics accounts for proteins requiring other proteins or cofactors to function, cellular localization and alternative transcript splicing.

*Techniques for Studying Proteomics*

There are many ways of studying proteomics including antibody based approaches such as Western Blotting and immunohistochemistry. A higher throughput assay that is capable of detecting and quantifying protein is enzyme linked immunosorbent assay (ELISA). Antibodies hold the advantage of high sensitivity and usually good specificity but are limited when distinguishing proteins sharing similar antigen regions and the lack of availability in some rarer protein targets. Immunohistochemistry (IHC) allows for the study of protein targets in patient biopsy samples. The utility of IHCs in proteomic studies has further been enhanced by the development of tumor microarrays (TMA) by Kononet et al., allowing for high throughput and consistent antibody staining across multiple tissue biopsy samples.¹ Such techniques have allowed for the study of protein localization within cells and tissues. Antibody techniques are also useful in identifying protein post-translational modifications. Finally, antibodies have proved useful in the study of protein-protein interactions through techniques such as Co-IP, which captures target proteins and their binding partners.
Mass spectrometry proteomics has grown increasingly popular in the past two decades as a high throughput technique. Bottom-up proteomics involves the enzymatic digestion of proteins into peptides that are then analyzed with a mass spectrometer coupled with a high performance liquid chromatography (HPLC) system.\(^2\) Bottom-up proteomics is useful in analyzing complex protein mixtures at the global cellular level due to the solubility and easy of separation of the peptides vs. proteins.\(^3\) Peptides are then fragmented and the observed peptide fragmentation patterns are compared with computer theoretical fragmentation patterns for peptide identification. The peptide sequences are then matched to the most likely proteins of origin. Alternatively, top-down proteomics involves the study of intact proteins. Top-down proteomics is useful in the study of post-translational modifications and protein primary structure.\(^4,5\) Top-down proteomics allow for the characterization of multiple different proteoforms of a single protein, some of which have different activities and functions.\(^6\) Top-down proteomics may also improve protein identification since there is no need to extrapolate the origins of non-unique peptides as in bottom-up proteomics. However, when dealing with intact proteins as in top-down proteomics separation of the intact proteins becomes a challenge limiting the throughput of this approach.\(^7\) Middle-down proteomics involves the study of larger peptides, greater than 3 kDa, and possesses some of the advantages of both bottom-up and top-down proteomics.\(^8,9\)

Various source material has been suitable in the study of proteomics in the context of cancer. Model organisms, such as *Saccharomyces cerevisiae*, has been used successfully in the investigation of histone cross talk and global effects of histone PTM states.\(^10\)
However, while yeast and other model organisms hold several orthologues with human, cancer studies are limited to very specific cellular processes that these organisms hold in common. Various human cell lines have been created to simulate different cancer subtypes. While cell lines easier to obtain and work with than actual patient tissues, there has been an ongoing debate as to their accuracy in representing the tumor and its microenvironment. Cell lines provide a simplified model for cancer as all cells in the culture are mostly homogenous whereas actual tumors are largely heterogenous. Furthermore, most cancer cell lines are fast growing and representative of poorly differentiated high grade metastatic tumors as opposed to the majority of tumors which are lower grade.

Another source of bio specimens for cancer proteomic studies involves the use of animal models. Genetically Engineered Mice (GEM) are used to simulate human cancers. Alternatively, tissue from human primary tumors can be grafted into immune deficient mice. The human mouse xenograft models have proven to be useful for preclinical drug treatment testing due to the tumor incorporated into the mouse has many of the characteristics of the primary human tumor of origin.

An increasing source of proteomic samples today is the use of human tissue itself. The advantages of using patient tissue for proteomic study is the direct representation of the sample to the actual disease. Use of patient tissue has drawn great attention because it accounts for heterogeneity both within the same patient and between patients. Tissues also include in vivo effects such as surrounding tissue as well as signaling molecules from other parts of the human body. Patient tissue can be derived from surgical resections and biopsies. Advancements in tissue storage has even allowed the use of samples collected
over time to obtain better sample groups. Tissues collected from patient can be immediately frozen and stored at -80°C for one year. Alternatively, standard pathology practice is to preserve tissues using formalin fixation and paraffin embedding (FFPE). Tissues preserved in this manner has been stable for up to two decades. Studies have shown that FFPE preserved tissues are suitable for mass spectrometry global proteomic analysis. Tissue can be further enriched from a biospecimen thanks to technological advances such as laser capture microdissection. When coupled with mass spectrometry, proteomics characterization of tissues can occur in samples with fewer than 10,000 cells. In addition to tissue use, a variety of human fluids, such as blood, serum, saliva and urine, have been investigated for potential cancer related markers.

The Proteomics of Cancer

Oncoproteins and Tumor Repressors

Oncogenes are genes that have the potential to cause cancer when over expressed. The proteins that these genes code for are known as onco-proteins. These protein play a role in enhancing cell survival such as blocking apoptotic signaling and increasing cellular division without regard for external signaling. Onco-proteins are typically overexpressed in cancers. In a normal cell, proliferation is carefully regulated by tumor suppressor proteins. When a cellular defect is detected, such as unrepaired DNA damage, the tumor suppressor proteins function to arrest the cell cycle to prevent cell division and initiate a pathway leading to cellular apoptosis. In many cases of cancer, tumor suppressor proteins have been inactivated or down regulated. Due to their delicate functional balance these two
groups of proteins, they have drawn much attention in proteomic cancer research. There are too many known onco-proteins and tumor suppressors to describe in the scope of this document, however, below are some examples of extensively studied protein that play a key role in tumors.

ERBB2/HER2 is an example of an onco-protein that has received much attention for its relationship with breast cancer, although it appears in other cancers as well. ERBB2 is a much studied epidermal growth factor receptor binding to factors found in the stroma leading to increased cellular survival.\(^{26-28}\) Perhaps the most studied tumor repressor is p53, which plays many roles including regulation of the cell cycle, induction of apoptosis, activation of DNA repair mechanisms and prevention of angiogenesis.\(^{29,30}\) Several studies have been conducted regarding the various pathways influenced by P53 regulation including glucose metabolism leading to the observed Warburg effect.\(^{31}\) It was shown that the P53 regulation can be through direct binding or indirectly through the use of miRNA.\(^{32}\)

The relationship between onco-proteins and tumor suppressors represent the complexity of protein interactions investigated through proteomics. Balance of individual proteins are necessary to maintain homeostasis of the various cellular processes, such as cell growth, division and apoptosis. Cancer is an example of the result of loss of protein homeostasis. An important concept to understand is that proteins do not function alone, they are a part of a large network in which alteration of one component affects several other proteins. The relationship between proteins is not limited to direct protein-protein interactions. As shown in the P53 studies, proteins may affect other proteins through
indirect mediators such as miRNA. Thus, unraveling the proteomics of cancer remains an interesting and ongoing challenge for researchers.

**Proteoforms and their Relationship to Cancer**

A unique advantage of proteomics is its ability to detect protein proteoforms. Proteoforms as generated from a single gene but with alternative splicing of the transcript. In the previous section, the importance of P53 as a tumor suppressor is discussed. Several studies have shown that through different translation initiation, alternative splicing and use of an internal promoter there are at least twelve P53 proteoforms dysregulation of which may lead to cancer. Phosphatidylinositol 3 kinase (PI3K) is a protein that translates external signals, such as growth factors, to various cellular signaling pathways. PI3K plays a role in proliferation, growth, survival, differentiation, motility and metabolism has described as overactive in many cancers. In a study by Utermark et al. the importance of selectively targeting PI3K isoforms was described. The difference in activity observed in the different proteoforms in tumor associated proteins such as P53 and PI3K illustrate the importance of proteomics in the distinguishing the proteoforms.

The study of proteoforms is a challenge to the field of proteomics where there are still a lot of unknowns. One challenge that researchers face is often there is cross reactivity between proteoforms in antibody based assays leading to difficulties in discerning the effects of individual proteoforms. With the advancement of top-down mass spectrometry proteomics individual proteoforms can now be identified and studied. With improvements
in technology we can expect elucidation of the roles proteoforms play in cancer and other diseases in the near future.

*Post Translational Modifications and Cancer*

Post-translational modification plays an important role on regulation protein function. Histones, proteins heavily associated with PTMs, are a group of DNA binding proteins consisting of the four core histones: H2A, H2B, H3 and H4 where two of each form this histone octamer and histone H1 which serves as a linker histone regulating DNA entry and exit of the histone/DNA complex stability the structure, known as the nucleosome.\(^{39}\) Post-translational modifications to these histones such as acetylation, methylation, phosphorylation, ubiquitination, ADP ribosylation, and sumoylation lead to alteration of chromatin structure altering the gene transcription that in turn leads to changes in proteomic expression.\(^{40}\) Core histones have a long N terminal tail where a large number of PTMs occur and a globular core domain. These protein expression changes are mediated by two known mechanisms: histone PTMs may change the interactions between histones and DNA allowing or denying access to transcriptional machinery, alternatively, histone PTMs may form structures facilitating the binding of further regulators of transcription. The effects of specific histone PTMs varies depending on the residue and histone which they occur. Acetylation, is a common histone PTM occurring on lysine which neutralizes the positive charge leading to decreased association with the negatively charged DNA. This modification is usually associated with increased transcription of associated genes. Methylation of histones is another common PTM which may lead to either up or down
regulation of protein expression. The effects of histone PTMs on cancer has been extensively reviewed by Füllgrabe et al.\textsuperscript{41} In summary, histone PTMs play a role in regulating many cellular processes in which disruption may lead to aberrant gene expression, genomic instability, loss of cell cycle checkpoints and alterations in DNA repair. However, understanding histone PTMs individually is only a small part of the big picture, as histones PTMs are known for cross talk. The PTM state of one histone amino acid residue can affect others creating a complex network of histone PTM interactions.

Histones are just one group of many proteins that are modified post translationally, several other proteins with PTMs have implications in cancer. One important example of a protein subject to PTM regulation is P53. As mentioned in the previous section, P53 plays a key role in stopping proliferation of cells exhibiting abnormalities. Post translational modifications known to occur on P53 include phosphorylation, ubiquitination, acetylation, methylation, glycosylation, farnesylation, hydroxylation, ADP ribosylation and PIN1 mediated prolyl isomerization.\textsuperscript{42-44} Currently, 40 amino acid residues on P53 has been described to be modified by a number of PTMs. The importance of acetylation in the regulation of P53 function in the processes of senescence, apoptosis and cell cycle arrest was demonstrated by Li et al.\textsuperscript{45}

Proteins are subject to many forms of regulation that dramatically affect their functions. As exemplified by P53, a mutation leading to the loss of a single PTM site can lead to loss of function for a protein. The complexity of protein PTMs can be seen through the histone family of proteins, not only do individual modifications affect the chromatin structure, intra- and inter- histone crosstalk additionally effect chromatin regulation, hence
global cellular function. Thus, the proteomics study of protein post translational modifications is essential in understanding protein function and their role in the disease process.

*Diagnosis and Prognosis Using Proteomics*

A proteomic biomarker is defined as a protein whose pattern is uniquely differentiable between a disease and healthy state. The ideal biomarker is both sensitive and specific. It also should be detectable in samples that are easily obtained, preferably through blood, serum or urine. Biomarkers may be involved in the actual cancer process or simply coexisting with the disease process. After a potential biomarker is identified in research it must undergo extensive validation before it can be approved by the US Food and Drug Administration. In 2011, 1,261 proteins had been identified as differentially expressed in tumors, however, of those only 5% were investigated as potential biomarkers and 3% were used in some clinical capacity. This represents the challenge of biomarkers making it from bench to bedside.\(^4^6\) As of 2013, the FDA had approved of 23 proteins to be used in the clinical setting as disease biomarkers.\(^4^7\) The majority of these biomarkers are immunoassay based hence requiring an antibody since alternative techniques such as mass spectrometry is currently not used in clinical practice.

One example of a proteomic biomarker success story is the FDA approval of OVA1 for ovarian cancer testing.\(^4^8\) The OVA1 assay actually consists of tests for five biomarkers associated with ovarian cancer: apolipoprotein A1, beta 2 microglobin, CA125, transferrin and transthyretin (also known as prealbumin). The investigators pursued the question of
what separated the malignant from the non-malignant ovarian tumors using a mass spectrometry proteomics approach with a SELDI TOF instrument for high throughput analysis of clinical samples. Their seven mass spectrometry candidates were eventually narrowed down to four that were suitable for immunoassay hence clinical use. The final four candidates were placed on a panel along with CA125 which was a predetermined ovarian cancer model. After one year of FDA review it was approved for clinical use. The final product has greater than 90% sensitivity and 90% negative predictive value for women diagnosed with ovarian cancer with planned treatment.

Post translational modifications states of proteins have been considered as diagnostic markers. One study assessed the phosphorylation levels in bladder cancer cell lines of linker histone H1. In this study, it was observed that there is an increase in histone several proteoforms of H1 global phosphorylation corresponding with increasing severity of bladder cancer the cell lines represent, from normal to non-invasive to high grade invasive. Further studies showed that percentage of phosphorylation of threonine 146 (pT146) on histone H1 correlated positively with tumor histopathological grade and proliferation rate. The value of H1 T146 as a diagnostic and prognostic marker was further assessed in breast cancer. Harshman et al. demonstrated the ability to distinguish breast cancer cell lines representing immortalized, neoplastic and metastatic breast tissue. They further demonstrated through tissue staining that pT146 stains differentially on breast tumor tissues of different grade and subtype. The potential of H1 pT146 as a therapeutic dose monitor was also assessed; experimental results showed that pT146 level changes when cells are exposed to external stimuli such as estrogen stimulus and therapeutic
LY294002. These two examples demonstrate the value of using proteomics and post translational modifications for clinical use as tumor markers.

Several proteins associated with cellular processes have proved useful in the assessment of suspected tumors. A protein that has been accepted in the medical field as a marker of proliferation is Ki-67, which is present in all cells progressing through the cell cycle but absent in resting cells. The prognostic value of Ki-67 in breast cancer was extensively assessed in a clinical outcome retrospective study by Inwald et al. They concluded that Ki-67 staining is of high value to prognosis, as heavier staining samples correlated with more aggressive tumors. It was also shown that tumors that stained higher for Ki-67 showed poorer disease free survival and overall survival when compared to cases which stained lower. While Ki-67 is useful in marking actively proliferating cells it is less useful in distinguishing the exact cell cycle phase, thus there is room for a marker that distinguishes cell cycle stages more specifically.

The search for markers which can accurately distinguish between normal and tumor cells remains a significant portion of total proteomic research. Proteins make attractive disease markers because that they are actively expressed in cells. Since immunoassays are well established in the clinical setting, assessing a patient sample for a marker is relatively simple. The real challenge for advancing diagnosis assays is pushing potential biomarkers discovered at the bench through the FDA approval process for clinical use.

*Establishing Therapeutic Targets Through Proteomics*
The FDA approval of drugs is a lengthy process, similar to the biomarker approval process. In addition to undergoing several tests of validation, prior to clinical use drugs must undergo lengthy clinical trials. There are three phases of clinical trials in addition to the preclinical research phase. The research at the bench that leads to a potential therapeutic target is considered the preclinical phase. Phase I of clinical trials involves patient participation to assess for safety of the proposed drug; information such as dosage is collected to assess for potential side effects. In Phase II, the therapy is tested for how well it works and is compared to current established treatment. This process takes up to two years to complete. Phase III involves a larger group of patient and the drug is compared against other treatments for a specific disease. From 2000-2009, the average annual FDA approval rate for drugs 4.7, however, in recent years the approval rate has increased to 8 and 10 for the years of 2011 and 2012 respectively. As a result, few drugs make it from bench to bedside.\textsuperscript{53} Despite the low rates of approval, proteomic based drugs such as protein kinase inhibitors and monoclonal antibodies. There are several classes of anticancer drugs that target proteins, some of which will be discussed in this section.

In previous sections, the importance of epigenetic regulation was discussed and how dysregulation can lead to cancer. In the past decade, several therapeutic agents have been developed to target PTM irregularities in an attempt to restore cellular function. Previous studies have shown that during cancer, histone H4 loses acetylation and trimethylation PTMs.\textsuperscript{54} Further studies have shown that this loss of acetylation may be associated with overactive histone deacetylases (HDAC), including prostate, lung, colorectal, breast, liver and gastric (reviewed by West et al.).\textsuperscript{55} Hence, the reasoning to use
HDAC inhibitors as treatment for cancer in order to restore normal PTM states. Two FDA approved HDAC inhibitors, vorinostat (suberoylanilide hydroxamic acid, SAHA) and romidepsin, have shown great success against T cell lymphomas.\(^{56-59}\) While other HDAC inhibitors exist, none have demonstrated significant improvement over these two established treatments, thus not many others have been approved by the FDA. Studies for use of HDAC inhibitors against additional tumor types, including solid tumors, are ongoing. For example, prostate cancer is indicated as a good cancer candidate for HDAC inhibitor treatment due to the high expression levels of HDACs in castrate resistant prostate cancer, however, several clinical trials have indicated HDACi’s, including SB939, is not sufficient alone but may be usable as part of a chemotherapy regiment.\(^{60}\) For testicular cancer, tests have been conducted using 5-aza-2’deoxyctydine (5-aza-dC) as a potential treatment agent.\(^{61}\) 5-aza-dC is a chromatin modifying drug that is gene specific with the ability to restore tumor suppressor (RASSF1A) function and inhibit oncogene (POU5F1).

Monoclonal antibody tumor therapies have received great attention due to their specificity as they are designed to target proteins that are specifically expressed on the tumor cell surface. Antibody mediated therapies can function in three ways: interaction with cell surface receptors in either agonistic or antagonistic functions, activating the immune system to destroy the tumor cell or carry a conjugated treatment agent to the targeted cell.\(^{62}\) The delivered agent may be a chemotherapeutic agent or a radioactive particle for targeted radiation therapy (radioimmunotherapy). Notable examples of monoclonal antibody cancer treatments include bevacizumab (Avastin) and cetuximab (Erbitux) which have been successfully used to great colorectal, breast and lung cancer...
patients by targeting vascular endothelial growth factors and epidermal growth factors respectively. Thus, continued proteomic research is essential in identifying these markers to improve specificity and expand the applicability of this treatment strategy to additional cancer types.

Protein kinase tumor therapies were developed on the rationale that several cellular processes involved in cancer: growth, survival, metastasis, invasion and angiogenesis are regulated by pathways containing these kinases. Two examples of kinase inhibitors are imatinib and trastuzumab. Imatinib (Gleevec) is used to treat chronic myeloid leukemia (CML) through occupation of the active site of the abnormally constitutively activation of the tyrosine kinase activity of abl when it becomes fused with bcr, forming the bcr-abl fusion protein. Previously, ERBB2/HER2 was discussed as a oncogene that is overexpressed in breast cancer. Trastuzumab (Herceptin) functions by binding to the ERBB2 to promote its endocytosis which prevents the downstream signaling of proliferation.

Several drugs target proteins that play a crucial role in cellular functions. Mitotic inhibitors such as taxanes (docetacel and paclitaxel) and vinca alkaloids (vincristine and vinblastine) prevent mitosis from occurring. Taxanes function by stabilizing microtubules leading to cell cycle arrest. Taxane based treatments have proven successful and is FDA approved in breast, head and neck, prostate, ovarian, gastric and non-small cell cancer. Vinca alkaloids act on tubulin preventing the formation of microtubules which are necessary for cell division. Vinblastine and vincristine are FDA approved for use with breast cancer, Kaposi sarcoma, non- Hodgekin’s lymphoma and testicular cancer and
leukemia, respectively. Another cellular process that is often targeted in chemotherapy is DNA replication. Anthracyclines, including doxorubicin, interferes with DNA replication in multiple mechanisms of action. Drugs of this class targets topoisomerase I and II preventing the separation of DNA strands for replication. In addition, anthracyclines also intercalate into the DNA strand further interfering with DNA replication and RNA transcription. Anthracyclines, however, affect cancer and non cancer cells alike requiring its judicial use only after other treatment options have been exhausted. Other topoisomerase inhibitor drugs have also proven very successful in the treatment of a variety of cancers.

It is important to keep in mind that the goal of cancer therapeutics is to destroy the abnormal cancer cells while sparing healthy cells. Thus, specificity of drug targeting is essential to minimize side effects through off target effects. Tumor cells are also greatly heterogeneous, even within a single tumor multiple different mutations may have accumulated due to loss of DNA repair and cell cycle regulators. As a result, it often is necessary to treat cancer from multiple angles through combination chemotherapy. There exists a constant need to find additional targets for the development of new therapeutic agents in an effort increase the effectiveness of existing treatments against cells with different mutations promoting increased proliferation and survival. Proteomics is an excellent candidate for this investigation since proteins are the effecters in many of these abnormal cellular processes, which if inhibited or corrected can lead to successful disease treatment.

*The Future of Proteomics in Cancer Research*
Like many other fields of science there is an increase on the emphasis on collaboration. In biomedical science, there has been increased use in terminology such as “clinical translational medicine” and “bench to bedside.” This terminology refers to the interface between the researcher and clinician, between science and clinical medicine.\textsuperscript{70} The necessity of translation is emphasized in the Journal of the American Medical Association, “effective translation of the new knowledge, mechanisms, and techniques generated by advances in basic science research into new approaches for prevention, diagnosis, and treatment of disease is essential for improving health.”\textsuperscript{71} As the goals of proteomic cancer research is to search for protein based biomarkers and therapeutic targets this trend is highly applicable.

At the turn of the century, biomedical research saw great advancements due to new technologies and funding. Dr. Elias Zerhouni, then director of the National Institute of Health (NIH) proposed a plan to move these biological science advances to applications in the clinical setting. This initiative became known as the NIH Roadmap for Medical Research.\textsuperscript{72-74} This initiative consisted of three main themes. The first theme is titled New Pathways to Discovery, where new innovations will aid biological researchers in understanding complex biological systems. Examples of this theme include improvements to data collection through new imaging technologies, reagents and standards as well as improved bioinformatics databases. The second theme is Research Teams of the Future which aims to set up interdisciplinary teams consisting of researchers such as engineers, physical scientists, mathematicians and computer scientists in addition to the traditional teams of biologists. This theme also supports the funding of high risk / high reward studies.
The final theme of the NIH Roadmap to Medical Research is Reengineering the Clinical Research Enterprise. This approach will restructure the research organization to create an infrastructure that will facilitate clinical translational research. For example, new partnerships between several academic centers and communities of physicians and patients will expand existing studies centered at a single academic center allowing for large scale clinical trials with a larger patient pool. Although a decade old, many of these research organization ideas are still visible in the biomedical research today.

This increased communication between researchers and clinicians should be beneficial for all parties. Researchers will have increased access to research samples. Previously, a challenge for translation of biomarker candidates to clinical practice is the lack of understanding researchers have for analytical, diagnostic and regulatory requirements of a clinical assay. Some of the existing barriers for clinicians include funding, recruiting research staff and participants, preparing reports and agreements, and maintaining a balance between clinical responsibilities and research. These barriers for clinicians interested in conducting research may be alleviated through partnership with established researchers. Additionally, dialogue between various parties promote the development of stronger experimental methods and setting goals of discovery that not only satisfy scientific curiosity but targets a specific need in the medical practice community. Through elimination of barriers by promoting increased collaboration between those who study basic science and those who treat the diseases, the distance between bench and bedside can be greatly shortened.
Conclusion

With the advancement in research techniques and understanding of human biology, scientific research in cancer has made large advances in recent years. High throughput analysis in proteomics has generated such large amounts of data, the need for bioinformatics and statistical data processing advancements have become necessary to make sense of all that is collected. Further advancements have allowed for the direct study of patient tumors without the need simulate the disease through cell lines and mouse models, improving the success rate of translating bench based studies to tools applicable to the clinical bedside. In the previous sections, a few successes of proteomics in discovery of new biomarkers leading to development of clinical diagnostic assays and therapeutic targets have been mentioned. These studies bring further focus to the importance of understanding proteomics in the pursuit of improving cancer outcomes.

It is important to remember the big picture of human disease biology when considering cancer research. The genomics, transcriptomics and proteomics all play a role in the dysregulation of cellular homeostasis leading to cancer. Along that same line of thought each of these processes offer check points which can be potentially targeted for therapeutics. Thus, it is critical that these processes be studied in parallel to further understanding their delicate relationship in maintaining the state of health. While the goal for the cure for cancer is somewhat beyond reach today due to the immense heterogeneity of the disease, proteomic research brings us one step closer as we improve diagnosis and treatment success rates to continue the optimistic trend of lowering cancer mortality.
In Chapter 2, a novel method for the collection of sectioned formalin fixed and paraffin embedded (FFPE) tissue for SuperFrost Plus slides. Chapter 3 describes the first global proteomic study of the thymus and thymoma subtype leading to the finding of desmoyokin expression is unique to the medulla but not the cortex of the thymus. This protein holds potential as a marker to objectively distinguish thymomas derived from the medulla (A) from those of cortex origin (B3) in case that are difficult to distinguish through histology alone. Finally, Chapter 4 of this dissertation discusses the use of yeast histone mutants to investigate the effect of histone post-translational modification on the global proteome.
Chapter 2: Anionic Detergent Mediated On-Slide Digestion: A Novel Approach to Process Archived Tissue for Bottom-Up Proteomics

Background

The use of patient tissue for proteomics research has become increasingly popular due to numerous advantages in the translation of discoveries into clinical practice. Large repositories of patient tissues collected over decades from patient biopsies and surgical resections are stored and preserved through formalin fixation and paraffin embedding (FFPE). Thus, to utilize these tissues effectively there is a need for the development of techniques in sample processing for proteomic study.

Introduction

Currently in the United States, 14.5 million Americans have a history of cancer. This year 1.6 million new cases will be diagnosed and over half a million people will die. Cancer is the second highest cause of mortality in the US behind heart disease and accounts for one in four deaths. For many forms of cancer, the key to successful treatment and optimal outcome is early diagnosis and treatment. However, studies show that up to 60% of cancers are not diagnosed until after metastasis has occurred, greatly complicating therapy. In the past decade, advances in high throughput technologies to allow for assessment of the DNA, RNA and proteins of whole cells in a single experiment have contributed greatly to new
diagnostic and treatment advancements in the –omics era. Proteomics play an important role in the clinical research setting in two ways: biomarker and therapeutic target discovery. Since proteins serve as the effectors of cellular function, a protein based assay to detect abnormal changes in protein structure or quantity may be used as an indicator of cellular dysfunction associated with disease.

There are several advantages to the use of patient tissue over other model systems. Cell culture of established cell lines has been a prominent source of bio specimens for research. However, there is debate as to how well a cell line represents human disease. Cell lines lack contact and signaling from neighboring stromal cells, which in a disease context can provide soluble factors, extracellular matrix molecules and cell-cell communications that can lead to differences in gene expression. Animal models are available to simulate various cancers in a mammalian system. However, these studies are limited in that the cells are non-human. Alternatively, xenografts involving the passage of human tumor cells into generations of mice have been used as a source of disease cells. However, this technique requires animals and is costly.

Patient bio specimens are the preferred source of material for cancer research for many reasons. First, human disease is heterogeneous. Molecular profiles from the cells of two patients with the same diagnosis are likely to show important differences. Thus, through use of samples from several different patients, researchers can account for heterogeneity. Patient tissue taken from its native environment includes all the associated factors such as cell-cell signaling, extracellular matrix and systemic signaling molecules transported by the bloodstream. Finally, patient tissue can be archived and stored for
decades. This allows for the collection of rare disease specimens over time. Also, patient tissue with longer history allows for longitudinal information to be collected detailing treatment outcome and prognosis.81

Freezing and Formalin Fixation & Paraffin Embedding (FFPE) are two predominant methods for tissue preservation and storage. Initially it was suggested that frozen tissue produces the best sample quality downstream for DNA, mRNA and proteins, however, with proper processing and advancements in macromolecule research techniques FFPE tissues studies have become more widespread.82,83 Tissue freezing is costly in space and energy storage, as they must be kept at -80º C and can only be stored for one year.84 Tissue freezing leads to loss of histological features, introducing challenges in the identification of areas of interest. Furthermore, patient tissue is typically not frozen unless experimental usage of the tissue is predetermined. The study described in this document focuses on application of proteomic analysis to the more common FFPE tissues.85-89 FFPE storage involves the fixation of tissue in formalin (4% formaldehyde) then embedding in a paraffin block.90 FFPE tissue storage preserves tissue histology allowing for identification of cellular boundaries.91,92 In histology labs, tissues are sectioned and mounted on slides that maintain adhesion of the tissue section during immunohistochemical workup. SuperFrost Plus slides are used for histology to provide excellent adhesion between the slide and tissue sections through a permanent positive charge without the use of chemical adhesives. Strong adhesion, advantageous for immunohistochemistry (IHC), hinders removal of tissues for downstream molecular analysis. A method of direct sample preparation from histological slides would greatly facilitate downstream molecular
analysis of archived tissues. The batch mounting of tissue specimens onto SuperFrost Plus slides is common practice in histology labs. The optimization of workflows that can use these batch produced slides reduces the amount of unnecessary sectioning of tissues blocks. The net effect is reduction in labor and waste. The lack of chemical adherent makes the SuperFrost Plus slide well suited downstream proteomic analysis.

Methods

Patient Tissue Collection

Patient tissue used for this study was formalin fixed and paraffin embedded. Tissues mounted on polyethylene napthalate (PEN) slides (Zeiss) were cut at 10 microns thickness. Tissues mounted on SuperFrost Plus slides were cut at 4 microns thickness. Staining studies were performed using reactive lymph node tissue mounted on PEN slides for laser capture microdissection of reactive lymph node regions. Tissue collection studies were performed using thymoma subtypes A and B3, lung squamous cell carcinoma and lymph node tissue mounted on both PEN slides and SuperFrost Plus slides. The use of patient tissue was approved by The Ohio State University Institutional Review Board (IRB) (approval on file).

Deparaffination and Staining

Tissues were deparaffinized with octane then rehydrated with an ethanol gradient (100%/90%/70%). Tissues were stained for 5 seconds followed immediately by a water wash. Single staining was performed using hematoxylin (Vector) and toluidine blue (Fisher
Scientific). The hematoxylin staining was tested at different concentrations of 5%, 10%, 15% and 20%. Toluidine blue was prepared as a stock solution using 2% (w/v) toluidine blue O mixed in 70% ethanol. The working solution was prepared by mixing the stock solution in 1% (w/v) sodium chloride at a 1:9 ratio. This process was repeated for double staining. Following staining the tissues were dehydrated with an ethanol gradient (70%/95%/100%).

**Laser Capture Microdissection**

Tissue collection for the staining experiment was performed using laser capture microdissection. Reactive lymph node tissue germinal center, mantle and interfollicular space were collected on a PALM Microbeam IV, release 4.2 (Zeiss) and pressure catapulted into digestion buffer consisting of 0.5% (w/v) RapiGest (Waters Corporation) in 50 mM ammonium bicarbonate. Laser energy was set at 51 and focus at 81. The goal for LCM usage was to collect a minimum of 20,000 cells from each of the aforementioned regions. Due to the small amount of mantle tissue on each slide serial sections were used to reach collection goals.

**On-Slide Digestion Sample Preparation**

The experimental schematic for the tissue collection methods are summarized in Figure 1. The addition of trypsin and digestion buffer along with detergent directly to the slide minimizes the number of sample processing steps resulting in improved analyte collection and minimal contamination. Blocks from three different B3 thymoma were cut and
mounted onto SuperFrost Plus slides in quadruplicate. Following deparaffination and manual removal of tissues of non-interest (by scraping), the slides were placed in lock mailer tubes and boiled for 20 minutes then heated at 60º C for two hours. On-slide digestion involved the direct deposition of 20 µL digestion buffer (0.5% w/v RapiGest) with trypsin (0.01% w/v). A coverslip was added to seal the digestion buffer and prevent evaporation. Slides were heated at 37º C for 18 hours with digestion buffer. Peptides were collected by rinsing the slide and coverslip with 50 µL of 50 mM ammonium bicarbonate. RapiGest and trypsin were inactivated through addition of formic acid lowering the pH to ~2 and the final peptide sample was dried and resuspended in 50 microliters of HPLC water.
Figure 1. Tissue Collection Methods Experimental Schematic. A) Collection of tissue from PEN slides. B) Mechanical collection of patient tissue from SuperFrost Plus slides using a dissecting instrument. C) On slide digestion of patient tissue.

Mechanical and PEN Collection Comparative Studies

The proteomic data obtained by on slide digestion was compared with data from in solution digestion of tissue removed from SuperFrost Slides by scraping and macrodissection from
PEN slides. Serial sections on SuperFrost and PEN slides from the B3 thymoma blocks were used to maximize tissue homogeneity for each comparison. For the PEN Slide, an incision surrounding the region of interest was made in the membrane allowing the desired tissue to be lifted from the slide and transferred to an Eppendorf tube for in solution digestion. Tissue placed on SuperFrost Plus slides were collected by mechanical removal after deparafinization. For mechanical removal, tissue surrounding the area of interest was removed using a scalpel. The slide was then washed in ethanol and the tissue of interest was collected. Formaldehyde induced cross links were reversed through heating and sample digestion with RapiGest and trypsin using the parameters described above in the on-slide digestion methods.

**Liquid Chromatography Mass Spectrometry**

Sample separation was performed using a Dionex 3000 UltiMate system using a C18 reverse phase column (Microm Bioresources Magic C18AQ, 200 µm x 150 mm, 200 Å). The mobile phases were water with 0.1% formic acid and acetonitrile with 0.1% formic acid. The HPLC gradient consisted of an increasing percentage of the acetonitrile mobile phase over 5 hours from a starting 2% to a final of 35%. Mass spectrometry data collection was performed using a Thermo Scientific LTQ Orbitrap XL mass spectrometer. Data analysis for protein identification and spectral count analysis was performed using an in-house search engine, MassMatrix, with a Swiss-Prot/TrEMBL human protein database (UniProt 3/30/2012).93-95
Assessment of Differences in Proteomic Data as a Function of the Quantity of Input Slides

The minimum number of patient tissue slides required for mass spectrometry proteomics was assessed using tissue collected from non-diseased thymus serial sections cut at 4 microns in thickness and placed on SuperFrost Plus slides. Tissue collection and removal were performed using on-slide digestion. LC-MS/MS was performed on samples processed using 1, 2 and 4 serial sectioned slides.

Detergent Removal of Tissue from Plus Slides

The ability of different detergents to effectively remove patient tissue from SuperFrost Plus slides was assessed using 2 anionic detergents (RapiGest and sodium dodecyl sulfate) and 2 nonionic detergents (NP40 and Triton X). The detergents were placed on slides then incubated for 18 hours at 37º C. The slides were then rinsed with water and assessed for tissue removal.

Tissue Collection for SuperFrost Plus Slides for Western Blotting

FFPE thymus tissue used for this study consisted of 3 serial sections cut at 4 µm thickness. Tissues were deparaffinized with octane and ethanol. Following deparaffinization, tissues were stained using hematoxylin followed by toluidine blue as previously described. Connective tissue and surrounding non-tumorous tissue was removed by dissection for cancer tissue enrichment. 50 mM ammonium bicarbonate was placed on the remaining tissue then sealed with a coverslip in preparation for formaldehyde crosslink reversal. The slide was heated at boiling temperatures for 20 minutes, followed by incubation at 60º C.
for two hours. 30 µL of 10% sodium dodecyl sulfate (SDS) was deposited on the tissue and the coverslip was reapplied. The slides were then incubated at 37º C for 18 hours. Tissue was then collected through 3 rinses of 50 µL of 50 mM ammonium bicarbonate for the mass spectrometry sample preparation. To ensure complete crosslink reversal necessary for Western blotting, a second round of crosslink reversal was performed. The collected sample pH was increased to 9 using 100 mM ammonium bicarbonate and then boiled for 20 minutes followed by incubation at 60º C. The sample was then dried using a Thermo SPD1010 speed vacuum concentrator. The result was mixed with 50 µL BioRad 2x Laemmli buffer and run in a 15% SDS PAGE gel. The proteins were transferred to nitrocellulose membrane via semi-dry transfer. The primary antibody used was mouse monoclonal antibody to histone H4 (AbCam ab31830).

**Results and Discussion**

*Tissue Staining with Hematoxylin and Toluidine Blue*

Due to the heterogeneous nature of human tissue, histological visualization of tissue and structures is required to separate tumor cells from adjacent tissue. The pathology gold standard for visualization is the use of hematoxylin and eosin staining. Hematoxylin is a naturally occurring stain that is oxidized into hematein and mixed with metal ions that carry a positive charge. As a result, it stains negatively charged molecules within cells, such as nucleic acids, with a purple blue color. Eosin is a synthetic stain targeting positively charged structures within cells such as lysine and arginine residues found in proteins. The binding of eosin to proteins leads to an alteration of mass, interfering with the mass
spectrometry proteomic process. Since coverslips are not used on slides for mass spectrometry proteomics, the issue is further complicated by darkening of stained tissue, leading to further loss of histological resolution. Historically, groups have used hematoxylin alone to stain tissues for mass spectrometry proteomics studies. However, in certain tissues the hematoxylin alone is insufficient in discerning the tissue structures, as demonstrated in reactive lymph nodes. Studies have shown that with a toluidine blue double stain, lymph node structures are distinguishable (Figure 2). Similar to hematoxylin, toluidine blue is positively charged and thus binds to the nucleic acid in cells. The combination of hematoxylin and toluidine blue was evaluated for its mass spectrometry compatibility in order to stain tissues that are difficult to visualize by hematoxylin alone.
Figure 2. Tissue Staining of Reactive Lymph Nodes. A) Hematoxylin and eosin stain of reactive lymph node tissue. H&E stains are the gold standard for staining in histology. B) Hematoxylin single stain. Note the lack of visibility of the tissue structures. C) Toluidine blue single stain. While the germinal centers can be visualized, the mantle region remains unclear. D) Hematoxylin and toluidine blue double stain. The germinal center, mantle and interfollicular space are visible using this staining method.
Tissue staining with hematoxylin alone was found to be insufficient to visualize different structures found within reactive lymph nodes (Figure 2a and b). Using hematoxylin at increasing concentrations of 5%, 10%, 15% and 20% improved visualization of tissue structures, but further improvement was necessary to visualize specific tissue structures (Figure 3). Use of toluidine blue allowed for visualization of the germinal centers (Figure 2c). When the tissue was stained with hematoxylin then counter stained with toluidine blue all structures of the reactive lymph node could be discerned (Figure 2d).

**Figure 3.** Hematoxylin Staining of Reactive Lymph with 20%, 15%, 10% and 5% Dilutions. Serial sections of reactive lymph nodes were stained using 20%, 15%, 10% and 5% dilutions of hematoxylin to optimize hematoxylin staining in order to reduce the darkening effect of stained tissue without coverslip for tissue dissection. A) Hematoxylin and eosin stain control. B) 20% hematoxylin stain. C) 15% hematoxylin stain. D) 10% hematoxylin stain. E) 5% hematoxylin stain.
The compatibility of staining methods with mass spectrometry based proteomic analysis was then compared. Data analysis of mass spectrometry results demonstrated that there was a high degree of correlation in protein identification between the three staining methods: hematoxylin, toluidine blue and double staining using hematoxylin and toluidine blue (Figure 4). A total of 398 protein groups were identified between the three staining methods, 347 protein groups (87%) were shared amongst the three methods. The protein spectral counts showed a high degree of correlation between the different staining methods (Figure 5). Some differences in protein IDs may be explained by the dark staining of hematoxylin alone leading to inadvertent collection of non-germinal center tissue due to lack of discernable borders. Consistent with this explanation, log-log plots of the spectral counts of the identified proteins also demonstrated a high degree of linear correlation between the toluidine blue single stain and the combined hematoxylin / toluidine blue stain (Figure 4B). The conclusion is that the number of protein identifications and spectral counts were not significantly affected by addition of toluidine blue.
Figure 4. Protein Identification and Spectral Count Quantitation in Reactive Lymph Nodes are Independent of Staining Method. A) The Venn diagram comparing the proteins identified between the three different staining methods: hematoxylin single stain, toluidine blue single stain and hematoxylin and toluidine blue double stain shows a large degree of overlap. This indicates that the proteins detected are consistent across each of the sample staining methods. B) The log-log plots of the spectral counts compare the quantification of proteins between the samples. The linear relationship indicates that the spectral counts between samples are comparable.
Figure 5. Spectral Count Correlation of Lymphatic Tissue Components and Tissue Staining. Plotting the spectral counts between two data samples show a high degree of correlation between the components of lymphatic tissue (germinal center, interfollicular space and mantle) and the stains used (hematoxylin, toluidine and hematoxylin/toluidine double stain) indicating relatively little difference in results from stain used (Top Right). The correlation matrices also show a high degree of correlation between the different data sets with the majority showing correlation values of 95 (Bottom Left).

Principle component analysis (PCA) was performed on the nine data sets from three lymph node tissue areas (germinal center, interfollicular space and mantle) stained with the
three different methods: hematoxylin, hematoxylin/toluidine blue and toluidine blue. The biplot of the largest components (Figure 6) illustrated a high degree of correlation between the staining methods for a given cell type with the exception of hematoxylin germina. Both Pearson and Spearman coefficients indicated positive correlations between the different techniques used indicating low effects from the staining in the data. Pearson coefficients indicated a strong positive linear relationship between the different staining methods used to distinguish between the three regions of the reactive lymph nodes. It was notable that the Pearson coefficients associated with hematoxylin staining of germinal centers in comparison to other data sets was lower than other associations. This difference was possibly due to challenges in discerning the germinal center borders when using hematoxylin staining leading to differences in the tissue collected. The Spearman coefficient also indicated a positive correlation with coefficients associated with hematoxylin germinal center stains lower than the other pairings as observed with the Pearson correlations.

Tissue Collection using PEN and SuperFrost Plus Slides
Having established that toluidine blue can be used as a counter stain for hematoxylin, we next evaluated tissue preparation methodology. The automatic sectioning of FFPE tissue onto adherent SuperFrost Plus slides is a common practice for histological characterization. Unlike slides with non-adherent membranes, tissue on SuperFrost Plus slides must be scraped free of the positively charged surface. To minimize loss of tissues due to scraping we evaluated a method that would allow for tissue digestion directly on the slide. On-slide digestion of tissues would eliminate the need for tissue sectioning onto nonadherent slides or the need for scraping, reducing the risk of contamination.

Lymph node, lung squamous cell carcinoma and thymoma subtype B3 (3 cases) and subtype A were sectioned onto SuperFrost Plus slides. Hematoxylin and eosin stained tissue were used for references to define areas of interest for collection. In this study, the use of different tissue collection techniques was compared. As described in the methods section, tissues on SuperFrost Plus slides were collected either by mechanical removal using a dissection tool or directly digested into peptides on-slide through direct application of digestion buffer and trypsin. Nonadherent PEN slides were used as a comparison as they are used for the well-established laser capture microdissection methods of tissue collection. Tissues were collected from PEN slides through scalpel dissection surrounding the tissue of interest and removal of tissue along with the associated PEN membrane.

Correlation of the spectral count data between preparation types is shown in Figure 7. Linear correlation of the log-log plots demonstrates that protein quantification was not adversely affected by the different tissue collection methods (Figure 7A). Combining the proteomic data from the 3 thymoma B3 cases showed 86% overlap in protein IDs across
all three tissue collection methods (Figure 7B). A total of 1,014 proteins (875 protein groups) were identified in all three sample collection methods. These data show that tissue digested on the surface of common SuperFrost Plus slides is compatible with proteomic analysis by MS analysis and compares well with other sample collection methods. The total tissue combined data showed that 92% of the proteins identified in all samples were present in all three tissue collection methods. In all tissue samples analyzed a total of 1,043 protein groups were identified of which 963 were common in all three sample preparation methods. The results suggested that proteomic analysis of tissue digested on the surface of common SuperFrost Plus slides was compatible with MS analysis, eliminating the need for scraping or non adherent slides.
Figure 7. Protein Identification and Spectral Count Quantitation in Reactive Lymph Nodes are Independent of Tissue Collection Method. A) The log-log plots of the spectral counts compare the quantitation of proteins between the samples. Top: Spectral count comparisons between the three different tissue collection methods using cumulative data from three thymoma B3 specimens. Bottom: Spectral count comparisons between the collection methods using cumulative data from thymoma B3 and A, lung squamous cell carcinoma and lymph node tissue. The linear relationship indicates that the spectral counts between samples are comparable. B) Venn diagram comparing the proteins identified between the three different tissue collection methods: PEN slide, mechanical collection and on-slide digestion show large overlap, indicating that the proteins detected are consistent across each of the sample collection methods.
Principle Component Analysis of tissue collection methods was performed using the R statistical tool for data sets of three thymoma B3 cases (biological replicates) denoted A, B and C. These cases were processed using the three tissue collection methods (PEN slides, on-slide digestion and mechanical tissue collection). The principle components did not show noticeable patterns leading to segregation by either sample or collection method. A high degree of correlation in the data was observed between the tissue collection methods used. Pearson correlation of mechanical tissue collection between samples A, B and C demonstrated strong correlation coefficient values of 0.8475, 0.8413 and 0.8423 for AvB, AvC and BvC respectively. On-slide sample collection also showed a high degree of correlation with Pearson coefficient values of 0.7872, 0.8231 and 0.7906 for AvB, AvC and BvC respectively. Data from the PEN slides also showed very strong correlations with Pearson values of 0.8405, 0.8800 and 0.9419 for AvB, AvC and BvC respectively. Pearson coefficients also showed high correlation between the samples that were collected using different methods: A (0.9076 mech v on-slide, 0.8239 mech v PEN and 0.7239 on-slide v PEN), B (0.8272, mech v on-slide, 0.9199 mech v PEN and 0.7629 on-slide v PEN) and C (0.7754 mech v on-slide, 0.8252 mech v PEN and 0.8652 on-slide v PEN) Taken together, the strong correlations suggest the different tissue collection methods are comparable without significant sources of bias.

Proteomic analysis of normal lymph node tissue was performed on samples generated by combining 1, 2 or 4 serial sections to assess the effect of sample size on proteomic depth. Correlation of the spectral count data between the different quantities of tissue input was performed (Figure 8A). Linear correlation of the log-log plots
demonstrated that quantification by spectral counting in mass spectrometry sample analysis showed correlation across the different number of combined tissue sections. As expected, protein depth improved as the amount of the number of tissue sections increased (Figure 8B). For FFPE lymph node tissue cut at 4 µm thickness mounted on SuperFrost Plus slides the numbers of unique proteins identified were 763, 787 and 852 for 1, 2 and 4 serial sections respectively. These data showed that increasing the number of serial slides of tissue input increased the protein identification.
Figure 8. Impact of Material Input on On-Slide Digestion. On-slide digestion was performed on serial sections of non-diseased lymph nodes. A) The log-log plots of the spectral counts compare the quantification of proteins between the samples. The linear relationship indicates that the spectral counts between samples are comparable. B) Venn diagram comparing the number of proteins identified for each different quantity of serial sections on SuperFrost Plus slides inputted.
Importance of Detergent for On-Slide Tissue Digestion

We propose that on-slide tissue preparation approach is facilitated using of an anionic detergent. The anionic detergent interacts with the positively charged slide surface, disrupting adhesion and allowing for tissue release and digestion. To test this hypothesis, we examined on-slide tissue digestion and release with two anionic detergents (Rapigest and SDS) and two nonionic detergents (NP40 and Triton X). Serial sections of thymus tissue were cut at 4 micron thickness and placed on SuperFrost Plus slides. The detergents and digestion buffers were added to separate slides and heated for 18 hours at 37º C. The slides were then rinsed with water and surveyed for extent of tissue digestion and release. The slides exposed to anionic detergents in the digestion buffer facilitated nearly complete displacement of tissue from the SuperFrost Plus slides while nonionic detergents lacked the ability to displace the tissue from slides (Figure 9).
Figure 9. Tissue Collection from SuperFrost Plus Slides is Mediated by Anionic Detergents. Incubation of patient tissue in anionic detergents (RapiGest and sodium dodecyl sulfate) and nonionic detergents (Triton X and NP40) followed by washes show that the tissue release from Plus slides only occurs in the presence of anionic detergents. This supports the hypothesis that the anionic detergents binds to the positively charged SuperFrost Plus slides mediating tissue release for collection.

Tissue Collection for SuperFrost Plus Slides for Western Blotting

The majority of studies involving FFPE patient tissue have been immunohistochemistry based. While they are useful in identifying the localization of the protein of interest, IHC experiments provide limited information about the quantity of that protein. While the use of FFPE tissue in Western blotting has been established, here we demonstrate the ability to combine the enrichment of tissue of interest through microscopy and dissection with deeper and more quantitative exploration of FFPE tissue through proteomics. In this example, we examined whether the anionic detergents would also enhance the removal of tissues
from slides for subsequent Western blots. The 10% SDS used in tissue collection has the added role of performing cell lysis and protein solubilization. Complete crosslink reversal is critical for Western blotting so two rounds of crosslink reversal were performed (one prior to tissue removal from the SuperFrost Plus slide and one immediately following tissue collection). While further studies are necessary using additional protein targets and controls, current evidence suggests that the collected specimen from the on-slide method may be suitable for Western blotting (Figure 10). The success of this application would be beneficial to the research field when IHC grade antibodies are not available for novel protein targets and validating results from LC-MS/MS studies.
Figure 10. Histone 4 Western Blot of Collected Patient Tissue Samples. Two technical replicates of FFPE patient tissue was collected from slides containing serial sections using the methods described in this manuscript. Western blotting using anti-Histone H4 antibody detects one band at the expected 14 kDa. This indicates the successful collection and reproducibility of tissue from the slides and crosslink reversal. This experiment suggests the possibility of applying our developed methodology to Western blotting.

Conclusion

This study clearly demonstrates that archived tissues sections on standard pathology SuperFrost Plus slides can be easily processed for bottom up proteomics experiments. Our methods allow for time and resource efficient assessment of patient tissue specimens for hypothesis generation in the search for novel biomarkers. Combined with a powerful high
throughput technique such as bottom up mass spectrometry proteomics a number of samples can be analyzed to generate a candidate list of potential targets of interest for further analysis. Improvements in archived tissue preparation methods increases options for translational research that used archived tissues in molecular pathology workflows.

Acknowledgements

The project described was supported by Award Number Grant TL1TR001069 from the National Center For Advancing Translational Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center For Advancing Translational Sciences or the National Institutes of Health.
Chapter 3: Proteomic Signatures of Thymomas

Background

The thymus is an integral part of the adaptive immune system, playing a major role in the production and development of lymphocytes starting at conception and continuing well into childhood. Thymoma is a rare thoracic epithelial tumor of the anterior mediastinum that occurs at a rate of 0.15 per 100,000 people (National Cancer Institute). Thymoma research is also complicated by the lack of tumor cell lines and animal models. In order to identify biomarkers associated with thymoma types an archival set of formalin fixed paraffin embedded (FFPE) patient tissues was used to profile protein expression differences by the use of liquid chromatography tandem mass spectrometry (LC-MS/MS) based proteomics.

The World Health Organization (WHO) devised classification system for thymoma is summarized in Table 1. This classification system is based on cell histology and lymphocyte to epithelial cell ratio. This system divides thymomas based on cellular morphology in which those with oval or spindle shaped cells are designated as type A and those with plump or dendritic cells are designated as type B. Any thymic carcinoma is classified as type C. The thymus consists of two regions, the central medulla and the lymphocyte rich cortex on the periphery. Type A thymomas originate from the medulla and type B thymomas originate from the cortex. Studies have shown significant differences
between survival and recurrence rates of the different subtypes, suggesting that the current WHO classification is useful in clinical practice. Subtypes AB, B1 and B2 have a large non-neoplastic lymphocyte presence that makes them easy to distinguish from subtypes A and B3 by histology. However, differentiating between type A and type B3 thymomas by histology is difficult in some cases because of the lack of T lymphocyte infiltration resulting in dependence of pathologist observation of overlapping cell shape and features. Thymomas A and B3 share similar immunohistochemical (IHC) staining patterns for panCK (cytokeratin), p63, EMA (epithelial membrane antigen) and TdT. Thymoma subtypes A and B3 represent the opposite extremes of 10 year survival at 95-100% vs. 36-40% respectively. B3 thymomas, have a higher metastatic potential and patients may benefit from radiation and/or chemotherapy in addition to thymectomy. For these reasons, a biomarker with the ability to distinguish between type A and type B3 thymomas will be clinically valuable in providing a supplementary objective diagnosis. Through the improvement of thymoma subtype diagnosis, an appropriate treatment plan is devised for the specific disease allowing for better patient outcome.
### Thymoma Subtype % of Total Thymoma 10 Year Survival Rate Morphological Features

<table>
<thead>
<tr>
<th>Thymoma Subtype</th>
<th>% of Total Thymoma</th>
<th>10 Year Survival Rate</th>
<th>Morphological Features</th>
</tr>
</thead>
</table>
| A               | 4-7%               | 95-100%               | • Spindle oval shaped cells without nuclear atypia  
|                 |                    |                       | • Few or none nonneoplastic lymphocytes  
|                 |                    |                       | • Encapsulated |
| AB              | 28-34%             | 90-100%               | • Same features as type A  
|                 |                    |                       | • Rich in nonneoplastic lymphocytes |
| B1              | 9-20%              | 83-85%                | • Resembles normal functioning thymus  
|                 |                    |                       | • Overwhelmingly rich in lymphocytes |
| B2              | 20-36%             | 71-83%                | • Plump cells with vesicular nuclei  
|                 |                    |                       | • Heavy population of nonneoplastic lymphocytes |
| B3              | 10-14%             | 36-40%                | • Round or polygonal epithelial cells with mild or no atypia (sheet like structure)  
|                 |                    |                       | • Very few lymphocytes |

**Table 1.** Histological Features of the Thymomas. While the thymoma subtypes often share overlapping histological features (National Cancer Institute and Stanford Surgical Pathology), the clinical outcomes differ.

Previously, Sun et al. determined that proteins were differentially expressed in thymomas B1 and B2 using homogenized resection specimens.\textsuperscript{105} Strobel et al. previously assessed the use of medullary and cortical markers to characterize thymoma subtypes.\textsuperscript{106} Unfortunately, these markers were not observed in the A and B3 subtypes. Another study identified proteasome subunit β5t as means of distinguishing medulla vs. cortex derived A and B thymomas respectively.\textsuperscript{107} Our study further assessed the protein expression
differences of thymomas by profiling the proteome of FFPE archival tissue specimens of healthy thymus and thymomas A, AB, B1, B2 and B3. Protein expression differences were assessed through pairwise comparisons between normal thymus tissue and thymoma tumors of the five A/B subtypes. A protein expression profile was identified to differentiate lymphocyte-rich tumors from lymphocyte-poor. It was confirmed that the onco-protein stathmin, identifies T lymphocyte content within the tumor rather than the thymoma itself. The proteomics analysis also revealed that desmoyokin was downregulated in subtype B3 relative to subtype A. Immunohistochemistry validation confirmed that desmoyokin was found in the medulla and not the cortex. Since type A thymomas are derived from the medulla while B type are derived from the cortex, subtypes A and B3 thymomas can be differentiated using desmoyokin IHC analysis.

This study performs the first comparative global proteomic profile of the thymus and the five different thymoma subtypes. While several of the findings are supportive of what is known of the thymus and thymomas, we propose the novel use of desmoyokin to distinguish between type A and B thymomas based on the tumors’ region of origin within the thymus.

**Methods**

*Formalin Fixed Paraffin Embedded Patient Tissue Processing and Digestion*

A total of 36 formalin fixed paraffin embedded patient tissues encompassing the five tumor subtypes and normal thymus tissue were profiled by shotgun proteomics. Biological replicates included 6 de-identified specimens for each diagnosis and no technical replicates
were performed, ultimately generating a total of 36 data files. The sample size was selected in consideration of the statistical power as well as the time and cost associated with the preparation and mass spectrometry experiment. Data from the mass spectrometry experiment was then used to select protein candidates for orthogonal validation with immunohistochemistry using a tumor microarray consisting of 71 tumor cores. Due to low sample quality of one thymoma A sample the data set was rejected leaving 5 thymoma A data sets and 35 data sets total. Patient specimens were selected from cases ranging from the years 2001 and 2012. The age range of the patients was from 20 to 70 years with an average age of 54. Male to female ratio of the specimens used for mass spectrometry studies was 16:15 (data unavailable for 5 specimens). The patient data for the specimens in this study is detailed in Table 2. All experimental work in this study was approved by The Ohio State University Institutional Review Board (IRB). Four serial sections from each specimen block were collected. One section was processed using a standard hematoxylin and eosin (H&E) protocol and specific tumorigenic areas were highlighted by a thoracic pathologist. Tissues were dried for 14 days prior to processing and digestion.
Table 2. Patient Details for LC-MS/MS Experiment. Patient age and gender for tissue specimens used for LC-MS/MS proteomics experiment.

FFPE tissues were deparaffinized using octane then rehydrated using a graded ethanol series (100%/90%/70%). Briefly, tissue slides were stained using 15% hematoxylin (Vector H-3404, Burlingame, CA, USA) then counterstained with toluidine blue (Fisher Scientific, Fair Lawn, NJ, USA). The toluidine blue was prepared as a stock solution by dissolving 500 mg of toluidine blue into 50 mL 70% ethanol. Working solution consisted of toluidine blue stock solution and 1% NaCl at a 1:9 ratio. Both hematoxylin and toluidine blue stains are compatible with mass spectrometry proteomics. Excess stain was removed with water and the tissue was dehydrated with a graded ethanol series (70%/95%/100%).

Using the matching H&E stained tissue slide, the connective tissue surrounding the tumor was removed with a scalpel and residual debris washed away with ethanol. The tissue crosslinks were reversed by incubating the tissue slides in boiling water for 20 minutes followed by 2 hours at 60°C. Each recovered tissue was digested on-slide with
0.01% (w/v) Trypsin (Promega, Madison, WI, USA) dissolved in 0.5% (w/v) RapiGest (Waters Corporation, Milford, MA, USA) for 18 hours at 37ºC. The digest was collected and debris removed by centrifugation. The supernatant was concentrated by vacuum centrifugation.

**LC-MS/MS Data Collection**

Tryptic peptide concentration was estimated by 280nm measurement using a Nanodrop spectrophotometer and for each sample 1500 ng of peptides were loaded onto a μ-precolumn (PepMap100, C18, 5 µm, 100 Å, 0.3 x 50 mm) at a flow rate of 20 µL/min for four minutes. Peptide separation was completed using a Dionex UltiMate 3000 RSLCnano HPLC system equipped with an Easy-Spray PepMap C18 3µm 100Å 0.75x150 mm column (1.8kV, 275°C). A linear gradient was used at a flow rate of 0.3 µL/minute from 2-28% Phase B (Phase A: 0.1% formic acid and Phase B: 0.1% formic acid in acetonitrile) from 4-145 minutes followed by a column wash and equilibration period. The total instrument acquisition time was 180 minutes. Data-dependent acquisition on an Orbitrap Fusion was completed in top speed mode. Orbitrap full scans (400-1600 m/z) were completed every four seconds (AGC:400K ions; 120K mass resolution; 50ms max injection time, 1 microscan) followed by quadruple isolation (1.4 Da) for CID fragmentation. Fragment ion masses were measured in the linear ion trap (NCE: 35%, AGC: 100 ions; 250ms max injection time, parallelizable time was enabled). Ion fragmentation selection was limited to precursor ion charges 2-7 that exceeded 10K signal intensity and isolated based on dynamic exclusion parameters (25s ± 10ppm, repeat count:1).
LC-MS/MS Data Analysis

RAW data files were analyzed directly using MaxQuant (v.1.5.2.8). Peptide-spectrum matches (PSMs) were determined using the Andromeda search engine. Database searches were conducted against FASTA database composed of a UniProt human database (taxon:9606, 20198 entries, May 2015), common Repository of Adventitious Proteins (cRAP V1.0, 116 entries, http://www.thegpm.org/crap) (20326 total entries). MaxQuant search parameters included: peptide length >5 amino acids, maximum of two tryptic missed cleavage events, dynamic modification for methionine oxidation and protein N-terminal acetylation, precursor mass tolerance of 20 ppm and fragment mass tolerance of 0.5 Da. Peptide false discovery rate (FDR) was determined using a reverse decoy database. PSMs and protein level FDR were both set at 1%. The corresponding protein identifications contained a minimum of one razor peptide. Total ion current (TIC) normalization signal to noise ratio was set to 2. The majority protein identifier was used to represent any protein groups. Precursor spectral feature information was shared across 36 RAW files using the ‘match-between-runs’ feature (20/0.7 minute alignment/match). A minimum of two unmodified unique/razor peptides was required for MaxLFQ quantitation. These features are limited to the MaxQuant environment.

Bioinformatic analysis was completed with the Perseus software (v.1.5.2.6) (http://www.perseus-framework.org). Due to the nature of this experimental design a series of two-tailed t-tests were used to determine differential protein expression with normal thymus being used as a common control. Reverse and contamination proteins were
removed and a final protein list was produced by filtering based on a positive MaxLFQ intensity in at least three biological replicates in either cancer or normal tissue. This process was repeated five times from a common matrix of 1193 protein entries. MaxLFQ values were log$_2$ transformed. Missing values were imputed under the assumption that the missing MaxLFQ intensity values are due to their corresponding peptides being present in the digest but under the limit of detection of the mass spectrometer. Missing value imputation was mimicked for low abundance proteins by bootstrapping MaxLFQ intensity values from the normal distribution of all values. As is commonly completed in this type of label-free proteomics experiment a 1.8 standard deviation downshift and 0.3 standard deviation width was used for the random reselection of data.\textsuperscript{108-110} A two-sample Student’s t-test was used to determine statistically significant differential protein expression. To control for multiple comparisons, the p-values were adjusted using a Benjamini-Hochberg approach.\textsuperscript{111}

\textit{Gene Ontology Term Enrichment Analyses}

Biological processes (GOTERM_BP_FAT) were investigated using the Database for Annotation, Visualization and Integrated Discovery (DAVID, v.6.7).\textsuperscript{112-114} Gene ontology (GO) analyses were performed on significant UniProt identifiers for each pairwise comparison independently (q-value <0.20). The entire human genome (approximately 30000 genes total, DAVID default) was used to assess gene ontology term enrichment when analyzing the complete subtype datasets for high throughput global experiments.\textsuperscript{112}

\textit{Immunohistochemistry}
Immunohistochemistry was performed on normal, thymoma A, AB, B1/B2, B3 and lung squamous cell carcinoma. Immunohistochemistry of tumor microarray (TMA) was selected as an appropriate validation since it is a routine clinical procedure. The ability to assess biomarkers through IHC facilitates potential future clinical use with existing equipment and protocols. FFPE tissues were cut at 4µm thickness on SuperFrost Plus slides. Slides were heated to 60ºC for one hour, cooled, deparaffinized then rehydrated using xylene and graded ethanol solutions. Slides were quenched for 5 minutes in a 3% hydrogen peroxide aqueous solution to block endogenous peroxidase. Tissue was pretreated with Target Retrieval Solution (Dako S1699, Dako, Carpinteria, CA, USA) at pH 6.0. Tissues were stained with the Stathmin antibody (1:150, Santa Cruz sc-48362 mouse monoclonal, Santa Cruz, Dallas, TX, USA) for 60 minutes at room temperature using the Intellipath Autostainer Immunostaining System. Detection was performed using 2 step Mach 3 Mouse HRP Polymer Detection (Biocare M3M530L, Biocare, Concord, CA, USA) 20/20 minutes. The slides were counterstained with Richard Allen hematoxylin (Thermo Scientific, Kalamazoo, MI, USA), dehydrated with graded ethanol solutions, cleared with xylene and coverslipped. A tumor microarray consisting of 71 total biopsy tissue cores (9 thymoma A, 18 thymoma AB, 9 thymoma B1/B2, 15 thymoma B3, 3 thymic carcinoma, 7 normal thymus, 5 lung squamous cell carcinoma and 5 lymph node) was used for IHC staining. The TMA specimens were derived from patients ages 20-69 with an average age of 54. Male to female ratio of the specimen sources was 29:24 (data unavailable for some specimens). The clinical data associated with samples used for the thymoma tumor microarray is summarized in Table 3. The tumor microarray was scanned.
desmoyokin IHC scans were performed using the AHNAK/desmoyokin antibody (1:400, Atlas Antibodies HPA026643 rabbit polyclonal, Atlas Antibodies, Stockholm, Sweden). Antibodies to terminal deoxynucleotidyl transferase (TdT) diluted at 1:25 were used to stain a tumor microarray to serve as a control to define the presence of immature T lymphocytes in patient tissues.
<table>
<thead>
<tr>
<th>Sample</th>
<th>Age (years)</th>
<th>Gender</th>
<th>Sample</th>
<th>Age (years)</th>
<th>Gender</th>
<th>Sample</th>
<th>Age (years)</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>64</td>
<td>M</td>
<td>B2</td>
<td>63</td>
<td>M</td>
<td>LN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>55</td>
<td>F</td>
<td>B2</td>
<td>55</td>
<td>M</td>
<td>LN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>67</td>
<td>F</td>
<td>B2</td>
<td>47</td>
<td>M</td>
<td>LN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td>B2</td>
<td>56</td>
<td>M</td>
<td>LN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td>B2</td>
<td>50</td>
<td>F</td>
<td>LN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td>B2</td>
<td>42</td>
<td>F</td>
<td>Metaplastic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>56</td>
<td>F</td>
<td>B2</td>
<td>62</td>
<td>M</td>
<td>MicroN-Thy</td>
<td>68</td>
<td>F</td>
</tr>
<tr>
<td>A</td>
<td>56</td>
<td>F</td>
<td>B2</td>
<td>52</td>
<td>M</td>
<td>MicroN-Thy</td>
<td>64</td>
<td>M</td>
</tr>
<tr>
<td>A</td>
<td>64</td>
<td>M</td>
<td>B2</td>
<td>54</td>
<td>M</td>
<td>NL Thymus</td>
<td>65</td>
<td>F</td>
</tr>
<tr>
<td>A</td>
<td>63</td>
<td>M</td>
<td>B2</td>
<td>70</td>
<td>F</td>
<td>NL Thymus</td>
<td>20</td>
<td>F</td>
</tr>
<tr>
<td>A</td>
<td>55</td>
<td>M</td>
<td>B2</td>
<td>36</td>
<td>F</td>
<td>NL Thymus</td>
<td>35</td>
<td>F</td>
</tr>
<tr>
<td>A</td>
<td>43</td>
<td>M</td>
<td>B3</td>
<td>71</td>
<td>F</td>
<td>NL Thymus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>51</td>
<td>M</td>
<td>B3</td>
<td>35</td>
<td>F</td>
<td>NL Thymus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>52</td>
<td>M</td>
<td>B3</td>
<td>35</td>
<td>F</td>
<td>NL Thymus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>60</td>
<td>F</td>
<td>B3</td>
<td>62</td>
<td>M</td>
<td>NL Thymus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B/B2</td>
<td>52</td>
<td>M</td>
<td>B3</td>
<td>40</td>
<td>F</td>
<td>SQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>56</td>
<td>F</td>
<td>B3</td>
<td>52</td>
<td>M</td>
<td>SQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>64</td>
<td>M</td>
<td>B3</td>
<td>54</td>
<td>M</td>
<td>SQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>55</td>
<td>M</td>
<td>B3</td>
<td>37</td>
<td>M</td>
<td>SQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>43</td>
<td>M</td>
<td>B3</td>
<td>70</td>
<td>F</td>
<td>SQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>51</td>
<td>M</td>
<td>B3</td>
<td>46</td>
<td>M</td>
<td>Thy Ca</td>
<td>46</td>
<td>M</td>
</tr>
<tr>
<td>B1</td>
<td>60</td>
<td>F</td>
<td>B3</td>
<td>50</td>
<td>F</td>
<td>Thy Ca</td>
<td>46</td>
<td>F</td>
</tr>
<tr>
<td>B1</td>
<td>56</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td>Thy Ca</td>
<td>69</td>
<td>M</td>
</tr>
<tr>
<td>B1</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>69</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>65</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Patient age and gender for tissue specimens used for tumor microarray studies.

Results

Protein in Thymus vs. Thymoma Comparison

A total of 1193 proteins were identified in our proteomic analysis of normal thymus and tumor subtypes. Multiple pairwise comparisons were completed to determine differential protein expression between normal thymus tissue and each thymoma subtype. After
Stringent filtering: 918, 999, 957, 1034 and 982 matched protein identifications were retained for A, AB, B1, B2 and B3 vs. normal thymus, respectively. There were 1,168 unique proteins in the union of these datasets. Implementing a significance cut-off: 215, 469, 507, 470 and 456 proteins were determined to be differentially expressed in subtypes A, AB, B1, B2 and B3, respectively. The union of these datasets yielded a total of 907 unique proteins. The union of significantly differentially expressed proteins were evaluated by hierarchical clustering and principle component analysis (Figure 11A and Figure 12). There were 61 proteins that were differentially expressed in all five tumor subtypes when compared to normal thymus tissue (Figure 11B). Of these 61 proteins, 7 proteins belonging to the collagen family are downregulated in thymomas in relation to normal thymus (Table 4). The collagens CO1A1, CO1A2, CO3A1, CO5A1, CO6A1, CO6A2 and CO6A3 were distinguished by 17, 18, 7, 2, 20, 24 and 97 unique peptides respectively.
Figure 11. Clustering and Venn Diagram of Significant Proteins (q-value <0.20). A) Unsupervised clustering of the 35 individual thymus samples reveal distinctive clustering within different subsets of normal and tumorous thymus tissue. This indicates the interpatient variability does not exceed intersubtype variability. B) Proteins identified in thymoma samples were matched against normal thymus tissue to obtain a fold change in protein expression. Proteins displaying a significant fold change of q-value 0.20 were compared between the different tumor subtypes.
Figure 12. Principal Component Analysis of Thymoma Subtypes. A) Scatterplot of the first two principal components of the PCA B) Scree plot of the first eight principal components.
Table 4. Collagen Proteins Differentially Expressed in Thymomas from Normal Thymus. Seven collagen proteins were significantly differentially expressed between normal thymus and thymoma. The associated fold changes and significance are listed in this table.

Proteins in Thymoma Subtypes Comparison

Cluster analysis of the 35 samples showed two distinct clusters. Cluster one consists of subtypes AB, B1 and B2, while cluster two consists of subtype A and B3 and normal thymus. Furthermore, cluster two can be sub-clustered into subtypes A and B3 and normal thymus. Cluster one was unable to be sub-clustered due to the heterogeneous proteomic signature from intermixed T lymphocytes. The presence of T lymphocytes was confirmed in both H&E and IHC imaging. DAVID was used to compare the gene ontology terms enriched in each thymoma subtype. A qualitative analysis of the proteins identified in the different thymoma samples was visualized as a Venn diagram (Figure 11B). It is notable that 122 proteins were differentially expressed in thymomas AB, B1 and B2 (intersection), the largest overlap in this analysis.

Contribution of Lymphocyte Component to the Tumor Protein Profile
The WHO categorization of thymoma subtypes assesses T lymphocyte content. Specifically, subtypes AB, B1 and B2 are lymphocyte-rich in contrast to lymphocyte-poor A and B3. The data from this study demonstrated a pattern, which allowed for identification of candidate T lymphocytes proteins found within thymoma samples. Prothymosin alpha, precursor of thymosin α1, is produced in the thymus and cleaved to regulate T lymphocyte proliferation.\textsuperscript{115,116} Thus, prothymosin alpha was selected as a T lymphocyte associated marker to establish a pattern of expression for potential lymphocyte markers. The log$_2$ fold changes of prothymosin alpha in AB, B1 and B2 in comparison to normal thymus were 1.73 (q-value 0.0107), 1.57 (q-value 0.0120) and 0.862 (q-value 0.0902) respectively. Downregulation of prothymosin alpha was observed in types A and B3 thymomas, with significant changes observed in B3, -1.68 (q-value 0.247) and -1.65 (q-value 0.0262) respectively. Proteins following this pattern of exhibiting downregulation or no change in subtype A and B3 thymomas while exhibiting upregulation in A, B1 and B2 thymomas were assessed as T lymphocyte proteins candidates. A total of 263 proteins followed this expression pattern and was subsequently analyzed using cluster analysis. A heatmap of the 263 proteins across the 29 thymoma samples was produced after the MaxLFQ values were $z$-normalized (Figure 13A). Cluster analysis revealed distinctive clustering based on T lymphocyte infiltration status with the exception of one thymoma B3 sample. Proteins with sample variation were observed and visualized in the top and bottom of the heatmap. GO term molecular functions were assessed by DAVID using the total detected protein set as background. From the protein candidates, six of the proteins from this list showed function associated with T lymphocytes in the GO term T Cell Differentiation (GO:0030217) (p-
value $4.89 \times 10^{-2}$): CD3D, CD3E, ZAP70, BCL11B, LCK, and RPL22. The T lymphocyte candidate proteins were visualized in a line plot of log$_2$ fold change, highlighting proposed T lymphocyte protein expression pattern between subtypes using prothymosin alpha and the proteins associated with T cell differentiation (Figure 13B). Lymphocyte infiltration status of the different thymoma types was validated using anti-terminal deoxynucleotidyl transferase (TdT) immunohistochemistry (Figure 14).

The established onco-protein stathmin was evaluated as a potential marker for distinguishing thymoma subtypes.$^{117-119}$ Stathmin was found to be upregulated in thymoma AB, B1 and B2 in relation to normal thymus tissue (Figure 15B and C). Immunohistochemical staining of thymoma tissues with anti-stathmin showed that the protein was localized to the lymphocytes infiltrating the thymus tissue (Figure 15A). In contrast, lung squamous cell carcinoma, an epithelial tumor used as a positive staining control, stained heavily for stathmin in the epithelial cells. Thus, data indicates that the elevation of stathmin detected in AB, B1 and B2 are T lymphocyte in origin and not from the epithelial tumor.
Figure 13. Proteomic Profile Contribution of T Lymphocytes. A) Heatmap plot of the 263 T lymphocyte candidates across the 29 thymoma samples produced after MaxLFQ values were z-normalized. Hierarchical clustering was performed using Euclidean distance and average linkage using the Perseus software. B) Line plot of log₂ fold change by subtype of the pattern associated with T lymphocyte infiltration. Prothymosin alpha (red) is a known T lymphocyte associated protein found in the thymus. Proteins matching DAVID GO Term T Cell Differentiation (GO:0030217) were highlighted in orange. Statmin (blue) a protein assessed for subtype diagnosis potential. Ki67 (green) is a known cell proliferation marker.
Figure 14. Intratumoral Lymphocytes Throughout Thymoma Subtypes. A) Histology of normal thymus and the thymoma types: A, AB, B1, B2 and B3. Tissue samples were stained with hematoxylin and eosin. Images were captured at 400x magnification. B) The different thymus specimens were stained with TdT (terminal deoxynucleotidyl transferase) an established maker for developing T lymphocytes (40x magnification).
Figure 15. Changes in Stathmin Expression in Different Thymoma Subtypes. A) Heavy stathmin staining was noted in the T lymphocytes of AB, B1 and B2 as well as in the epithelial derived lung squamous cell carcinoma tumor (40x magnification). B) Bar graph displaying the intensity quantitation of stathmin for the thymus and thymoma samples. Standard deviation is denoted in the error bars. (q-values: *** <0.05, ** 0.05-0.10 and * 0.10-0.20) C) Log₂ fold changes of stathmin expression of the different thymomas in comparison to normal thymus.
**Proteins associated with type A and B3 thymomas**

Pairwise log<sub>2</sub> fold comparison of A and B3 thymomas was performed to assess for candidates that differentiate between the two subtypes, and 90 proteins were determined to show significant differences in expression. Gene ontology analysis through DAVID revealed proteins associated with the terms anti-apoptosis (GO:0006916, p-value 1.50 x 10<sup>-3</sup>), negative regulation of apoptosis (GO:0043066, p-value 1.90 x 10<sup>-2</sup>), negative regulation of programmed cell death (GO:0043069, p-value 2.10 x 10<sup>-2</sup>) and negative regulation of cell death (GO:0060548, p-value 2.10 x 10<sup>-2</sup>). These GO terms consists of an overlapping set of 7 proteins: annexin 1, annexin 4, alpha B crystallin, DDAH (dimethylarginine dimethylamaminohydrolase 2), heat shock protein 70, heat shock protein 90 and NFKB1 (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1) (Table 5).

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>UniProt Header</th>
<th>q-value</th>
<th>Log2 Fold Change (A/B3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annexin 1</td>
<td>P04083</td>
<td>0.138</td>
<td>-1.784</td>
</tr>
<tr>
<td>Annexin 4</td>
<td>P09525</td>
<td>0.121</td>
<td>-0.998</td>
</tr>
<tr>
<td>Alpha b crystallin</td>
<td>P02511</td>
<td>0.087</td>
<td>-4.181</td>
</tr>
<tr>
<td>DDAH 2</td>
<td>O95865</td>
<td>0.084</td>
<td>-2.019</td>
</tr>
<tr>
<td>HSP 70</td>
<td>P38646</td>
<td>0.023</td>
<td>0.798</td>
</tr>
<tr>
<td>HSP 90B1</td>
<td>P14625</td>
<td>0.135</td>
<td>-0.726</td>
</tr>
<tr>
<td>NFKB1</td>
<td>P19838</td>
<td>0.072</td>
<td>1.701</td>
</tr>
</tbody>
</table>

**Table 5.** Anti-apoptotic Proteins were Differentially Expressed in Thymomas A and B3. Gene ontology analysis determined that anti-apoptotic proteins in the set of 90 proteins showing significant differences in protein expressions thymomas A and B3. Fold changes and significance are listed in this table.
Desmoyokin is associated with medulla of normal thymus and type A thymomas

Desmoyokin, a protein encoded by the AHNAK (neuroblast differentiation-associated) gene, was a protein of interest exhibiting a log2 fold change of -1.81 (q-value 0.0668) in B3 thymomas in relation to A thymomas with known associations with cancer. Expanding the analysis across all tumor subtypes, desmoyokin was determined to be significantly downregulated in B1, B2 and B3 in comparison to normal thymus, while it was not determined to be significantly different in A and AB tissues from normal thymus tissue (Figure 16B and C). Immunohistochemistry staining of thymoma and normal thymus tissue using an antibody to desmoyokin confirmed that its expression was increased in subtype A (Figure 16A). A tumor microarray consisting of 36 viable cores was stained for desmoyokin expression and evaluated as negative, low or high. These 36 cores were subtyped and 8 were subtype A, 7 were B1, 10 were B2 and 11 were B3. AB thymoma cores were excluded due to the heterogeneity of the tumors. Results of this evaluation were in agreement with mass spectrometry data as the tumor microarrays for type A thymoma stained significantly higher for desmoyokin than Type B thymomas. The chi-square tests for associations between thymoma subtypes and desmoyokin resulted in a Pearson chi squared value of 15.9, indicating that its abundance was significantly associated with subtype (p-value 0.014). Association between thymoma A vs. grouped thymoma B (B1, B2 and B3) and desmoyokin was also performed. The resulting Pearson chi squared value obtained was 15.8, demonstrating significance (p-value 0.00038). We compared H&E, anti-desmoyokin and anti-proteasome subunit β5t stained normal thymus tissue (Fig 5D). From this staining we observed desmoyokin staining exclusively within the thymus
medulla and found proteasome subunit β5t staining uniquely within the cortex, as expected. In summary, we show that the protein desmoyokin is a potentially useful protein marker in difficult cases for differentiating the diagnosis in thymomas A and B3 clinically.
Figure 16. Desmoyokin is Differentially Expressed in and A and B Thymomas. A) Thymus tissue of A, AB, B1, B2 and B3 were stained with anti-desmoyokin antibody. The type A thymomas (A and AB) stained heavily while the type B (B1, B2 and B3) and normal thymus tissue expressed low amounts of the protein (40x magnification). B) Bar graph displaying the intensity quantitation of desmoyokin for the thymus and thymoma samples. Standard deviation is denoted in the error bars. (q-values: *** <0.05, ** 0.05-0.10 and * 0.10-0.20) C) Log2 fold changes of desmoyokin expression of the different thymomas in comparison to normal thymus. D) Normal thymus tissue is stained with hematoxylin and eosin, anti-desmoyokin and anti-proteasome β5t.
Discussion

This study is the first global proteomic profiling of normal thymus and thymomas through the use of archival patient tissues and LC-MS/MS. In this study we were able to identify protein profiles of the different disease subtypes. We identified desmooyokin as a protein associated with the thymus medulla, hence, a novel identification marker for type A thymomas.

It was determined that 61 proteins showed significant changes in expression level from normal thymus tissue. Seven of these 61 proteins belong to the collagen family, all of which were at lower abundance in thymomas. Changes in collagen in cancer have been extensively studied and reviewed. During the cancer development process, remodeled collagen I, III, IV is deposited while collagen that forms a physical barrier is degraded promoting tumor cell proliferation and invasion. Dysregulation of collagen levels in the stroma can be oncogenic. Existing studies have shown that COL1A1 (encoded by the collagen type 1 alpha 1 gene) has some predicative value in tumor diagnosis and prognosis where it has been found to be upregulated and downregulated in oral squamous cell carcinoma and hepatocellular carcinoma respectively. Another explanation for the observed collagen changes is the reversion of the atrophied adult thymus composed mostly of adipocytes and stroma. However, few studies have been conducted on collagen composition changes in association with the thymus. To date, this is the first evidence of collagen changes in association with thymomas. Future study on the changes in collagen distribution factors in thymoma may further elucidate its role in tumorigenesis.
GO term analysis showed that B3 thymoma samples were enriched for the term Proteasome Accessory Complex. This finding was confirmed through analysis using DAVID. Closer examination of this term revealed that several differentially expressed proteins were subunits responsible for forming the 26S proteasome. B3 showed significant overexpression of subunits 1, 3, 4 and 5. PSMD 4 was significantly upregulated in thymoma subtypes AB, B1, B2 and B3. PSMD 4 plays a key role in the targeted transport of ubiquitinated p53 into proteasomes for degradation.124,125 These data suggested the possibility of p53 dysfunction in certain thymoma subtypes. The IPA upstream regulator analysis further supported this hypothesis. Using the list of differentially expressed proteins of thymomas generated in the study, IPA predicted a downregulation of the effects of p53 in all of the thymomas, most confidently in the AB, B1 and B2 subtypes. However, previous studies have shown little value in assessment of quantity of the p53 protein in thymoma since loss of function can be associated with mutation.126,127 It has also been noted that the p53 mutation may play a role in the tumorigenesis of thymoma.128 Thus, a future study would be to elucidate the role of p53 in thymomas through a study of protein quantity by immunohistochemistry, assessment of p53 ubiquitination status and protein half-life, combined with sequencing for mutations of the p53 gene.

Hierarchical cluster analysis revealed a distinctive separation of the different thymoma subtypes into two groups, with AB, B1 and B2 grouping together while A, B3 and normal thymus formed a second group. The proteomic results in this study were then validated using IHC. We concluded that lymphocyte presence contributed significantly to the clustering results. The histology based subtype classification devised by the WHO
defines AB, B1 and B2 thymomas as possessing heavy non-neoplastic T lymphocyte infiltration. Thymoma AB contains large pockets of lymphocytes within the epithelial cell structure in the type B-like regions. To establish a pattern associated with T lymphocyte content, a marker with a known relationship to T lymphocytes with specificity to the thymus was needed. Prothymosin alpha, the precursor to thymosin α1, was selected for this purpose. Thymosin α1 is a peptide signaling molecule mainly produced in the thymus that plays a role in the proliferation and maturation of T cells. Due to the immersion of the biopsy tissue during the FFPE preservation process, free active biomolecules were likely lost. However, thymosin α1 can be monitored through its uncleaved precursor prothymosin alpha found inside epithelial tumor cells. The use of prothymosin alpha as a surrogate for thymosin α1 was previously demonstrated in literature for radioimmunoassays. Our proteomics analysis of FFPE archival tissue revealed that in comparison to normal thymus tissue thymoma of AB, B1 and B2 types show an increase in prothymosin alpha while thymomas A and B3 show a lower level of prothymosin alpha, consistent with expected levels of T lymphocytes. Thus, using the expression pattern of prothymosin alpha as a template, candidates for other T lymphocyte associated proteins were identified. This finding also offers a potential explanation that tumor tissue may be abnormally producing signaling factors leading to the proliferation of the infiltrating T lymphocytes.

A total of 263 unique proteins were identified as T lymphocyte associated protein candidates. The clustering of this data set revealed segregation between thymoma samples associated with the degree of lymphocyte infiltration with the exception of one B3 outlier.
As thymoma B3 originates from the thymus cortex, the site of lymphocyte maturation, a varying amount of T lymphocytes may be present. It can be hypothesized that the outlier sample possessed a higher lymphocyte content resulting in the observed grouping. Through DAVID GO term analysis 6 of the proteins expressed multiple hits in functions associated with T lymphocytes in the GO term T cell differentiation. CD3 delta (CD3D) and epsilon (CD3E) are chains of the CD3 complex. CD3 is a component of the T-cell receptor complex, which also consists of the ζ chain and either αβ T-cell receptor or γδ T-cell receptor. The T-cell receptor complex conveys signals from the extracellular T cell receptor to intracellular signaling pathways. CD3D is required for transition of the T cell from the double positive to the single positive stage. CD3E is necessary for progression of T cells past the double negative stages. The loss of CD3E leads to absence of mature double and single positive thymocytes implicating its role in pre-TCR development. Brodeur et al. showed that the intracytoplasmic tail of CD3E is essential for double negative to double positive transition. Tyrosine kinase LCK (LCK), tyrosine kinase ZAP-70 (ZAP70) and B-cell lymphoma/leukemia 11B (BCL11B), play roles in the differentiation and survival of T lymphocytes. LCK plays a role in pre-TCR signaling at the double negative 3 (DN3) transition to double negative 4 (DN4) for the commitment of alpha-beta T cells development. Ribosomal protein, RPL22 is also essential for the alpha-beta differentiation of T cells during the DN3 stage. It is also notable that Ki67, a well-established proliferation marker, was found in all 18 of the AB, B1 and B2 samples but not in any of the A and B3 samples. In further support of the importance of lymphocyte contribution to the protein profile, a high degree of overlapping
proteins was found among thymoma AB, B1 and B2 samples which are characteristically lymphocyte-rich, as confirmed by TdT IHC staining.

We report a basis for differentiating between these subtypes and normal tissue using an expression pattern derived from markers unique to the thymus that are known to be associated with the T lymphocyte processes. Following GO term analysis of potential T lymphocyte candidates, six proteins were highlighted to be overexpressed in thymomas: LCK, BCL11B, RPL22, CD3D, CD3E and ZAP70. Interestingly, these proteins are associated with the T lymphocyte differentiation process which occurs in the thymus cortex. The statistical similarity of the low coexistence of thymoma and myasthenia gravis in patients diagnosed with thymomas A and AB may be due to the fact that both of these subtypes contain thymus medulla components where negative selection elimination of potential autoimmune cells occur. On the other hand, B type thymomas are derived from the thymus cortex, the lack of thymus medullary components for negative selection is a possible explanation for the higher rate of autoimmunity observed in these cases. Myasthenia gravis is a disease caused by abnormal antibody binding to post synaptic muscle acetylcholine receptors. This blocks or weakens the action potential signal from the somatic nervous system leading to a weaker contraction. The thymus is the site of T lymphocyte maturation, for both positive and negative selection, to ensure the cells can interact with major histocompatibility molecules without reacting to antigens leading to autoimmunity. It is currently hypothesized that an abnormality of the thymus, such as thymus hyperplasia or thymoma, causes misprogramming of T lymphocytes leading to the observed autoimmunity. \(^\text{151}\) Statistics from the National Cancer Institute show that
complications from myasthenia gravis occur at different rates for the different thymoma subtypes: A (17%), AB (16%) B1 (57%), B2 (71%) and B3 (46%). Thus, we hypothesize that the proliferation of T lymphocytes in the presence of B type thymoma epithelial tumor cells is enhanced by signaling factors leading to abnormal T lymphocyte development and autoimmunity.

The thymus produces a set of signaling molecules known as thymosins. Thymosin α1 is associated with stimulating T lymphocyte response, thus making it an attractive candidate for further study. As an individual ages the thymus normally atrophies with increasing adipose tissue. The level of thymosin also decrease with age.\textsuperscript{131} Studies have been conducted in the past regarding the levels of thymosins, thymomas and myasthenia gravis. However, the results have been somewhat contradictory on whether thymosin α1 is elevated or decreased in thymoma cases.\textsuperscript{152-155} Additionally, there is disagreement on whether thymosin alpha is normal or upregulated in myasthenia gravis. This may be attributed to the fact that these studies were conducted prior to the WHO subtyping of thymomas, hence differences between the diseases were not taken into account.

Our mass spectrometry data shows a significant increase in the precursor prothymosin alpha, suggesting the possibility of an increase in thymosin α1 as well, in thymoma AB, B1 and B2 in comparison to normal thymus tissue. However, there is a significant decrease of prothymosin alpha in thymoma B3 in relation to normal thymus tissue. Thus, it may be possible that the presence of thymosin α1 only plays a role in the presence of T lymphocytes infiltrating the tissue but is not necessarily involved in the
autoimmunity observed in certain thymoma cases. This is supported by the fact that not all cases of AB, B1 and B2 thymomas with T lymphocyte infiltrate has autoimmune reactions.

At this time there have been no proteomic or other molecular studies investigating the differences in autoimmune response in the different thymoma subtypes. A necessary step to directly determine the signaling molecules involved in autoimmunity requires patient clinical data as well as bloodwork during the period of illness to search for the antibody effectors of disease.\textsuperscript{156} Combining data from these processes will allow a better picture elucidating the nature of all disease processes occurring in an individual at a given time. Thus, elucidating the association between the different thymoma subtypes and myasthenia gravis complications warrants further investigation in the future.

Stathmin, a protein involved in the regulation of microtubule dynamics, has been well established as a marker for high cell proliferation rate.\textsuperscript{157,158} Stathmin is over-expressed in several different cancer types.\textsuperscript{119,159-162} High levels of stathmin have been associated with poor prognosis in multiple cancer types and resistance to drugs stabilizing microtubules, such as taxane.\textsuperscript{163} Furthermore, depletion of stathmin leads to cell cycle arrest in the G2 phase and apoptosis.\textsuperscript{164} Due to the known correlation between stathmin expression level and other epithelial cancers, the differential expression of stathmin detected between the thymoma tumor types was further investigated using IHC. The IHC staining showed that the stathmin was localized to the T lymphocytes infiltrating the biopsy tissue, leading to the conclusion that stathmin is a lymphocyte associated protein. The positive staining of stathmin in T lymphocytes is supported by previous studies, as stathmin plays a critical role in the activation of T lymphocytes as a regulator of cell polarization.
and T lymphocyte migration from the vascular compartment across tissue barriers.\textsuperscript{164,165} The use of stathmin to distinguish the different thymoma subtypes is limited due to its staining of lymphocytes as opposed to the tumorous epithelial cells themselves, as CD4 and CD8 are well-established IHC targets for T lymphocytes.\textsuperscript{166}

Comparison of the proteomic profiles between A and B3 revealed 90 proteins to be significantly expressed differentially between the subtypes. Of these 90 proteins most were involved in biological functions not obviously associated with cancer including sugar and amino acid metabolism. Seven of these proteins are involved in anti-apoptotic processes. It is known that some proteins may serve roles as either onco-proteins or tumor suppressors in different diseases. Annexins 1 and 4 were found to be significantly downregulated in thymoma B3 relative to thymoma A. Interestingly, decreases in Annexin 1 expression have been correlated to esophageal and prostate cancer severity while showing the opposite relationship with pancreatic cancer.\textsuperscript{167,168} Annexin 4 while typically found over expressed in most tumors was found to have the opposite relationship in prostate cancer.\textsuperscript{169,170} HSP70 expression is also higher in thymoma B3 than A, consistent with its activity promoting cell survival.\textsuperscript{171,172} We found that NFKB1, which controls genes associated with processes such as apoptosis inhibition and cell cycle progression, was expressed significantly higher in B3 thymoma.\textsuperscript{173-175} DDAH2 is a protein that was found relatively expressed at a greater level in type A thymoma than in type B3. There exists evidence that DDAH2 expression may be up or down regulated in different tumor types, as exemplified by ovarian carcinoma and oral squamous cell carcinoma respectively.\textsuperscript{176,177} High levels of proteins HSP90b1 and αB-crystallin are typically associated with cancer progression.\textsuperscript{169,178-180} Surprisingly, these
proteins were found to be significantly more highly expressed in thymoma A over thymoma B3, despite B3 being the more severe disease. Differences in the site of origin of thymoma A and B3 may play a role in these unexpected results. Since there is no previous work on these proteins in the context of thymoma, further study and validation of these proteins is required to understand their specific roles in the disease process and to assess their potential as diagnostic and/or prognostic markers.

The search for a protein marker to differentiate types A and B3 thymomas for potential use in clinical diagnosis led to the investigation of desmoyokin, the protein encoded by the AHNAK (neuroblast differentiation-associated) gene. Protein candidates were selected for further evaluation based on highest absolute fold change difference between thymomas A and B3. Candidates included αB-crystallin (log_2 -4.18 fold change), protein s100-a10 (log_2 -2.13 fold change), protein s100-a6 (log_2 -2.05 fold change), desmoyokin (log_2 -1.81 fold change) and galectin 7 (log_2 3.58 fold change). The protein candidates were then further evaluated through assessment of fold change between all thymoma subtypes and normal thymus tissue for expression level patterns. Protein s100-a10 and galectin 7 were only present in A and B3 thymomas respectively when matched with normal thymus tissue. Protein s100-a6 was found to be elevated in A, AB, B2 and B3. While s100-a6 is a known tumor marker, its enhanced expression in four of the thymoma subtypes excluded it as a useful marker to distinguish between the different subtypes. αB-crystallin showed elevated expression level in type A thymoma while decreased expression in AB, B1, B2 and B3, significantly in B1 and B3.
Desmoyokin was significantly downregulated in all three B thymomas while exhibiting little changes in type A and AB, supporting the fact that the protein is associated with the epithelial cell derived thymoma tumors as opposed to T lymphocytes. Furthermore, we hypothesized that desmoyokin serves as a marker distinguishing the thymus medulla from the cortex. Desmoyokin was selected for further study through IHC of tumor microarray due to its significance and expression level pattern between the A and B thymoma subtypes.

Comparison of desmoyokin staining of thymus tissue with staining by hematoxylin and eosin, and anti-proteasome subunit β5t revealed that regions were stained either exclusively with desmoyokin or with proteasome subunit β5t. Proteasome subunit β5t is a protein previously described as expressed exclusively in the thymic cortical epithelial cells.\textsuperscript{107} Use of proteasome subunit β5 was selected as a negative control since has been shown to be unique to the cortex. It is notable that Keratin 5 and Keratin 8 expression are also used as markers for medulla and cortex but other studies have cast doubt on their specificity.\textsuperscript{182-185} Our proteomic data showed no significant difference in Keratin 5 or Keratin 8 expression in A vs. B3 thymoma subtypes. Other markers that have been associated with cortical thymus epithelial cells, such as LY51 and CD205, were also not observed in our proteomic results. Thus, it was concluded that desmoyokin is a marker present in the thymus medulla but not the cortex, providing objective identification of thymus cortex and medulla. Since type A thymomas are derived from medullary thymus tissue and type B thymomas are derived from cortical tissue, desmoyokin serves as a marker distinguishing tumor subtypes based on tissue of origin. Whether desmoyokin
observed in the thymus plays a role in tumorigenesis or the reason for this alteration has yet to be determined. The desmoyokin protein may be downregulated in cancers.\textsuperscript{186,187} Previous studies have shown that its levels negatively correlate with cell division rate. This protein is post-translationally modified through phosphorylation of serine and threonine, with phosphorylation occurring in actively growing cells. A study by Lee et al. showed that desmoyokin can function as a tumor suppressor through regulation of the TGF\(\beta\)/Smad pathway leading to cell cycle arrest.\textsuperscript{188} This study showed that desmoyokin expression was significantly downregulated in B subtype thymomas compared to the A subtype thymomas and normal thymus tissue. The A subtype thymomas did not show a significant difference in desmoyokin expression when compared to normal thymus tissue. It is clear that much remains to be learned about the desmoyokin protein in the context of cancer.\textsuperscript{188} One potential future course of study is to study the phosphorylation states of the desmoyokin protein in thymomas to assess for possible correlation with abnormal cell growth patterns. Thus, this protein is a target for future study not only as a marker for clinical function to distinguish between thymoma types A and B3 but also to further elucidate the changes associated with tumorigenesis of thymomas from the different regions of the thymus.

**Conclusion**

This study is the first global proteomic characterization of all epithelial derived thymoma subtypes compared to normal thymus, overcoming the challenges associated with studying a disease with no representative cell line or animal model. Challenges due to the rarity of this disease were addressed through the utilization of FFPE tissues collected over two
decades. The resulting proteomic characterization has led to the discovery of interesting potential biomarker proteins: desmoyokin and stathmin. This study revealed that desmoyokin was downregulated in thymus cortex tissue, the site of origin for subtype B thymomas but not subtype A thymomas. This protein may allow for differentiation between thymomas based on their site of origin. The stathmin protein ultimately proved to be unsuitable for a disease biomarker to separate thymoma subtypes, but its changes in abundance across these thymomas supports existing knowledge of associated lymphocytic infiltration. These findings on desmoyokin and stathmin have opened the possibility for several future studies exploring the role of the immune system, the cell cycle, and cell motility pathways in the development of this disease. The molecular data from this study also provide a new opportunity for comparison to other cancers in an effort to elucidate the process of tumorigenesis.
Chapter 4: Measuring the Impact of Chromatin Modifications on Global Protein Expression

Introduction

Histones are an attractive molecular target for ongoing cancer research for several reasons. Histones are a group of proteins that wrap DNA to form chromatin fibers. They serve as epigenetic regulators for controlling access of transcriptional machinery to DNA and create binding sites to other transcriptional factors. Two of each of the four core histones proteins (H2A, H2B, H3 and H4) are arranged in an octamer with linker histone H1 regulating the DNA entry and exit to the complex. The core histones possess amino acid tails directed outwards from the histone octamer where enzymes reversibly add and remove post-translational modifications (PTMs) to specific residues within the histone tail leading to changes in histone binding properties to DNA and alterations in protein binding sites. Histone PTMs include acetylation, methylation, ubiquitination, phosphorylation, ADP ribosylation and sumoylation. Histone PTMs have been associated with critical cellular functions such as DNA repair and stability, development and apoptosis, all processes which deregulation can lead to cancer. Epigenetic regulation through histone PTM is furthermore complicated by the presence of histone crosstalk with other histone PTM sites as well as direct DNA methylation. There is much that remains unknown about epigenetic gene regulation at this time.
Due to large number of cellular processes regulated by histone PTMs, diseases in several organ systems have been associated with deregulation. Conditions in the neurological, endocrine, cardiological, gastrointestinal, immunological and pulmonary systems have been linked to alterations in histone PTM states (Figure 17A). Several histone PTMs have been linked to specific cancer types and processes (Figure 17B). Use of drugs to restore histone acetylation at specific residues such as Histone Deacetylase (HDAC) Inhibitors is already in practice to treat several cancers, most notably acute myeloid leukemia (AML). For our study, we investigated several specific histone PTM sites that have been associated with cancer or processes leading to or resulting in the progression of cancer: H4K16 (breast cancer and cancer cell sensitivity to chemotherapy), H3K4 (prostate carcinoma), H3K56 (preserving genomic integrity) and H4K91 (DNA replication and repair). Further evidence that linked cancer to histone PTMs include studies that have shown oncogene activation and tumor repressor loss due to abnormal histone PTMs which are reversible with the restoration of modification states. Study of the global proteomic effects of these specific histone PTM sites will allow for better understanding of their downstream effects, playing an important role in therapeutic target and disease biomarker discovery not limited to cancer but to a great variety of clinical conditions.
Figure 17. Examples of Human Disease Associated with Histone Post Translational Modifications. A) Histone post translational modifications are involved in many clinical conditions. B) Histone post translational modifications play an important role in many cancers. Consequences of altered histone modifications include: changes in gene expression, silencing at heterochromatin, cell cycle checkpoint instability and impaired DNA repair.
Methods

The overall experimental design is summarized in Figure 18.


Yeast Histone Mutant Creation

Yeast histone mutant plasmids were created in the laboratory of Dr. Mark R. Parthun. The plasmids contained wild type yeast histone genes H3 and H4 (HHT2 and HHF2). These genes were mutated to contain the desired point mutation mutating lysine to arginine to simulate the constitutive non-acetylated state, and lysine to glutamine to simulate the constitutive acetylated state and lysine to alanine to simulate a null mutation. Plasmids contain an ampicillin resistance gene for bacterial screening during amplification of plasmids in E. coli. The plasmids also contained a gene for the synthesis of tryptophan, for
screening of yeast colonies that had expressed plasmid bound genes. Plasmids were collected using miniprep by ethanol precipitation. The cells used for these transformations were UCC1111 yeast cells which had the endogenous H3 and H4 gene deleted from the genome. These cells possess a plasmid containing the endogenous gene allowing for viability. The UCC1111 cells were transformed though heat shock at 42º C for 40 minutes with an in-house made transformation mix containing 50% polyethylene glycol, 1 M lithium acetate, plasmid DNA and non-specific DNA from salmon sperm. The resulting cells are plated on a HC (Hartwell’s Complete) minus tryptophan plate for three days to select for cells containing and expressing the plasmid. Colonies are then further screened and stored on YPD (Yeast Extract Peptone Dextrose) plates. A further screen is the expression of the histone mutant plasmid disrupts the pigment producing ADE2 gene leading to colonies with mutant histones to appear pink.

**SILAC Labeling**

Yeast cells were stored and grown on YPD plates. For experimentation, one colony was taken from the plate and grown in 5 mL YPD media overnight. Mutant yeast cells were placed in HC media containing labeled lysine ($^{13}$C$_6^{15}$N$_2$) while wild type yeast cells were grown in regular HC media. Cells were grown to log phase (~0.5-0.7 OD) and harvested by centrifugation. Mutant yeast cells were mixed with wild type cells on a 1:1 ratio by weight of cells harvested.

**Yeast Global Protein Extraction**
The cell pellet was suspended in 250 µL of 20% TCA. Sample was spun for 3 minutes at 13.4 k RPM remove TCA then frozen pellet in -80 degrees for 30 minutes. Pellet was then thawed on ice and 250 µL of 20% TCA was added with add 250 µL of glass beads and kept on ice. Samples were alternatively vortexed for 1 minute then placed on ice for 1 minute 10 times. 300 µL of 5% TCA is added then spun for 2 minutes using minispin supranatant was collected then spun down at 15,000 RPM for 15 minutes to collect the pellet. Sample was spun at 13.4k RPM in 700 µL of 5% TCA for 15 minutes and pellet was collected then washed in 750 µL of EtOH and spun down. The sample was redissolved in Rapigest (in 25 mM Ammonium Bicarbonate) (Waters). The sample underwent trypsin digestion for 17 hours at 37 degrees. Rapigest was removed per protocol provided by manufacturer. Centrifuge sample for 15 minutes at 13.4k RPM. Supernatant was then dried down by vacuum centrifugation. Sample was resuspended in HPLC water.

**HPLC and Mass Spectrometry**

HPLC separation of the sample was carried out on a Dionex UltiMate 3000 with reversed-phase C18 column (Michrom Bioresources Magic C18AQ, 200 µm x 150 mm, 3 µm, 200 Å) at a flow rate of 2 µL/min and detection was carried out on a Thermo Scientific LTQ Orbitrap mass spectrometer with a Michrom source for high-resolution and high-mass-accuracy results. The HPLC solvents used were H2O with 0.1% Formic Acid for the aqueous phase and Acetonitrile with 0.1% Formic Acid for the organic phase. The gradient used was as follows: 0-7 min 2% organic phase, 7-367 min linear increase from 2% to 35% organic phase, 367-377 min linear increase from 35%-50% organic phase, 377-387 min
linear increase from 50%-90% organic phase, 387-398 min 90% organic phase, 397-400 min decrease to 2% organic phase until the end of the run at 430 min.

Mass Spectrometry Data Analysis

Raw mass spectrometry were converted to mzXML and searched through MassMatrix, an in-house search engine against the yeast protein database from YeastMine (February 2011) (yeastmine.yeastgenome.org), to obtain peptide sequence information for protein identification and SILAC quantitation results (Figure 19).93-95,198,199 SILAC relative quantitation data was expressed as a ratio of heavy protein to unlabeled protein. Normalization of data was performed as follows using the statistical package R (http://www.r-project.org/) and the MASS add on package. First the ratios were log₂ transformed. The mean ratio was then subtracted from every ratio to generate a Gaussian distribution with mean of 0, because most proteins are assumed to remain unchanged. The normalized kernel density plots for the ratios were then plotted to determine skewness and kurtosis.
Figure 19. Raw LC-MS/MS data from LTQ Orbitrap. A) Chromatogram displaying total peak sum as a function of time as sample is separated by the HPLC. B) Mass spectrometry data displaying signal abundance of given masses at a certain time point. C) Example of SILAC peaks separated by 8 Da due to the isotope label.

The online bioinformatics tools YeastMine, Database for Annotation, Visualization and Integrated Discovery (DAVID), and Panther were used to assess and group the proteins exhibiting significant change on the basis of cellular location, molecular function and role in key biological processes. All three tools were used concurrently for multiple analyses and for validation of each individual search results. YeastMine allowed for a Gene Ontology, publication, pathway, and interaction search. Additionally, YeastMine displayed orthologues of the gene list to various organisms including humans. Panther was used to cluster proteins by molecular function, biological processes, cellular component, protein
class and pathways. DAVID allowed for GO terms searches, pathway analysis (Kyoto Encyclopedia of Genes and Genomes), and protein domain analysis.

_Real Time PCR Quantitation of mRNA Transcripts_

mRNA was extracted from cell lysate using TriZol following TriZol protocol. Reverse transcription was performed using the High Capacity cDNA Reverse Transcription Kit (Invitrogen) with MultiScribe reverse transcriptase. TaqMan Real Time PCR was run by the TaqMan universal master mix (Life Technologies, 4304437) with the gene specific TaqMan detection assays. TaqMan RT-PCR assays were used to detect and quantify the target genes SUI1 (Life Technologies, Sc04160161), RTC3 (Life Technologies, Sc04130164) and HSP12 (Life Technologies, Sc04120148). The control used for the RT-PCR runs was using a TaqMan transcript detection assay (Life Technologies, Sc04120488) to actin. Samples were plated on 96 well plates for TaqMan PCR and run using an Applied Biosystems 7300 Real Time PCR System.

_Polyribosome Profiling_

Yeast cells were grown in 250 uL of YPD media until log phase (~0.5-0.7 OD) then protein synthesis is arrested with cyclohexamide (100 ug/mL) for 10 minutes then harvested by centrifugation. Cells were weighed then resuspended in 1.5 times lysis buffer per gram of cells (10mM Tris pH 7.4, 150mM NaCl, 20mM MgCl2, 20mM KCl, Protease Inhibitor, Phosphatase Inhibitor, 100µg/mL cycloheximide, 2 mM DTT and RNAse out (5µL/ml, Thermo Fisher Scientific, 10777-019)). Cells were lysed using glass beads alternating
between 3 minutes vortex and 2 minutes on ice three times total. The supernatant was collected for polyribosome profiling. Samples were run across a sucrose gradient and collected into 24 fraction with the absorbance recorded.

Real Time PCR Quantitation of Polyribosome Fractions

Internal control was provided by spiking Firefly Lulicerase control mRNA (Promega, L4561) into the TriZol (Thermo Fisher) at 5 picograms per mL TriZol. mRNA was extracted from polyribosome fractions using TriZol following TriZol protocol. Odd number fractions that were collected from the polyribosome profiling was used for this experiment. Reverse transcription was performed using the High Capacity cDNA Reverse Transcription Kit (Invitrogen) with MultiScribe reverse transcriptase. TaqMan Real Time PCR was run by the TaqMan universal master mix (Life Technologies, 4304437) with the gene specific TaqMan detection assays. Luciferase TaqMan assay used was MR03987587 by Life Technologies (now Thermo Fisher Scientific). TaqMan RT-PCR assays were used to detect and quantify the target genes SUI1 (Life Technologies, Sc04160161), RTC3 (Life Technologies, Sc04130164) and HSP12 (Life Technologies, Sc04120148). The control used for the RT-PCR runs was using an actin TaqMan transcript detection assay (Life Technologies, Sc04120488). Samples were plated on 96 well plates for TaqMan PCR and run using an Applied Biosystems 7300 Real Time PCR System.

Results

SILAC Yeast Histone Mutant Mass Spectrometry
Yeast SILAC mass spectrometry experiments were performed on several yeast histone mutant cell lines: H3K4A, H4K16Q, H4K16R and H4K91Q. Later yeast histone mutants H3K56R and H4K56Q experiments were conducted to assess for potential translational effects associated with the mutations. The collected SILAC data was normalized then plotted as density plots (Figure 20). Several proteins were observed to be downregulated after the histones were mutated. Initially we analyzed only proteins displaying significant (p<0.05) differences between the heavy and light ratio (Table 6, Table 7, Table 8 and Table 9). However, due to the small number of proteins fitting this criterion we included proteins exhibiting a two-fold or greater change in abundance. The proteins were clustered by Gene Ontology (GO) terms (Table 10) and pathways (Table 11) using DAVID. Panther was also used to analyze the proteins expressing significant change in the yeast mutants to assess for patterns in function (Figure 21).
Figure 20. Density Plots of SILAC Mutant to Wild Type Ratios. Ratio data obtained from the MassMatrix database search is converted to $\log_2$ and plotted using the statistics tool R for data normalization. Note the number of proteins displaying down regulation in the histone mutants.
<table>
<thead>
<tr>
<th>ID</th>
<th>Protein Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAP4</td>
<td>Vacuolar aminopeptidase</td>
<td>Vacuole targeting (CVT) pathways</td>
</tr>
<tr>
<td>ALD3</td>
<td>Cytoplasmic aldehyde dehydrogenase</td>
<td>Involved in beta alanine synthesis</td>
</tr>
<tr>
<td>PRE6</td>
<td>Alpha 4 subunit of the 20S proteasome</td>
<td>Relocates from cytosol to the mitochondrial surface upon oxidative stress</td>
</tr>
<tr>
<td>CRP1</td>
<td>Protein that binds to cruciform DNA structures</td>
<td>Binds to cruciform DNA structures</td>
</tr>
<tr>
<td>SPF1</td>
<td>P type ATPase</td>
<td>Involved in ER function and Ca2+ homeostasis</td>
</tr>
<tr>
<td>PET9</td>
<td>Major ADP ATP carrier of the mitochondrial inner membrane</td>
<td>Major ADP ATP carrier of the mitochondrial inner membrane</td>
</tr>
<tr>
<td>WSS1</td>
<td>Sumoylated protein</td>
<td>DNA damage response</td>
</tr>
<tr>
<td>TAF5</td>
<td>Subunit (90kDa) of TFIID and SAGA complexes</td>
<td>Involved in RNA polymerase II transcription initiation and in chromatin modification</td>
</tr>
<tr>
<td>YPD1</td>
<td>Phosphorelay intermediate protein</td>
<td>Osmotic stress response</td>
</tr>
<tr>
<td>ADE6</td>
<td>Formylglycinamidine ribonucleotide (FGAM) synthetase</td>
<td>De novo purine nucleotide biosynthetic pathway</td>
</tr>
<tr>
<td>CTS1</td>
<td>Endochitinase</td>
<td>Required for cell separation after mitosis</td>
</tr>
</tbody>
</table>

Table 6. Proteins Exhibiting Significant Change in SILAC Ratios in H3K4A.
<table>
<thead>
<tr>
<th>ID</th>
<th>Protein Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDC6</td>
<td>Pyruvate decarboxylase</td>
<td>Involved in amino acid catabolism</td>
</tr>
<tr>
<td>PDC5</td>
<td>Pyruvate decarboxylase</td>
<td>Decarboxylates pyruvate to acetaldehyde</td>
</tr>
<tr>
<td>ILV5</td>
<td>Acetohydroxyacid reductoisomerase</td>
<td>Branched chain amino acid biosynthesis</td>
</tr>
<tr>
<td>EMI2</td>
<td>Non essential protein</td>
<td>Transcriptional induction of the early meiotic specific transcription factor IME1</td>
</tr>
<tr>
<td>HXT3</td>
<td>Low affinity glucose transporter</td>
<td>Induced in low or high glucose conditions</td>
</tr>
<tr>
<td>ALD3</td>
<td>Cytoplasmic aldehyde dehydrogenase</td>
<td>Involved in beta alanine synthesis</td>
</tr>
<tr>
<td>ACO1</td>
<td>Aconitase</td>
<td>Required for the tricarboxylic acid (TCA) cycle</td>
</tr>
<tr>
<td>HSP78</td>
<td>Oligomeric mitochondrial matrix chaperone</td>
<td>Prevent the aggregation of misfolded proteins, resolubilize protein aggregates</td>
</tr>
<tr>
<td>YMR196W</td>
<td>Putative protein of unknown function</td>
<td>Unknown</td>
</tr>
<tr>
<td>GDI1</td>
<td>GDP dissociation inhibitor</td>
<td>Regulates vesicle traffic in secretory pathways</td>
</tr>
<tr>
<td>ARG1</td>
<td>Arginosuccinate synthetase</td>
<td>Catalyzes the formation of L arginosuccinate</td>
</tr>
<tr>
<td>ADE16</td>
<td>Enzyme of de novo purine biosynthesis</td>
<td>Enzyme of de novo purine biosynthesis</td>
</tr>
<tr>
<td>SEC26</td>
<td>Essential beta coat protein of the COPI coatamer</td>
<td>Involved in ER to Golgi protein trafficking</td>
</tr>
</tbody>
</table>

Table 7. Proteins Exhibiting Significant Change in SILAC Ratios in H4K16Q.
<table>
<thead>
<tr>
<th>ID</th>
<th>Protein Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>YEL047C</td>
<td>Soluble fumarate reductase</td>
<td>Required for anaerobic growth</td>
</tr>
<tr>
<td>ADE3</td>
<td>Cytoplasmic trifunctional enzyme C1 tetrahydrofoloate synthase</td>
<td>Involved in single carbon metabolism</td>
</tr>
<tr>
<td>LAP4</td>
<td>Vacuolar aminopeptidase</td>
<td>Cytosol to vacuole targeting (CVT) pathway</td>
</tr>
<tr>
<td>KAP95</td>
<td>Karyopherin beta</td>
<td>Mediate nuclear import of NLS containing cargo proteins via the nuclear pore complex</td>
</tr>
<tr>
<td>HXT3</td>
<td>Low affinity glucose transporter</td>
<td>Induced in low or high glucose conditions</td>
</tr>
<tr>
<td>YRA1</td>
<td>RNA binding protein</td>
<td>Required for export of poly(A)+ mRNA from the nucleus</td>
</tr>
<tr>
<td>ARO2</td>
<td>Bifunctional chorismate synthase and flavin reductase</td>
<td>Catalyzes the conversion of 5 enolpyruylshikimate 3 phosphate (EPSP) to chorismate</td>
</tr>
<tr>
<td>NUP2</td>
<td>Nucleoporin</td>
<td>Involved in nucleocytoplasmic transport</td>
</tr>
<tr>
<td>RNR1</td>
<td>Regulatory subunit of ribonucleotide diphosphate reductase</td>
<td>Catalyzes rate limiting step in dNTP synthesis</td>
</tr>
<tr>
<td>GRX4</td>
<td>Glutathione dependent oxidoreductase</td>
<td>Protects cells from oxidative damage</td>
</tr>
<tr>
<td>SPT16</td>
<td>Subunit of the heterodimeric FACT complex</td>
<td>Facilitates RNA Polymerase II transcription elongation</td>
</tr>
<tr>
<td>KEL1</td>
<td>Protein required for proper cell fusion and cell morphology</td>
<td>Protein required for proper cell fusion and cell morphology</td>
</tr>
</tbody>
</table>

*Table 8. Proteins Exhibiting Significant Change in SILAC Ratios in H4K91Q.*
<table>
<thead>
<tr>
<th>ID</th>
<th>Protein Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAP4</td>
<td>Vacuolar aminopeptidase</td>
<td>Cytosol to vacuole targeting (CVT) pathway</td>
</tr>
<tr>
<td>ADE17</td>
<td>Enzyme of de novo purine biosynthesis</td>
<td>De novo purine biosynthesis</td>
</tr>
<tr>
<td>FAA1</td>
<td>Long chain fatty acyl CoA synthetase</td>
<td>Activation of imported fatty acids</td>
</tr>
<tr>
<td>PAB1</td>
<td>Poly(A) binding protein</td>
<td>Mediates interactions between the 5 cap structure and the 3 mRNA poly(A) tail</td>
</tr>
<tr>
<td>GRX4</td>
<td>Glutathione dependent oxidoreductase</td>
<td>Protects cells from oxidative damage</td>
</tr>
<tr>
<td>ILS1</td>
<td>Cytoplasmic isoleucine tRNA synthetase</td>
<td>tRNA synthetase</td>
</tr>
<tr>
<td>UGP1</td>
<td>UDP glucose pyrophosphorylase</td>
<td>Catalyzes the reversible formation of UDP Glc from glucose 1 phosphate and UTP</td>
</tr>
</tbody>
</table>

**Table 9.** Proteins Exhibiting Significant Change in SILAC Ratios in H4K16R.
<table>
<thead>
<tr>
<th>Gene Ontology Term</th>
<th>H3K4A</th>
<th>H4K16Q</th>
<th>H4K91Q</th>
<th>H4K16R</th>
</tr>
</thead>
<tbody>
<tr>
<td>oxidation reduction</td>
<td>oxidation reduction</td>
<td>nitrogen compound</td>
<td>regulation of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>biosynthetic process</td>
<td>translation</td>
<td></td>
</tr>
<tr>
<td>response to temperature stimulus</td>
<td>purine</td>
<td>amine biosynthetic</td>
<td>de novo IMP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ribonucleoside</td>
<td>process</td>
<td>biosynthetic process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>monophosphate biosynthesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>purine nucleoside</td>
<td>cellular metabolic</td>
<td>nitrogen compound</td>
<td></td>
</tr>
<tr>
<td></td>
<td>monophosphate biosynthesis</td>
<td>compound salvage</td>
<td>biosynthetic process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>purine nucleoside</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ribonucleoside</td>
<td>oxidization reduction</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>metabolism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>response to abiotic stimulus</td>
<td>nitrogen compound biosynthesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nucleotide biosynthetic process</td>
<td>purine nucleoside</td>
<td>cellular amino acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>monophosphate metabolism</td>
<td>biosynthetic process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nucleobase and nucleic acid biosynthesis</td>
<td>GMP biosynthetic process</td>
<td>deoxyribonucleotide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nucleoside and nucleotide biosynthesis</td>
<td>generation of precursor</td>
<td>e metabolic process</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>metabolites and energy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regulation of translation</td>
<td>GMP metabolic process</td>
<td>organic acid biosynthetic process</td>
<td></td>
<td>tRNA aminoacylation for protein translation</td>
</tr>
<tr>
<td>posttranslational regulation of gene expression</td>
<td>cell redox homeostasis</td>
<td>carboxylic acid biosynthetic process</td>
<td></td>
<td>tRNA aminoacylation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>generation of precursor</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>metabolites and energy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nitrogen compound biosynthetic process</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 10.** Gene Ontology for Yeast Histone Mutant Proteome Changes. Top 10 GO term matches for list of proteins showing significant expression change for yeast strains H3K4A, H4K16Q, H4K91Q, and H4K16R.
<table>
<thead>
<tr>
<th>Pathways</th>
<th>H3K4A</th>
<th>H4K16Q</th>
<th>H4K91Q</th>
<th>H4K16R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycolysis/ Gluconeogenesis</td>
<td>Purine metabolism</td>
<td>Purine metabolism</td>
<td>Amino-tRNA biosynthesis</td>
<td>Purine metabolism</td>
</tr>
<tr>
<td>Amino/nucleotide sugar metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histidine metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galactose metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatty acid biosynthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11. DAVID Pathway Analysis. Pathways associated with proteins showing significant expression change for yeast strains H3K4A, H4K16Q, H4K91Q, and H4K16Q.

Figure 21. Panther Assessment of Molecular Function. Protein expressions of significant change were assessed for molecular function using the Panther online tool.

Correlations Between RNA Transcription and Protein Expression

106
Levels of protein expression was next compared to mRNA transcript levels using available published data. Dion et al. performed a series of microarray experiments with several yeast histone 4 mutants. Of the mutants they examined we performed SILAC labeled mass spectrometry on H4K16R. We plotted protein expression level against the reported mRNA transcript level (Figure 22). While most protein expression levels correlated with detected mRNA levels a small subset showed increased level of mRNA compared to decreased level in protein abundance. To validate the microarray experiment, three genes that showed a discrepancy between the mRNA and protein levels were selected (SUI1, RTC3 and HSP12) and TaqMan Real Time PCR was performed. SUI1 is a protein involved in the recognition of the initiator codon and modulates translational accuracy at the initiation phase. RTC3 is associated with cellular stress response, deletion of this gene leads to lower levels of Heat Shock Proteins (HSP). Finally, HSP12 is a heat shock protein associated with cell membrane stability in high temperatures. Validation using TaqMan PCR revealed that while the mRNA levels remain unchanged the protein levels decreased, hence suggesting either a decrease in translation or loss of protein stability due to the alterations in histone PTM regulation (Figure 23). In depth GO term search for total proteomic data of H4K16R revealed that proteins showing a significant change in expression level match with several processes that can affect overall protein level (Table 12).
Figure 22. Protein Expression Levels vs. mRNA Levels in H4K16R. Protein expression data from the mass spectrometry experiment is compared to previously published microarray data for the H4K16R histone mutant (Dion et al., 2005)
Figure 23. Comparison of Mass Spectrometry, Microarray, and RT PCR Data. Transcription and translation of select genes are compared using data from mass spectrometry (protein), microarray (RNA), and TaqMan RT-PCR (RNA). Note: Microarray data is not available for H4K16Q. (Dion et al., 2005)
Table 12. Gene Ontology Terms for H4K16R Protein Expression Changes. DAVID gene ontology enrichment reveals that protein changes associated with H4K16R are greatly involved in protein translation.
Polyribosome profiling was performed to assess ribosome binding to mRNA transcripts of the target genes in order to determine if binding deficiencies have occurred due to mutation of PTM sites of histones. Polyribosome profiling of the wild type, H4K16R and H4K16Q initially revealed an elevation in the peak representing the 80S ribosomal subunit in the H4K16R mutant (Figure 24A). This prompted us to extract the transcripts from every other polyribosome fraction to assess the distribution of the target gene transcripts (Figure 24B). RT-PCR gene expression values were normalized to the firefly luciferase internal standard placed in each sample. The distribution of the target gene transcripts SUI1, RTC3 and HSP12 did not show a notable difference between the control transcripts from actin. Additionally, the mutant transcript distribution was not noticeably different from the wild type. An independent mutant, H3K56R, was tested in the same experiment to determine if this observed distribution pattern is consistent (Figure 25). The H3K56R polyribosome plot also showed a relative increase in the free 80S ribosomal subunit as observed in the H4K16R yeast mutant. However, the distribution of the three selected target genes remain consistent with both the actin control as well as the wild type yeast samples.
Figure 24. Polyribosome Profiling of H4K16. A) Polyribosome Profile Analysis of H4K16. Polyribosome profiling comparison of the wild type, H4K16Q, and H4K16R reveals a notable increase in the amount of 80 S ribosomal subunit suggesting a decrease in translational initiation in H4K16R. B) TaqMan Real Time PCR of Polyribosome Fractions. qPCR of polyribosome fractions was performed on genes HSP12, RTC3, and SUI1 to assess mRNA distribution.
Figure 25. Polyribosome Profiling of H3K56R. A) Polyribosome Profile Analysis of H3K56R. Polyribosome profiling comparison of the wild type and H3K56R reveals a notable increase in the amount of 80 S ribosomal subunit suggesting a decrease in translational initiation in H3K56R. B) TaqMan Real Time PCR of Polyribosome Fractions. qPCR of polyribosome fractions was performed on genes HSP12, RTC3, and SUI1 to assess mRNA distribution.
Discussion

The processes of converting the information stored in DNA to RNA transcripts to functionally active proteins is a complicated process with several different levels of regulation. It is well known that histone post-translational modifications can regulate the transcription of DNA into RNA through changes in chromatin structure and degrees of association with DNA, controlling access of RNA polymerase to DNA. Additionally, histone post-translational modifications alter the charge and structural environment facilitating or inhibiting the interactions of other regulators. Little study has been conducted on whether histone PTM states can regulate transcript translation as well. When global mass spectrometry proteomic data compared to microarray data suggested a translational regulatory effect in addition to what was previously described we pursued the possibility. There are several processes which can lead to the observed effect which will be discussed in this section.

The initial hypothesis we tested was assessing for a deficiency in translational initiation. It is known that histone PTMs provide binding scaffolds for a variety of regulators, thus it is possible that one of these regulators can modify newly transcribed mRNA. These modifications facilitate the binding of ribosomes serving as a promoter for translation initiation. The increase in mRNA transcripts can be explained by the lack of protein product causing feedback by the cell to increase transcription. Initial data from polyribosome profiling experiments supported this hypothesis as the 80S ribosomal peak appeared increased in the H4K16R samples. An explanation for this observation is
deficiencies in translational initiation leads to a buildup of the 80S (combined 40S and 60S ribosomes) on the transcript without forward translation. However, assessment of the distribution of transcript for the three selected target genes SUI1, RTC3 and HSP12 and actin as a control revealed that there was no substantial difference in the distribution of transcripts in the polyribosome fractions. Furthermore, the distribution of the target genes in the yeast histone mutant were similar to that of the wild type. This data shows that polyribosomes (multiple ribosomes on each transcript) were forming regardless of gene (target or control) or phenotype (histone mutant or wild type). The formation of these polyribosomes does not support the hypothesis as the presence of polyribosomes suggests there is no initiation deficiency in the histone mutants.

A second hypothesis explaining the discrepancy is protein stability. Lower abundance of a protein can be caused by decreased stability leading to degradation. This hypothesis is supported by the results of the SILAC global proteomic study of H4K16R. DAVID analysis for GO terms of proteins showing two fold or greater change in abundance between the wild type and histone mutant were clustered mapped to protein folding. One important example of a protein folding chaperone is HSP12 which we observed downregulated in the H4K16R mutant. Proteins that are misfolded are targeted for degradation through various pathways. Once again the increase in mRNA transcripts may be caused by a cellular feedback mechanism responding to the low protein level. A future goal would be to test this hypothesis using a pulse-chase experiment with TAP tagged proteins.
A third alternative hypothesis is the possibility that while translational initiation and ribosome loading on the mRNA is normal, there may be delays or interruptions in the translational elongation process. Global yeast proteomic data indicated that tRNA aminoacylation was downregulated in the H4K16R mutant. Without the necessary tRNA bound to the corresponding amino acid the ribosome cannot advance. Thus, while the amount of polyribosomes bound to the mRNA transcripts remain constant, increased time is needed for the translational machinery to translate the protein. This hypothesis fits the data observed in the polyribosome profile traces. The increase 80S subunit, the state where one ribosome is bound to the mRNA transcript, takes more time to move forward to allow for a second ribosome to attach to the transcript to begin the translation process. The deficiency in properly charged tRNA will affect specific proteins that require the missing amino acid. Again, the increase in transcript level is the cellular feedback mechanisms are increasing transcription or delaying mRNA degradation in response to the lack of the affected proteins. A future goal would be to test this hypothesis by determining the rate of protein synthesis using radiolabeled methionine and measuring its incorporation into newly synthesized proteins.

**Conclusion**

Ultimately this project demonstrates the incredible complexity of the biological regulation process of converting the information stored in DNA to the proteins that perform the functions of cellular processes. Our experimental results showed a discrepancy between mRNA transcript levels and measured protein levels. We hypothesized that this
discrepancy is due to a post-transcriptional regulatory effect of the mRNA due to the acetylation patterns of histones. However, this hypothesis was not supported by real-time PCR of the polyribosome fractions of wild type and mutant samples. That experiment suggested there was no difference in the polyribosome formation between wild type and H4K16R yeast mutants. It may be worthwhile in the future to test the two alternative hypotheses: differences in protein stability and rate of protein translation for the affected proteins.
Summary

In the course of my graduate studies, experiments focused on the overarching themes of biomarker and therapeutic target discovery. The two main lines of experimentation pursued involved yeast modeling of histone post-translational modifications and global proteomics of thymomas, a rare epithelial tumor that has not been previously characterized in proteomics.

Patient tissue proteomics yielded exciting results paving the way for future studies. Highlights of this line of research include the development of a new method to collect and enrich for patient tissue of interest archived over decades on their native positively charged slides and the description of a marker, desmoyokin, unique to the medullary regions of the thymus while absent from the cortical regions. Thymoma is a tumor divided into 5 epithelial tumors with A (medullary derived) types and B types (cortex derived). This marker shows great potential in utility as an objective method to distinguish A and B thymomas which sometimes may have histologically overlapping features. This is especially useful when identifying thymomas A and B3 which have significantly different clinical outcomes.

Yeast histone PTM studies initially yielded interesting results suggestive of a previously undescribed mechanism of translational regulation associated with different states of histone acetylation primarily at histone 4 lysine 16. Through mass spectrometry
of global cellular lysate we observed downregulation of proteins. Previously published microarray data corresponding to the mRNA transcripts of our detected protein was used in quantitative comparison. It was observed that some proteins while downregulated had elevated mRNA transcripts. This microarray results were validated using TaqMan real time PCR for genes HSP12, RTC3 and SUI1. However, through polyribosome profiling and subsequent real time PCR of resulting fractions did not show a deficiency in the translational initiation. Due to constraints of testing for rate of translation and protein half-life this studied was concluded at this point.

In conclusion, with the availability of high throughput proteomic research techniques it is possible to observe the big picture of the processes occurring in cells and tissue. Through analysis of differences in mutant vs. wildtype or disease vs. normal, new hypothesis generation occurs leading to innovations in the search for diagnostic biomarkers and therapeutic targets.
References


Kim, Y. S., Hwan, J. D., Bae, S., Bae, D. H. & Shick, W. A. Identification of differentially expressed genes using an annealing control primer system in stage III


Appendix A: Future Applications of Yeast Epigenetics

Histones are alkaline proteins that wrap DNA for compaction such that all of the genomic components can fit inside the nucleus of the cell. There are four histones H2A, H2B, H3 and H4 which are known as the core histones. Two of each core histone forms the histone octamer which DNA is wrapped around. Histone H1 is known as a linker histone, regulating the entry and exit point of DNA into the histone/DNA complex, known as the nucleosome. Histone tails play an important role in the chromatin regulation. One specific example is the histone H4 tail which contains 4 lysine residues on which acetylation is known to occur, at lysines 5, 8, 12 and 16. Increase in acetylation of the H4 tail decreases its affinity for DNA, allowing the cell to regulate the chromatin structure. Of particular interest in the H4 tail is the H4K16 residue. Acetylation of H4K16 leads to the open chromatin state, facilitating transcription. This is mediated mainly by the histone acetyltransferase MOF (Sas2 in yeast). Disruption of this gene has implications associated with cancer.

In Chapter 4, the variety of histone PTMs that play a role in epigenetic regulation was introduced focusing entirely on acetylation primarily in Histone 4. It is important to note that other histone PTMs play a crucial role as well. Histone methylation on lysine residues can occur involving one to three methyl groups as mono-, di- and tri-methylation respectively. Methylation may play a role in activating or repression of genes depending
on the specific location of the modified amino acid. Disruption of histone methylation levels have been documented to be associated with human disease. A much studied example of histone methylation involves pontine gliomas where it was discovered that loss of trimethylation on H3.3K27 due to mutation of lysine to methionine and isoleucine strongly correlated with disease phenotype.\textsuperscript{208,209} H3K36M has been described as an oncogenic mutation occurring in head neck squamous cell carcinoma and colon cancer leading to decreased overall methylation at the H3K36 site.\textsuperscript{210} Additionally, sites where loss of methylation has been associated with cancer include H3K4. Interestingly, both H3K36 and H3K4 trimethylation has been associated with gene activation, the modification is often lost in association with cancer.

Crosstalk between histone acetylation and methylation has been previously documented. For example, H3 acetylation positively correlates with the level of methylation found on H3K4.\textsuperscript{211} Another histone PTM site that is of interest for this ongoing investigation is histone H3K79, which is the site for mono-, di- and tri-methylation which is involved in crosstalk with other histone residues. Methylation of H3K79 by Dot1 reduces Sir protein binding, leading to decrease of silent chromatin.\textsuperscript{212} Histone crosstalk has been extensively studies and reviewed over the last several years. Post-translational modifications include acetylation, methylation, ubiquitination and phosphorylation on specific amino acid residues. One PTM excludes the presence of another on the same residue. Crosstalk between different histone PTMs can occur either cis- or trans-, on same or different histones.\textsuperscript{213} One key example of histone crosstalk is the mono ubiquitination of H2BK123 is involved with the methylation patterns of H3K4.\textsuperscript{214} In recent years, groups
were able to quantify combinations of K27 and K36 PTMs coexisting on histone H3 using mass spectrometry.\textsuperscript{215,216} This advancement made it possible to observe histone crosstalk on the same peptide. However, much remains to be understood in the realm of the histone crosstalk networks.

A previous study in our laboratory assessed the PTM states of several histone residues using a series of 44 yeast mutants modified at sites of known histone PTM regulation such as H3K56 (acetylation) and H3K79 (mono-, di- and tri- methylation).\textsuperscript{10} This study noted a decrease in the tri- methylation of H3K79 in histone H4 mutants at K12 and K16 when the lysine is mutated to glutamine to simulate the constitutively acetylated PTM state. The authors proposed the possibility of H4K16 acetylation leading to the recruitment of Sir2-4, playing a role in the sequential formation of the Sir complex ultimately leading to the silent chromatin.\textsuperscript{217} This process may eventually lead to the changes in H3K79 methylation levels. However, an alternative hypothesis regarding the interplay between H4K16 acetylation and H3K79 methylation was proposed by Altaf et al.\textsuperscript{218} The mechanism proposed is that H4K16ac leads to the displacement of Sir3 which allows competitive binding of Dot1, the histone methyltransferase associated with H3K79 methylation, facilitating this process. However, this does not account for the loss of tri-methylation associated with the mutation of H4K12 to the constitutively acetylated state. It is further hypothesized that on the Histone 4 N-terminal tail, acetylation states of K12 and K16 play a role in the level of H3K79 methylation through charge mediated interactions with Dot1. A study by Fingerman et al. showed that a basic patch formed by R17, H18 and R19 on the N-terminal tail of histone H4 was necessary for the activity of
Dot1 through interaction with an acidic domain on the methyltransferase.\textsuperscript{219} While this study included the H4K16 mutants Q and R, his conclusion was that the basic properties of the lysine in K16 did not play a role in the trimethylation of H3K79 as measured by Western blotting. We propose to reexamine this crosstalk to follow up on our previous mass spectrometry study.

**Preliminary Studies**

Original study design aimed to examine whether the positioning of nucleosomes is altered as a result of H3K79 trimethylation changes due to potential crosstalk with the histone H4 N-terminal tail. H4K16 acetylation and H3K79 trimethylation are both activating marks, however, data suggests that the level of these PTMs are inversely proportional. Briefly, the study design utilizes ChIP-seq with antibodies targeting H3K79me3 to identify the genes associated with the modified histones. An additional study design involves the collection of inter-nucleosome DNA for sequencing to determine which genes are accessible due to histone positioning and the changes which occur due to simulated histone PTMs.

**Yeast Mutant Creation**

Yeast histone mutant plasmids were created by the laboratory of Dr. Mark R. Parthun. For this series of studies yeast histone mutants were created for the acetylation sites on the H4 N-terminal tail: H4K5, H4K8, H4K12 and H4K16. A detailed processes of the design and creation of the yeast mutants appears in Chapter 4.
**Yeast Growth**

Yeast cells were grown in 200 mL of YPD media until reaching 0.45-0.50 optical density at 600 nm light absorbance. The histones were crosslinked to the DNA through 37% formaldehyde (5.5 mL for 200 mL of cell culture) for 15 minutes at 30º C at constant shaking. Following the crosslink time the formaldehyde reaction was quenched using 2.5 M glycine (11 mL for 200 mL cell culture) for 10 minutes shaking at 30º C. Cells were pelleted through centrifugation at 4700 g for 5 minutes at 4º C. Immediately after cell collection the cells were washed once and re-pelleted.

**Chip-Seq Sample Preparation**

Yeast samples were prepared for Chip-Seq using a modified protocol published by Lefrançois et al.\textsuperscript{220} Briefly, collected cells were mixed with 1 mL lysis/IP buffer (50 mM HEPES/KOH pH7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate) and approximately 100 µL glass beads. The mixture was vortexed for 10 cycles (1 minute vortex/1 minute on ice). Lysate was collected and isolated from the beads through centrifugation into a larger collection tube (5 mL Eppendorf) at 400 x g for 3 minutes. DNA was sonicated for 5 minutes (30 second on/30 second off) at 45% amplitude using a Fisher Scientific Sonic Dismembrator Model 500.

During optimization, the sonicated samples were treated with proteinase K at a concentration of 0.4 mg/mL, then purified using phenol/cholorform. DNA was precipitated overnight with the addition of 60 µg glycogen and 45 µL of 45 M LiCl in 100% ethanol. The pellet was collected through centrifugation and washed with 70% ethanol. Sonication
efficiency was assessed by running the DNA on a 2% agarose gel and visualized using a LiCor Odyssey FC system.

**Preliminary Results**

Early results suggest that there was a difference between the sonicated DNA extracts from wild type, H4K16R and H4K16Q mutants. Sonicated DNA extracts showed different profiles of DNA fragment lengths despite samples being subjected to the same sonication treatment. It was hypothesized that this difference could be caused differential protection of the DNA from sonication sheer due to differences in positioning of associated histones. Assuming that no crosstalk was occurring, H4K16Q was predicted to be associated with the open chromatin state. Thus, sonication should result in small sheared DNA fragments due to the exposed DNA. Unexpectedly, however, the data showed the opposite effect. H4K16R showed the smallest DNA fragments as a result of the shearing, wild type extracted DNA shows a mix and H4K16Q DNA shows the longest fragments. These data suggest that PTM status of H4K16 alone does not determine the organization of nucleosomes and that downstream crosstalk may be involved. Thus further investigation of the crosstalk between H4K16ac and H3K79me3 was warranted.

**Further Studies**

Due to the preliminary results collected while optimizing the ChIP-seq sample preparation, a micrococcal nuclease ladder experiment was added to assess for differential lengths in DNA between nucleosomes between the different yeast histone mutants. It was hypothesized that mutation of the H4K16 residue to simulate constitutive acetylation or
non-acetylation will lead to a change in H3K79 trimethylation ultimately leading to alteration in the nucleosomal positioning. It was also hypothesized that the proposed crosstalk between H4K16 and H3K79 is mediated through the basicity of the H4 N-terminal tail and that this regulation mechanism may also include H4K12.

**Yeast Cell Lysis and Micrococcal Nuclease Ladder**

Yeast cell walls are to be removed using Zymolase in spheroplasting buffer (1.2 M sorbitol and 20 mM HEPES pH 7.4) until 90% of the cell population are spheroplasts. Cells are lysed using lysis buffer (10 mM Tris-HCl pH 7.5 and 18% Ficoll). Following lysis, nuclei are isolated using a 60% sucrose gradient in wash buffer (10 mM Tris-HCl pH 8, 30 mM sodium butyrate, 0.5% NP-40, 1 mM PMSF and 75 mM NaCl). The nuclei are washed in Wash Buffer A (10 mM Tris-HCl, 50 mM NaCl, 1.0 mM MgCl₂) followed by Wash Buffer B (10 mM Tris-HCl, 400 mM NaCl, 1.0 mM MgCl₂). 2 Worthington units of micrococcal nuclease are used for each sample and a ladder optimization is to be performed prior to final experiments using digestion times for 15, 20, 25 and 30 minutes. The digestion reaction is quenched using 0.5 M EDTA. Following digestion, the resulting DNA is purified using phenol/chloroform. DNA is precipitated overnight at 4º C in ethanol. Collected DNA is run on a 2% agarose gel (with ethidium bromide) and imaged using a LiCor Odyssey FC system.

**Chip-Seq of Exposed and Chromatin Bound DNA**
Samples are prepared as stated above. For ChIP-Seq assessment of exposed DNA, the protein bound DNA is removed through centrifugation followed by phenol/chloroform purification and collected for assessment. It should be noted that this approach may prove to be challenging due to the variable lengths of internucleosomal DNA and the possibility that the sizes are incompatible with the Illumina system available at The Ohio State University Nucleic Acids Core Facilities.

For the collection of DNA associated with nucleosomes the sonicated lysate is centrifuged prior to immunoprecipitation. ChIP is performed using Protein A/G magnetic beads (Pierce) bound to rabbit polyclonal H3K79me3 antibody (Abcam ab2621) per manufacturer protocol. Immunoprecipitation is performed overnight at 4°C. Sample is eluted off the beads using elution buffer (1 x TE/ 1% SDS solution) and washes a second time using 1 x TE/ 0.67% SDS solution. Proteinase K treatment and DNA precipitation is as described in the previous optimization section.

**RT-PCR Validation of Changes in Exposed DNA**

Validation of study results will be conducted using real time PCR. Probes will be designed to target genes with differential association with the nucleosomes as determined by the ChIP-Seq experiments. RT PCR will be performed using an Applied Biosystems Real Time PCR system.

**Western Blot Validation of Changes in H3K79me3 Level**
In validation that the level of H3K79me3 has been altered as a result of the modification state of H4K16, total cell lysate is run on a 15% acrylamide gel. The resulting Western blot will then be probed with the H3K79me3 antibody (Abcam) to determine and compare the levels of this modification between the different yeast histone mutants.