ABSTRACT

NEPOMUCENO, ANGELITO INOCENCIO. Development and Application of Discovery-Based Proteomic Strategies Needed for the Study of Small Populations of Cells in Model Biological Systems. (Under the direction of Dr. David Charles Muddiman).

In the era of "-omics" based technologies, mass spectrometry (MS) has been synonymous with proteomics. This has only been accomplished with the advent of electrospray ionization and matrix-assisted laser desorption ionization techniques, whom Dr. John B. Fenn and Dr. Koichi Tanaka received the Nobel Prize in chemistry in 2004. A variety of proteomics based mass spectrometric workflows are utilized to define an organism's proteome. Each workflow is dependent on many necessary tools. First, a reference database comprised of proteins, expressed sequence tags and/or a genome sequence database. Secondly, analytical separation techniques are necessary for complex mixtures. A wide variety of techniques can be used to separate complex mixtures, *i.e.* SDS-PAGE, HPLC, Filter Aided Sample Preparation, and Stage-tip fractionation. To analyze these complex mixtures, mass spectrometers that can provide the necessary sensitivity, robustness and accuracy are necessary. Lastly, the identification of proteins is then feasible with a series of software suites that match the MS data to that of the protein sequences in the provided database. Herein, this dissertation describes several approaches used to define global proteomes of tissue samples obtained from model biological systems.

Animal models are essential towards advancements in basic and clinical research. The domestic hen (G*allus gallus*) was used as an animal model to

investigate ovarian cancer and basic reproductive studies. There have been mammalian models which have been used to investigate ovarian cancer. However, scientists have not found a suitable animal model that spontaneously develops ovarian cancer. The chicken on the other hand not only develops ovarian cancer spontaneously, the types of cancers found are similar to those that occur in humans. The ovaries of egg laying hens were also used to study follicular development and recruitment. Small white follicles were excised from the cortex of the ovary in order to obtain global proteome profiles of the white yolk, follicular wall and stromal cells. The proteomes provided were used to investigate follicular recruitment.

Proteins extracted from tissue samples were processed in several methods prior to analysis via LC-MS/MS. Protein samples obtained from the ovarian cancer study were processed on a SDS-PAGE. Gel lanes were fractionated in order to maximize the proteome coverage. Samples were then analyzed using a nanoLC coupled to a LTQ-FTICR mass spectrometer. In a short number of years, many advancement in techniques used for sample preparation were conceived. Limited by the amount of samples obtained using laser microdissection experiments; a filter aided sample preparation (FASP) protocol was used. To further increase proteome coverage, samples were analyzed on a linear quadrupole orbitrap mass analyzer due to the increase in sensitivity, mass accuracy of both MS and MS/MS spectra, and speed of analysis.

Development and Application of Discovery-Based Proteomic Strategies Needed for the Study of Small Populations of Cells in Model Biological Systems

by
Angelito Inocencio Nepomuceno

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Chemistry

Raleigh, North Carolina

2014

APPROVED BY:

_____	_____
David C. Muddiman Professor, Chemistry Committee Chair	James N. Petitte Professor, Chemistry
_____	_____
Gufeng Wang Professor, Chemistry	H. Troy Ghashghaei Professor, Molecular Biomedical Sciences

DEDICATION

I'd like to dedicate this dissertation to my grandfather Antonio C. Nepomuceno. Without his unconditional love, support, guidance, and encouragement this journey would have never begun. I know he's watching over me.

BIOGRAPHY

Angelito Inocencio Nepomuceno was born in Queens, New York on January 8, 1978 to his parents Angelito H. and Carolina Nepomuceno and is the eldest of two children. Angelito attended Staten Island Academy from kindergarten to high school and graduated in 1996. After high school he chose to attend Stetson University in Deland Florida. He earned two Bachelor of Science degrees in the fields of Chemistry and Biochemistry. He then moved to Richmond Virginia and started his graduate career in 2000 at Virginia Commonwealth University where he obtained his Master of Science degree in Analytical Chemistry under the mentorship of Dr. David C. Muddiman. After his master's thesis, Angelito went to work in the Mass Spectroscopy industry at IonSpec Inc. in Lake Forest, California. After 6 wonderful years of in California, Angelito began his Ph.D. career in Raleigh, North Carolina at North Carolina State University in 2010 under the direction of Dr. David C. Muddiman.

ACKNOWLEDGMENTS

I would like to firstly acknowledge my family; my parents Angelito H. and Carolina and my sister Angela for their inspirational support and encouragement. Without your loving support I would not have made it this far.

There have been many teachers, professors, and role models who have been instrumental in my life, and I sincerely thank each one of you. However, I would like to individually thank those who have directly impacted my scientific career. During my undergraduate education at the Stetson University, I was fortunate enough to become involved in undergraduate research during my senior year. It is here that I realized the passion I have for research. Dr. Ramee Indralingam was helpful, motivating, and provided a perspective on science that is not found in the classroom alone. She was the person who most encouraged me to apply to graduate school.

When I first started at VCU, I knew from the start that I wanted to work with Dr. David Muddiman. I was excited to be part of his research program. I'm not going to lie; working with a motivated and brilliant young professor like Dr. Muddiman was tough. I truly learned from him what it takes be a successful graduate student.

I then went to work for IonSpec Inc. which was later purchased by Varian Inc. I interacted with many people during my career. There are just too many people to name; I want to thank them all for their support even when I chose to go back to graduate school.

When the decision was made to go back to graduate school to obtain my Ph.D., I knew I wanted to work for Dr. David Muddiman again. This second go around was a much different experience. I came in with the motivation to learn and was given the opportunity to grow as an independent researcher and develop critical thinking skills. I would also like to thank Dr. Adam Hawkridge and Dr. James Petitte for their support and patients. Then the Muddiman alumni and group members who have been tremendously supportive over the past several years and I can't thank them enough.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS

1. Angelito Nepomuceno; David C. Muddiman; James N. Petitte, Global Proteomics Analysis of Functional Compartments in Immature Avian Follicles Using Laser Microdissection Coupled to LC-MS/MS, *J. Biol. Reproduction*. (submitted, June, 2014)

2. Jonathan Hall; Michael Bereman; Angelito Nepomuceno; Elizabeth Thompson, David Muddiman; Robert Smart, C/EBPα prevents the proteasomal degradation of $p21^{WAF1/CIP1}$ in response to UVB radiation and controls subsequent p21/CDK interactions and the DNA damage-induced G_1 checkpoint. Nucl. Acid Res. (submitted 4/9/2014)

3. Hsi-Chuan Chen; Jina Song; Jack P. Wang; Ying-Chung Lin; Joel Ducoste; Christopher M. Shuford; Jie Liu; Quanzi Li; Rui Shi; Angelito Nepomuceno; Fikret Isik; David C. Muddiman; Cranos Williams; Ronald R. Sederoff; Vincent L. Chiang. Systems Biology of Lignin Biosynthesis in *Populus trichocarpa*: Heteromeric 4-Coumaric Acid:Coenzyme A Ligase Protein Complex Formation, Regulation, and Numerical Modeling. Plant Cell **2014**, 26, 876-893

4. Angelito I. Nepomuceno; Radiance J. Gibson; Shan M. Randall; David C. Muddiman. Accurate Identification of Deamidated Peptides in Global Proteomics Using a Quadrupole Orbitrap Mass Spectrometer. *J. Prot. Res.,* **2014**, 13(2), 775-785

5. Justin Schilling; Angelito Nepomuceno; Jennifer E. Schaff; David C. Muddiman; Harry V. Daniels; Benjamin J. Reading. Compartment Proteomics Analysis of White Perch (Morone Americana) Ovary Using Support Vector Machines. *J. Prot. Res.* **2014**, 13(3), 1515-1526

6. Brittany L. White; Emine Gökce; Angelito I. Nepomuceno; David C. Muddiman; Timothy H. Sanders; Jack P. Davis. Comparative Proteome Analysis and IgE Binding Properties of Peanut Seed and Testa (Skin). *J. Agric. Food. Chem.* **2013**, *61*(16), 3957-3968.

7. Angelito I. Nepomuceno; Huanjie Shao; Kai Jing; David C. Muddiman; James N. Petitte; Michael O. Idowu; Xianjun Fang; Adam M. Hawkridge. Ovostatin 2 (OVOS2) is Up-regulated in Chicken and Human Ovarian Cancer. *Proteomics Clinical Applications.* **2013** (Submitted)

LIST OF PRESENTATIONS

1. **Poster** - "Global Proteome Analysis of Wild-Type and FoxJ1 knock-out Mouse Brain Tissues Using a Quadrupole Orbitrap Mass Spectrometer" Radiance J. Gibson; Angelito I. Nepomuceno; Nagendran Muthusamy; Shan M. Randall; Philip L. Loziuk; H. Troy Ghashghaei; David C. Muddiman. *61st American Society for Mass Spectrometry Conference*. Minneapolis, Minnesota June 2013.

2. **Poster** - "Mass Spectrometry-Based Proteome Characterization of Laser Microdissected Ovary and Oviduct Tissues from Chickens with Ovarian Cancer" Angelito I. Nepomuceno, James N. Petitte, Adam M. Hawkridge, David C. Muddiman. *61st American Society for Mass Spectrometry Conference*. Minneapolis, Minnesota, June 2013.

3. **Oral** - *"*Proteome Characterization of Chicken Ovary Tissue and Plasma using One-Dimensional SDS-PAGE nLC-MS/MS*"* Angelito I. Nepomuceno, James N. Petitte, David C. Muddiman, Adam M. Hawkridge. *63rd Southeastern Regional Meeting of the American Chemical Society*. Richmond, Virginia, October 2011.

4. **Poster** – "In-Depth Proteome Characterization of healthy and Cancerous Chicken Ovarian Tissues" Angelito I. Nepomuceno, James N. Petitte, David C. Muddiman, Adam M. Hawkridge. *59th American Society for Mass Spectrometry Conference*. Denver, Colorado, June 2011.

5. **Poster** - ""Proteome Characterization of Chicken Ovary Tissue and Plasma using One-Dimensional SDS-PAGE nLC-MS/MS" Angelito I. Nepomuceno, James N. Petitte, David C. Muddiman, Adam M. Hawkridge. *7th Annual United States Human Proteome Organization Conference*. Raleigh, North Carolina, March 2011.

CHAPTER 1

An Introduction to Quantitative Global Proteomics Measurements of Tissue Lysate Samples by Liquid Chromatography Tandem Mass Spectrometry

1.1 Chicken Model

1.1.1 Ovarian Cancer

Ovarian cancer (OVC) is among the most lethal gynecological cancers in the Western World due to a combination of ineffective early-stage detection methods and late-stage treatment strategies. More than 22,000 women are diagnosed with OVC per year, approximately 70% of which present with advanced stages of the disease (Stage III and IV) when surgical intervention and chemotherapeutic treatments are least effective. Early diagnosis is the single most important determinant for survival, with 5-year mortality rates less than 10%, for women with Stage I OVC. Significant effort has been focused on identifying early-stage markers for OVC including trans-vaginal ultrasound, blood-based proteins (e.g., *CA-125* and *HE-4*), metabolites (e.g., *LPA*), and tissue-specific genetic alterations (e.g., *p53*).[1] Although these detection methods have important clinical value for managing OVC, they do not possess the predictive power necessary to be useful in a population-based screen.[2] OVC is a heterogeneous disease with four main histopathological subtypes including serous, endometriod, clear cell, and mucinous.[3-5] Each tumor subtype has distinct cellular morphology, genetic abnormalities, and malignant potential all of which contribute to the complexity of finding a common, early-stage marker for OVC.

The domestic chicken is an emerging experimental model for OVC since they develop OVC tumors spontaneously[6,7] with similar morphological[8] and molecular similarities[9-19] to humans. Important molecular-level similarities include *CA-125* expression[12], frequency of *p53* mutations[10], and E-cadherin up-regulation[19]. OVC prevalence in the aging hen can exceed 35% depending on the birds' age, genetic background (i.e., strain), and number of eggs produced.[7] There are several advantages to studying OVC in the chicken, such as a well-defined genetic background that is maintained in the agricultural industry, a predictable OVC onset window, control over environment and diet, and a well-defined genome.

For biomarker discovery, these advantages are critical along with additional strengths such as, the ability to draw large amounts of blood longitudinally with matched tissue samples from the ovary, oviduct, and neighboring tissues in the abdominal cavity. Longitudinal sampling facilitates the determination of both intra- and inter-individual variability, the former of which has been shown to be more sensitive to OVC detection using *CA-125*.[20-24] A recent study generated a large-scale biorepository that included longitudinal plasma samples, matched ovary, and oviduct tissue samples, from approximately 250 2.5 year old birds over a one-year period.[9] From this biorepository, extensive global and targeted proteomics investigations have been carried out. A critical finding in this earlier work was the identification of a predicted form of ovomacroglobulin that increased significantly with the onset of ovarian cancer. Importantly, the companion form of this protein is

expressed solely in the oviduct raising important questions about the origin of the ovarian tumors in the study.[25]

1.1.2 Follicular Development

In chickens only the left ovary and oviduct develop. Ovulatory cycle take approximately 24-26 hours in length depending on the age of the chicken. The avian ovary contains several hundred of follicles arranged in a hierarchy system based upon size. There are 5 preovulatory follicles (F1-F5) ranging in roughly 9 mm to 40 mm in diameter. When the F1 is released into the oviduct, the preovulatory follicles (F2-F5) mature to the next size to take place of the F1. While a new F5 follicle is recruited from a pool of small yellow follicles. The small yellow follicles are replenished from a pool of large white follicles (2-5 mm) in size. The smallest of follicles and the most immature are the small white follicles that are less than 2 mm.

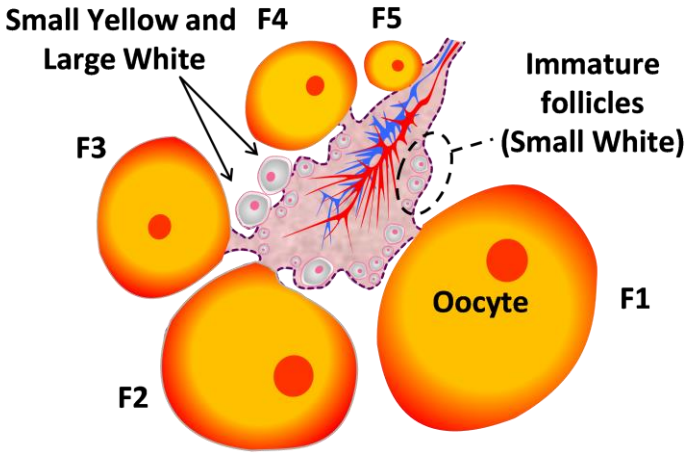


Figure 1.1 – Illustration of the chicken ovary, and the follicle classification.

and are embedded within the cortex of the ovary. **Figure 1.1** is an illustration of the chicken ovary detailing the follicle classifications.

Many visible changes occur to the follicle during follicular development, *ie*. color and size. Researchers have been studying the cellular changes that occur in follicles during development. There are two processes in which selection of the follicles occur in development. The first is within the initial recruitment of small white follicles which are embedded within the ovarian cortex. It is not well understood the signaling that occurs for premature follicles to enter the next stage of development. The second is the cyclic recruitment (or follicle selection) at which the follicle enters a state of differentiation and rapid growth occurs before ovulation.[26] It is these signaling pathways that have become heavily studied in more recent years. In particular, the stroidogenic pathway has been heavily examined due to the particular pathways taken by various follicles during development.[27] For example, small follicles will produce dehydroepiandrosterone (DHEA), estradiol and androstenedione through the Δ5 pathway. While steroidogenesis shifts from the Δ5 to the Δ4 pathway in large yolk follicles.[27]

1.2 *Sample Preparation of Tissue Lysate prior to Mass Spectrometry*

1.2.1 In-Gel Digestion

In 1996, Shevchenko and co-workers incorporated a method to enzymatically digest proteins from stained gels.[28] The complex protein mixture is loaded onto a one dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (1D-

SDS-PAGE) to fractionate the proteins by molecular weight. The gel lane is then cut into portions and an in-gel digestion is performed prior to analysis by LC-MS/MS. First an ionic detergent such as SDS allows for the protein to unfold while applying a negative charge. Proteins will then migrate through the pores of the polyacrylamide gel towards the positively charged electrode. Separation is then based on the protein's electrophoretic mobility.

Each gel lane is then cut into a number of fractions (12) to reduce the complexity of each sample prior to analysis. In order to extract the embedded proteins within the gel, a series of chemical modifications are used. Cysteine bonds between intramolecular and intermolecular proteins are reduced using dithiothreitol (DTT). To prevent disulfide bonds from reforming, thiols are then capped using iodoacetamide alkylating agent. This also allows for all cysteines to be searched as a carbamidomethyl fixed modification. A serine protease (trypsin) is used to cleave all arginines and lysines on the C-terminal end. Following digestion, the newly formed peptides are extracted from the gel. These fractions are then able to be analyzed using LC-MS/MS.

1.2.2 Filter Aided Sample Preparation (FASP)

The disadvantage of using in-gel digestion has been the low recovery of peptides from the gels. This method has also proven very tedious and time consuming when dealing with large scale experiments. Filter aided sample preparation (FASP) followed by stage tip fractionation has proven to be more

efficient method allowing for maximum recovery of peptides.[29] Protein extraction methods from whole tissue can require a variety of reagents, ie. urea, thiourea, DTT, and SDS. The sample lysates are placed on top of a regenerated cellulose filter. The samples solutions are spun down leaving the proteins on top of the filter and allowing all the reagents to pass through. It is very important to remove these regents as they can interfere with digestion. This is then followed by reduction, alkylation and digestion which all take place on top of the filter. After the tryptic digestion the peptides are allowed passage through the filter where they are collected for fractionation. Fractionation of peptides is accomplished using anion exchange resin that has been inserted into a pipette tip. The isoelectric point (*pI*) of the peptide determines which pH buffer the peptide will elute into. Six different buffers ranging from pH 3 – 11 are typically used resulting in six peptide fractions.[30]

1.3 RP-HPLC-nano-ESI-MS/MS

1.3.1 Reversed-Phase High Performance Liquid Chromatography (RP-HPLC)

In complex protein samples, to reduce the complexity of the samples a separation technique prior to MS analysis often performed. The most common separation coupled to a MS system has been liquid chromatography. A reversed phase (RP) column is frequently used. RP-HPLC separation is based upon the hydrophobicity of the peptides. Peptides are retained on an analytical column made typically of a resin or carbon particle or a silica particle which is bonded with a non-polar molecule, *e.g.* alkyl chains. The alkyl chains create a strong affinity towards

hydrophobic peptides. Peptides are then eluted from the column by increasing the concentration of organic solvent used. The properties that affect the resolution are particle size, pore size, bonded phase, and column length. The eluent of the LC can either be capture off-line or directed online with the electrospray tip.

1.3.2 Electrospray Ionization

Online RP chromatography and electrospray ionization (ESI) is the most commonly coupled technique for analyzing peptides. The soft ionization of ESI allows for multiple charging of biomolecules. The exact mechanism for ESI is much under debate, however the underlying theme for both arguments begin with a voltage at ~+2 kV applied to the capillary. The voltage induces a surplus of charges (H+) in positive mode from the oxidation of H_2O. Droplets are ejected from the electrospray emitter tip in the form of a "Taylor Cone".[31] These droplets begin to decrease in radius through desolvation and a series of coulombic fission processes. As these droplets begin to decrease in size, the charge density will increase to a point where it can no longer contain the charges, called the Rayleigh Limit[32]. Once this limit is reached progeny droplets are formed, with some containing analytes. The process in which these small charged droplets containing analyte molecules become gas phase ions has been heavily disputed. There are two existing models, the charge residue model (CRM) proposed by Dole *et. al.,*[33,34] and the ion evaporation model (IEM) stated by Iribarne and Thompson.[35-37] The Dole model suggest through a series of columbic fissions the droplet sizes will reach a radius

small enough for the solvent to evaporate leaving only a gas-phase ion. The IEM model hypothesized that the stress caused by the amount of charges on the surface of the droplet causes the ejection of gas phase ions. Evidence later presented by Fenn also supports the IEM model.[38] In his work Fenn presented supporting evidence behind the IEM by conducting electrospray of analytes with differing hydrophobic properties. Solvent, analyte, and excess charge are ejected from the surface of the parent droplet forming the progeny droplets. Since the more hydrophobic analytes will have a tendency to be at the surface of the droplet, they will be ejected from the parent droplet and go on to form gas phase ions. These ions will have a lower charge state due to the lower energy of solvation. The analytes that remain in the droplet will have a decreased charge to volume ratio causing a reduction in the chance to be ionized, which would make these ions have a higher charge state.

1.4 Mass Spectrometers

1.4.1 LTQ-FTICR-MS

Hybrid mass spectrometers such as the Linear Ion Trap Fourier Transform Ion Cyclotron Mass Spectrometer (LTQ FTICR-MS) has a linear ion trap front end coupled to a *FTICR* mass analyzer. The ion trap has two modes of operation; 1) a low resolving power and low mass accuracy mass analyzer with MS^n capabilities and 2) collection of ions to be injected into the FTICR mass analyzer. The high

magnetic field of a 7T magnet is used to maintain the ion at the center of the *ICR* cell. The ions are excited to a specific radius within the cell by applying an RF voltage through the excitation plates. After the ions are set in their cyclotron motion,[39] they will induce a current on the detection plates resulting in a transient. The transient is fast Fourier transformed to provide a spectrum in the frequency domain. **Equation 1** allows for the relationship between frequency and *m/z* based on the applied magnet field strength (B_0).

$$\omega = \frac{qB_0}{m} \qquad \textbf{Equation 1.1}$$

$$Resolving\ Power = \frac{m}{\Delta m} = \frac{1.274 \times 10^7 B_0 T_{acq'n}}{m/z} \qquad \textbf{Equation 1.2}$$

$$MMA_{(ppm)} = \frac{M_{exp} - M_{theo}}{M_{theo}} \times 10^6 \qquad \textbf{Equation 1.3}$$

FTICR-MS is a very powerful mass analyzer. Its ability to obtain high resolving power and low mass measurement accuracy (MMA) that make it an ideal tool for examining complex mixtures. Resolving power using *FTICR* results from the length of the transient as well as the applied magnetic field. The MMA is defined to be the difference between the experimental mass (M_{exp}) and theoretical mass (M_{theo}) divided by (M_{theo}). The routine achievable mass is in the low part per million range (<5 ppm) and resolving power is ~100,000 RP.

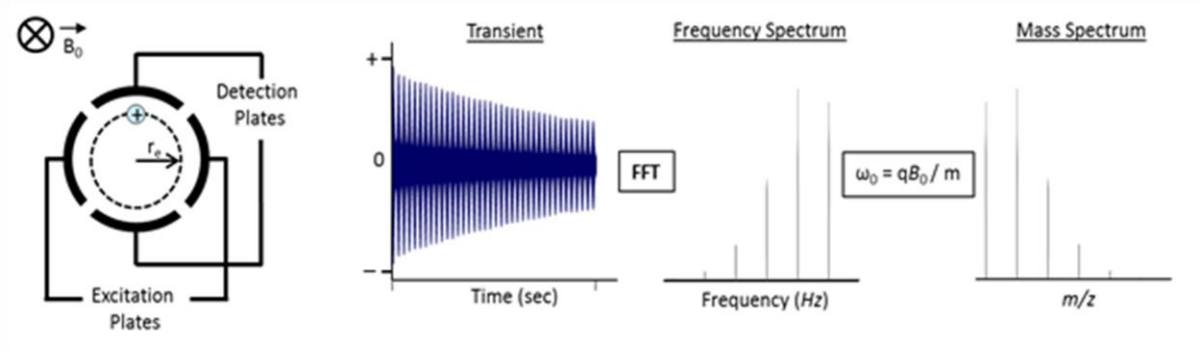


Figure 1.2 – Schematic of the ion detection in FTICR. The induced ion current on the detection plates produce an image currant (transient). The transient is the fourier transformed resulting in a frequency domain which is then converted to a *m/z* spectrum.

1.4.2 Linear Quadrupole Orbitrap Mass Analyzer

Mass spectrometry instrumentation has been continuously evolving. The latest of mass spectrometry instruments is the hybrid linear quadrupole obritrap MS system (Q-Exactive). Advantages of the new instrument range from improved ion transmission from the electrospray source to the vacuum through the use of an S-lens, higher mass accuracy of MS/MS spectra, and fast MS/MS fragmentation due to parallel filling and detection modes. It has also been stated that fragmentation is more efficient since MSMS occurs in the high energy collision dissociation (HCD) cell rather than in an Ion Trap. One disadvantage of using a Q-Exactive is the achievable resolving power when comparing the same length transient times. However, for bottom-up experiments the resolving power is more than efficient.

Figure 2 is a schematic of the linear quadrupole orbitrap MS. Unlike the hybrid mass analyzer LTQ-FTICR-MS, the orbitrap is the only mass analyzer.[40,41] And in place of an LTQ, there is a quadrupole that can be used as a mass filter. The ions are collected in the C-trap and then injected into the orbitrap. The injection of the ions provides an initial phase of ion oscillation around the inner spindle electrode. The ions orbit around the center electrode and simultaneously oscillate in the *z*-direction. It's this frequency of the ions in the *z*-direction that is used to determine the ion mass/charge ratio. The relationship is demonstrated in the **equation 4** below. Unlike FTICR the equation becomes independent of *r* and ϕ motion.[42]

$$\omega = \sqrt{\frac{m}{z} \times k} \qquad \textbf{Equation 1.4}$$

Where *k* is a defined constant based upon the field curvature. The ω is detected by an induced current on outer detector electrodes. The signal is recorded in the time-domain that is then processed using an 'enhanced Fourier transformation' (eFT) algorithm. The new feature of the eFT algorithm allows for doubling mass spectrometric resolution.

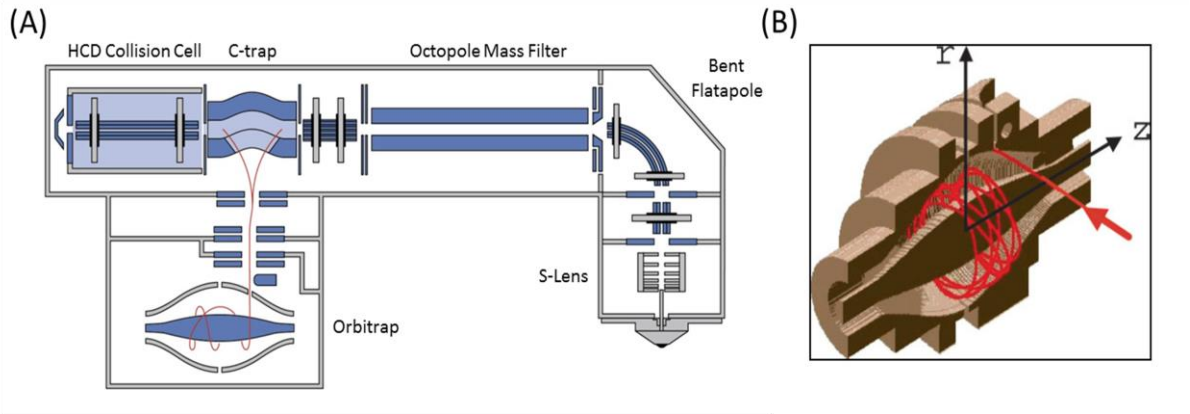


Figure 1.3 – (A) Schematic of the Linear Quadrupole Orbitrap Mass Spectrometer. (B) A schematic of the ion motion within an Orbitrap. (Picture taken from J. Mass Spectrom. 2005; 40: 430-443)

1.5 Bioinformatics

The 'bottom-up' approach for protein identification and quantification is used for all of the experiments. The proteins are purified from the samples and then enzymatically digested. The generated peptides are separated using liquid chromatography coupled with tandem mass spectrometry. Data processing can be handled by several different programs that assign peptide sequences from the acquired MS/MS data. SEQUEST[43] and MASCOT[44] are the most popular programs, in addition to others such as Andromeda[45], X!-tandem,[46] and OMSSA[47]. MASCOT is utilized to provide a theoretical in-silico digestion of each protein and a given database is used based on certain search criteria; digestive enzymes, number of

missed cleavages, and fixed and/or variable modifications. The data is processed based on the determination of possible peptide matches found in MS spectra within a defined mass tolerance. Then each complementing MS/MS spectra are evaluated to determine the peptide sequence. A subsequent score is given based a fitting of the theoretical spectrum of the candidate peptide matches, with the observed spectrum. The resulting peptides are grouped to propose a corresponding protein in the target and decoy database.

Protein grouping, statistical filtering and quantification of MASCOT DAT files were accomplished using ProteoIQ.[48-50] Identified proteins (protein groups) were considered valid using a protein false discovery rate (FDR) of 1%. A label free approach for relative quantification was accomplished by using the spectral counts obtained from ProteoIQ. The advantage of using label free approaches for relative quantification is the need for added tagging or labeling techniques is discarded. Studies have proven that relative abundances can be determined by using spectral counting and/or ion abundances. However, discrepancies in spectral counts can be seen between runs and between samples. In order to compensate for these differences, normalization of spectral counts is performed between runs and samples. To obtain the normalized spectral counts (NSpC) for a given protein, spectral counts are first normalized within replicates based upon the run with the highest total number of spectral counts (TSpC). It is then followed by normalizing the spectral counts between samples using the replicate with the largest TSpC.

Larger proteins will inherently have more spectral counts because of the number of peptides are generated is larger. Relative abundances can be calculated by obtained normalized spectral abundance factors (NSAF). This is calculated by the number of spectral counts for a given protein (SpC) divided by the number of amino acids in the protein (L) and then divided by the sum of SpC/L for all proteins in the dataset.

$$NSAF = \frac{\left({SpC}/{Length}\right)x}{\sum_{i=1}^{N}\left({SpC}/{Length}\right)x} \qquad \textbf{Equation 1.5}$$

1.6 Laser Microdissection

The analysis of heterogenous tissues can be an overwhelming task due to the difficulty to discern which cells contribute which cellular component to a given tissue lysate. However, by employing laser microdissection (LMD), users are able to attain subpopulations of tissues under direct microscopic visualization.[51,52] LMD technology utilizes an ultraviolet (UV) laser to isolate specific microscopic regions from a tissue sample without the contamination of surrounding cells. The initial applications for LMD were intended to isolate lowest amounts of cells needed for amplifying DNA or RNA. In more recent applications, researchers have isolated cells in hopes to define proteomic profiles of healthy and diseased cells.[53] Discrepancies in the number of cells needed for a global proteomic study using LMD based experiments has been much debated. Several authors have demonstrated the

identification of 905 proteins from 500 cells.[53] The amount of cells will vary due to the fact that cells can vary in size.

The analysis of biological and clinical samples is limited by their availability. Samples can either be preserved by fresh frozen or formalin-fixed and paraffin embedding (FFPE). FFPE samples are preferred method of preserving the tissues since they can be maintained at room temperature and for longer periods of time. However in the formalin fixation process, formalin adds methylene hydrate groups to the side chains of amino acids resulting in methylene bridge formation causing protein cross-linking. The result of these cross-links can lead to polymer formation which may lead to a misidentification of the native protein. Nevertheless, researchers have shown robust results in protein identification of *FFPE* tissues. Several groups have demonstrated in-depth proteomic analysis of FFPE tissues from lung adenocarcinoma and squamous cell carcinoma located in the head and neck.[54,55] Studies have also been conducted on FFPE compared to matched OCT frozen tissue samples, concluding that the FFPE did not have any bearing on the detection and quantification of targeted protein identifications.[56]

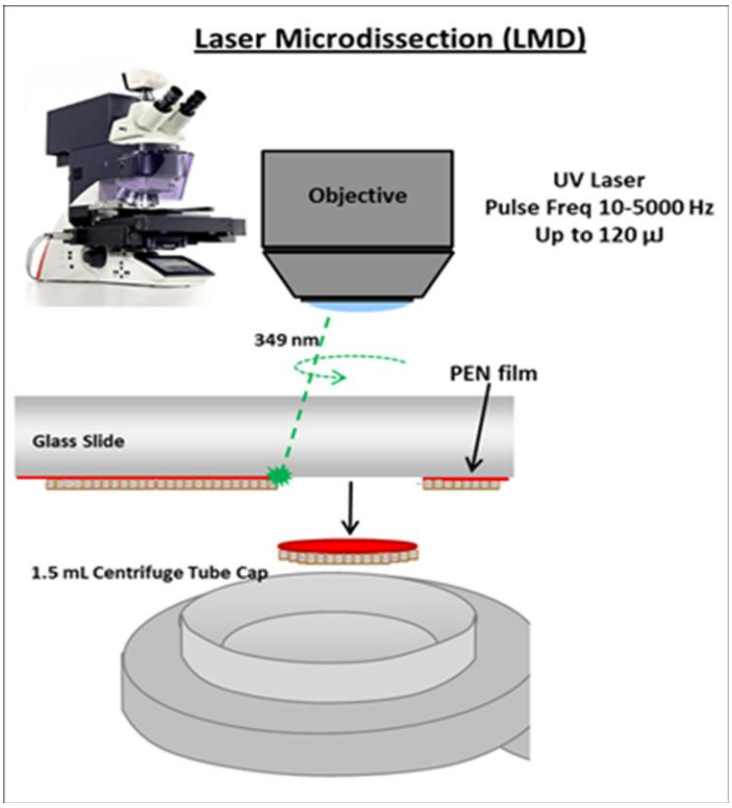


Figure 1.4 – Schematic of laser microdissection instrument.

1.7 Synopsis of Completed Research

The work presented within this dissertation encompasses a global proteomic approach to using tissues obtained from model biological system. Herein, several protein extraction methods have been employed prior to analysis using LC-MS/MS technologies. **Chapter 2** describes the role of bioinformatics in large scale global proteomic studies. Users must confidently identify proteins as developments in mass spectrometry allow for more accurate and sensitive instrumentation. This study confidently identified more than 4800 proteins from mouse brain tissues.

Chapter 3 utilizes the domestic chicken as a model for studying ovarian cancer. The initial goal of this study was to identify novel biomarkers from the chicken model using quantitative mass spectrometry-based proteomics. The global proteomes of plasma, ovary and oviduct tissues were compared in Heatlhy Egg Laying Hens, Early stage and Late Stage OVC Hens. This resulted in the identifying the OVOS2 protein as significantly up-regulated in late-stage ovarian cancer chickens. The identification of the OVOS2 as a candidate biomarker was further investigated in human ovarian cancer cell lines. Levels of OVOS2 were significantly up-regulated at the mRNA and protein levels providing evidence of OVOS2 as a conceivable ovarian cancer biomarker.

For the studies conducted in **Chapter 3**, the entire tissues were consumed. In order to use less of the tissue, Laser Microdissection (LMD) was employed. LMD was used in removing the yolk, follicular wall, and stromal cells of small white follicles embedded in the ovarian cortex. In **Chapter 4**, the proteomic profiles obtained from each segregated compartment were quantitatively compared. This technique provides an *in-situ* proteomics based approach in studying follicular development. This novel approach has proven complementary to previously used mRNA expression and western blots techniques in studying follicular development. The methodology allows for less of the tissue to be consumed, allowing the remaining tissue to be preserved and examined later for any future targeted experiments that may result from the global proteomics data.

1.8 References

(1) Williams, T. I.; Toups, K. L.; Saggese, D. A.; Kalli, K. R.; Cliby, W. A.; Muddiman, D. C. Epithelial ovarian cancer: disease etiology, treatment, detection, and investigational gene, metabolite, and protein biomarkers. *J Proteome Res* **2007**, *6*(8), 2936-62.

(2) Bast, R. C., Jr.; Hennessy, B.; Mills, G. B. The biology of ovarian cancer: new opportunities for translation. *Nat Rev Cancer* **2009**, *9*(6), 415-28.

(3) Cho, K. R.; Shih Ie, M. Ovarian cancer. *Annu Rev Pathol* **2009**, *4*287-313.

(4) Scully, R. E. Classification of human ovarian tumors. *Environ Health Perspect* **1987**, *73*15-25.

(5) Auersperg, N.; Wong, A. S.; Choi, K. C.; Kang, S. K.; Leung, P. C. Ovarian surface epithelium: biology, endocrinology, and pathology. *Endocrine Reviews* **2001**, *22*(2), 255-288.

(6) Wilson, J. E. Adeno-carcinomata in hens kept in a constant environment. *Poult Sci* **1958**, *37*1253.

(7) Fredrickson, T. N. Ovarian Tumors of the Hen. *Environmental Health Perspectives* **1987**, *73*35-51.

(8) Barua, A.; Bitterman, P.; Abramowicz, J. S.; Dirks, A. L.; Bahr, J. M.; Hales, D. B.; Bradaric, M. J.; Edassery, S. L.; Rotmensch, J.; Luborsky, J. L. Histopathology of ovarian tumors in laying hens: a preclinical model of human ovarian cancer. *Int J Gynecol Cancer* **2009**, *19*(4), 531-9.

(9) Hawkridge, A. M.; Wysocky, R. B.; Petitte, J. N.; Anderson, K. E.; Mozdziak, P. E.; Fletcher, O. J.; Horowitz, J. M.; Muddiman, D. C. Measuring the intra-individual variability of the plasma proteome in the chicken model of spontaneous ovarian adenocarcinoma. *Anal Bioanal Chem* **2010**, *398*(2), 737-749.

(10) Hakim, A. A.; Barry, C. P.; Barnes, H. J.; Anderson, K. E.; Petitte, J.; Whitaker, R.; Lancaster, J. M.; Wenham, R. M.; Carver, D. K.; Turbov, J.; Berchuck, A.; Kopelovich, L.; Rodriguez, G. C. Ovarian adenocarcinomas in the laying hen and women share similar alterations in p53, ras, and HER-2/neu. *Cancer Prev Res (Phila)* **2009**, *2*(2), 114-21.

(11) Stammer, K.; Edassery, S. L.; Barua, A.; Bitterman, P.; Bahr, J. M.; Hales, D. B.; Luborsky, J. L. Selenium-binding protein 1 expression in ovaries and ovarian tumors in the laying hen, a spontaneous model of human ovarian cancer. *Gynecological Oncology* **2008**, *109*(1), 115-121.

(12) Jackson, E.; Anderson, K.; Ashwell, C.; Petitte, J.; Mozdziak, P. E. CA125 expression in spontaneous ovarian adenocarcinomas from laying hens. *Gynecological Oncology* **2007**, *104*(1), 192-198.

(13) Urick, M. E.; Johnson, P. A. Cyclooxygenase 1 and 2 mRNA and protein expression in the Gallus domesticus model of ovarian cancer. *Gynecological Oncology* **2006**, *103*(2), 673-678.

(14) Johnson, P. A.; Giles, J. R. Use of genetic strains of chickens in studies of ovarian cancer. *Poult Sci* **2006**, *85*(2), 246-50.

(15) Giles, J. R.; Olson, L. M.; Johnson, P. A. Characterization of ovarian surface epithelial cells from the hen: a unique model for ovarian cancer. *Exp Biol Med* **2006**, *231*(11), 1718-1725.

(16) Giles, J. R.; Shivaprasad, H. L.; Johnson, P. A. Ovarian tumor expression of an oviductal protein in the hen: a model for human serous ovarian adenocarcinoma. *Gynecological Oncology* **2004**, *95*(3), 530-533.

(17) Rodriguez-Burford, C.; Barnes, M. N.; Berry, W.; Partridge, E. E.; Grizzle, W. E. Immunohistochemical expression of molecular markers in an avian model: a potential model for preclinical evaluation of agents for ovarian cancer chemoprevention. *Gynecological Oncology* **2001**, *81*(3), 373-379.

(18) Dixon, R. B.; Bereman, M. S.; Petitte, J. N.; Hawkridge, A. M.; Muddiman, D. C. One-Year Plasma N-linked Glycome Intra-individual and Inter-individual Variability in the Chicken Model of Spontaneous Ovarian Adenocarcinoma. *Int. J. Mass Spectrom. Ion Proc.* **2011**, *305*(2-3), 79-86.

(19) Ansenberger, K.; Zhuge, Y.; Lagman, J. A.; Richards, C.; Barua, A.; Bahr, J. M.; Hales, D. B. E-cadherin expression in ovarian cancer in the laying hen, Gallus domesticus, compared to human ovarian cancer. *Gynecologic oncology* **2009**, *113*(3), 362-9.

(20) Skates, S.; Troiano, R.; Knapp, R. C. Longitudinal CA125 detection of sporadic papillary serous carcinoma of the peritoneum. *Int J Gynecol Cancer* **2003**, *13*(5), 693-6.

(21) Skates, S. J.; Menon, U.; MacDonald, N.; Rosenthal, A. N.; Oram, D. H.; Knapp, R. C.; Jacobs, I. J. Calculation of the risk of ovarian cancer from serial CA-125 values for preclinical detection in postmenopausal women. *J Clin Oncol* **2003**, *21*(10 Suppl), 206s-210s.

(22) Tuxen, M. K.; Soletormos, G.; Petersen, P. H.; Schioler, V.; Dombernowsky, P. Assessment of biological variation and analytical imprecision of CA 125, CEA, and TPA in relation to monitoring of ovarian cancer. *Gynecologic oncology* **1999**, *74*(1), 12-22.

(23) Tuxen, M. K.; Soletormos, G.; Rustin, G. J.; Nelstrop, A. E.; Dombernowsky, P. Biological variation and analytical imprecision of CA 125 in patients with ovarian cancer. *Scand J Clin Lab Invest* **2000**, *60*(8), 713-21.

(24) Skates, S. J.; Xu, F. J.; Yu, Y. H.; Sjovall, K.; Einhorn, N.; Chang, Y.; Bast, R. C., Jr.; Knapp, R. C. Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers. *Cancer* **1995**, *76*(10), 2004-2010.

(25) Lim, W.; Jeong, W.; Kim, J. H.; Lee, J. Y.; Kim, J.; Bazer, F. W.; Han, J. Y.; Song, G. Differential expression of alpha 2 macroglobulin in response to dietylstilbestrol and in ovarian carcinomas in chickens. *Reprod Biol Endocrin* **2011**, *9*.

(26) Ocon-Grove, O. M.; Poole, D. H.; Johnson, A. L. Bone morphogenetic protein 6 promotes FSH receptor and anti-Mullerian hormone mRNA expression in granulosa cells from hen prehierarchal follicles. *Reproduction* **2012**, *143*(6), 825-33.

(27) Robinson, F. E.; Etches, R. J. Ovarian steroidogenesis during follicular maturation in the domestic fowl (Gallus domesticus). *Biol Reprod* **1986**, *35*(5), 1096-105.

(28) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal Chem* **1996**, *68*(5), 850-858.

(29) *Saccharomyces* Genome Database. http://www.yeastgenome.org/.

(30) Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2007**, *2*(8), 1896-1906.

(31) Taylor, G. *Proc. Royal Soc. London. Series a. Math and Phys. Sci.* **1964**, *280*(138).

(32) Rayleigh, L.; F.R.S. Philosophical Magazine Series 5. **1882**, *14*.

(33) Dole, M.; Mack, L. L.; Hines, R. L. Molecular Beams of Macroions. *J Chem Phys* **1968**, *49*(5), 2240-&.

(34) Mack, L. L.; Kralik, P.; Rheude, A.; Dole, M. Molecular Beams of Macroions .2. *J Chem Phys* **1970**, *52*(10), 4977-&.

(35) Iribarne, J. V.; Thomson, B. A. Evaporation of Small Ions from Charged Droplets. *J Chem Phys* **1976**, *64*(6), 2287-2294.

(36) Thomson, B. A.; Iribarne, J. V. Field-Induced Ion Evaporation from Liquid Surfaces at Atmospheric-Pressure. *J Chem Phys* **1979**, *71*(11), 4451-4463.

(37) Iribarne, J. V.; Dziedzic, P. J.; Thomson, B. A. Atmospheric-Pressure Ion Evaporation Mass-Spectrometry. *International Journal of Mass Spectrometry and Ion Processes* **1983**, *50*(3), 331-347.

(38) Fenn, J. B. *J. Am. Soc. Mass Spectrom* **1993**, *4*(7), 524-535.

(39) Comisarow, M. B. Theory of Fourier-Transform Ion-Cyclotron Resonance Mass-Spectroscopy .2. Signal Modeling for Ion-Cyclotron Resonance. *J Chem Phys* **1978**, *69*(9), 4097-4104.

(40) Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal Chem* **2000**, *72*(6), 1156-1162.

(41) Hu, Q. Z.; Noll, R. J.; Li, H. Y.; Makarov, A.; Hardman, M.; Cooks, R. G. The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **2005**, *40*(4), 430-443.

(42) Perry, R. H.; Cooks, R. G.; Noll, R. J. Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications. *Mass Spectrom Rev* **2008**, *27*(6), 661-699.

(43) Eng, J. K.; Mccormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J Am Soc Mass Spectrom* **1994**, *5*(11), 976-989.

(44) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*(18), 3551-3567.

(45) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J Proteome Res* **2011**, *10*(4), 1794-1805.

(46) Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*(9), 1466-1467.

(47) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X. Y.; Shi, W. Y.; Bryant, S. H. Open mass spectrometry search algorithm. *J Proteome Res* **2004**, *3*(5), 958-964.

(48) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **2002**, *74*(20), 5383-5392.

(49) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **2003**, *75*(17), 4646-4658.

(50) Weatherly, D. B.; Atwood, J. A.; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R. A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol Cell Proteomics* **2005**, *4*(6), 762-772.

(51) EmmertBuck, M. R.; Bonner, R. F.; Smith, P. D.; Chuaqui, R. F.; Zhuang, Z. P.; Goldstein, S. R.; Weiss, R. A.; Liotta, L. A. Laser capture microdissection. *Science* **1996**, *274*(5289), 998-1001.

(52) Espina, V.; Heiby, M.; Pierobon, M.; Liotta, L. A. Laser capture microdissection technology. *Expert Rev Mol Diagn* **2007**, *7*(5), 647-657.

(53) Wisniewski, J. R.; Ostasiewicz, P.; Mann, M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J Proteome Res* **2011**, *10*(7), 3040-9.

(54) Kawamura, T.; Nomura, M.; Tojo, H.; Fujii, K.; Hamasaki, H.; Mikami, S.; Bando, Y.; Kato, H.; Nishimura, T. Proteomic analysis of laser-microdissected

paraffin-embedded tissues: (1) Stage-related protein candidates upon non-metastatic lung adenocarcinoma. *J Proteomics* **2010**, *73*(6), 1089-1099.

(55) Reimel, B. A.; Pan, S.; May, D. H.; Shaffer, S. A.; Goodlett, D. R.; McIntosh, M. W.; Yerian, L. M.; Bronner, M. P.; Chen, R.; Brentnall, T. A. Proteomics on Fixed Tissue Specimens - A Review. *Curr Proteomics* **2009**, *6*(1), 63-69.

(56) Scicchitano, M. S.; Dalmas, D. A.; Boyce, R. W.; Thomas, H. C.; Frazier, K. S. Protein Extraction of Formalin-fixed, Paraffin-embedded Tissue Enables Robust Proteomic Profiles by Mass Spectrometry. *J Histochem Cytochem* **2009**, *57*(9), 849-860.

CHAPTER 2

Accurate Identification of Deamidated Peptides in Global Proteomics using a Quadrupole Orbitrap Mass Spectrometer

2.1 Introduction

Mass spectrometry continues to be a leading technology used in the field of proteomics.[1] The widely used shotgun "bottom-up proteomics" approach allows for the identification of proteins from complex mixtures. This method relies on the enzymatic digestion of proteins into peptides, followed by separation using nanoscale liquid chromatography (LC) coupled to a mass spectrometer. Data is often collected in a data dependent acquisition (DDA) mode, in which a full precursor spectrum is acquired and then followed by a series of MS/MS product ion spectra. The ions chosen for MS/MS are based on the ion abundance with a defined threshold, and the number of consecutive product ion spectra is defined by the user. The series of precursor ions and resulting MS/MS ion peaks can be searched using a variety of search engines. Several programs such as SEQUEST[2,3], Mascot[4] and X!Tandem[5] are often utilized.

Database searching requires a user-defined protein database obtained for a specified organism. Each program performs an *in-silico* digestion of the protein database to produce a peptide list based on specified parameters such as digestive

enzymes, number of missed cleavages, and fixed/variable modifications. A precursor *m/z* is first matched to possible candidate peptides within a defined mass tolerance. MS/MS spectra are then matched to possible theoretical sequence information from the candidate peptides and then given a confidence score. These identified peptides are then grouped into their corresponding protein(s).

Challenges in global proteomics arise with accurate identification of low-abundant peptides with insufficient precursor ion abundance. Advancements in instrument acquisition speeds, resolving power, and mass measurement accuracies have greatly improved the number of proteins identified. To achieve the maximum number of proteins identified, a quadrupole orbitrap mass spectrometer (Q-Exactive) was utilized in these experiments. The Q-Exactive offers many attributes that allow for a comprehensive analysis of proteins extracted from a complex biological matrix. Unlike hybrid instruments with dual mass analyzers, all spectra are obtained using an orbitrap mass analyzer with acquisition rates as fast as >12 Hz at 17.5 k_{FWHM} resolving power.[6,7] High resolving power spectra and high mass measurement accuracy can thus be obtained with the use of the orbitrap mass analyzer. Many researchers have developed strategies that account for high resolution tandem mass spectra[8] and high precursor mass accuracy[9].

Researchers commonly process their protein samples first by using reducing agents such as dithiothreitol to break any intramolecular and intermolecular disulfide bridges. The addition of iodoacetamide (IAM) is often followed to covalently bond with the free thiol groups to prevent any cysteine disulfide formations from

reoccurring. Typical trypsin digestion protocols use buffers with a pH value between 7-8 for 1-24 hour duration at 37 ^{0}C. During the peptide process, peptides are frequently searched with post-translation modifications. Peptides are searched with a fixed carbamidomethyl modification on cysteine residues that accounts for the iodoacetamide (+ 57.02 Da). However, data is often searched with variable modification such as oxidation of methionine (+ 15.99 Da) and deamidation of asparagines (N) and glutamines (Q) (+ 0.98 Da). Researchers have optimized trypsin proteolysis in order to maximize detection of targeted peptides.[10-12]

Oxidation of methionine residues is an important post-translational modification; it has been related to oxidative stress and aging.[13] However, oxidation can be induced during the electrospray process.[14] This is the reason for searching oxidation on methionine residues as a variable modification. Differentiating between the oxidized and unmodified peptides can easily be distinguished due to the + 15.99 Da mass shift. Deamidation of asparagine and glutamine on the other hand results in only a + 0.98 Da mass shift. Deamidation has been related to protein degradation and is thought to play an important part of aging studies.[15,16] Deamidation has also resulted in loss of protein structure, solubility and function.[17] Further, accurate identification of deamidated peptides impacts the assignment of *N*-glycosylated sites.

The deamidation of asparagine to aspartic acid and glutamine to glutamic acid is a non-enzymatic process, the half-life can be between 1-500 days.[18] Current protocols for tryptic digestion are favorable for deamidation due to the basic pH

condition and can cause the misrepresentation of deamidation of native proteins. To circumvent this challenge, Li *et al.* has implemented an O^{18} labeling strategy to monitor deamidation during sample preparation.[19] MALDI and HPLC-MALDI studies resulted in approximately 70-80% of –Asn-Gly- sites as being deamidated after a standard 12 hr tryptic digestion at 37 ^{0}C.[20] Hao and coworkers have shown that the ^{13}C peaks of amidated peptides can be misassigned as monoisotopic peaks of the corresponding deamidated peptides using a linear quadrupole ion trap fourier transform mass spectrometer, and they discuss an arbitrary method for the determination of deamidated or amidated peptides..[10]

Search parameters were optimized in order to increase the number of identified proteins for a global proteome study on mouse brain tissue lysate. Increasing the value on the precursor mass tolerance beyond 5 ppm resulted in an increase in proteins identified. It was also determined that a MS/MS tolerance of 0.02 Da was optimal for the number of proteins identified. Further investigation of the increase in number of proteins identified revealed a large contingency of peptides that were inaccurately assigned as deamidated. These misassigned deamidated peptides increased the number of proteins identified, resulting in falsely identified proteins. It is inherently important not only to use a mass spectrometer with high mass measurement accuracy and high resolving power but to also carefully consider search parameters in the bioinformatics software in order to avoid false positive protein identifications.

2.2 Experimental

2.2.1 Animals

Animal testing was performed under Institutional Care and Use Committee regulations and approval at North Carolina State University. Mice were housed at Laboratory Animal Research facilities at the College of Veterinary Medicine. Mice were sacrificed at 5 days of age, and the brains were removed. Following the removal of the brain, the brains were washed with 0.1 M phosphate buffered saline (PBS) and were snap frozen in liquid nitrogen, weighed and then stored at -80°C.

2.2.2 Sample Preparation and Digestion

Lysis buffer at a ratio of 5 µL:1 mg of tissue was added to extract the proteins from the brain tissue. All reagents were purchased from Sigma Aldrich (St. Louis, MO, USA). The lysis buffer consisted of 50 mM Tris, 8 M Urea, 2 M Thiourea, 10 mM EDTA, 10 mM DTT, 0.001% sodium azide and was adjusted to a pH of 7.8 using hydrochloric acid. An OMNI TIP Homogenizing Kit (Omni International, Kennesaw, Georgia) was employed to break apart frozen tissue that was suspended in the lysis buffer. The homogenizer is a mechanical blade that was used for no longer than 1 minute to prevent the tissue from heating up.This was followed by adding sodium dodecyl sulfate (SDS) to make up a final concentration of 2% (w/v) and then placed onto the Genie Disrupter (Scientific Industries) for 1 minute followed by 5 minutes of incubation period on ice. This step was then repeated twice more. The sample was then centrifuged for 30 minutes at 14,000 $\times$ *g*. The soluble fraction

was removed from the centrifuge tube without disrupting the cellular debris pellet or the top lipid layer.

The samples were digested following the filter-aided sample preparation (FASP) method published by Wisniewski and co-workers with minor changes.[21] 21 µL of ~7 mM dithiothreitol (DTT) was added to 9 µL of brain tissue lysate for a final concentration of 5 mM DTT. The sample was incubated for 30 min at 56 ^{o}C to reduce the protein disulfide bonds. The sample was then mixed with 200 µL of 8 M urea in 0.1 M Tris/HCl (pH 8.5) and transferred onto a Vivacon 500 30 kDa MW cutoff filter (Vivaconproducts, Littleton, MA) and centrifuged at a constant 14,000g and 21 ^{o}C to prevent carbamylation using a refrigerated benchtop centrifuge (Eppendorf, Hauppauge, NY) for 15 min. This step was repeated once more and then the flow-through solvent was discarded. Alkylation step was performed when 100 µL of 0.05 M iodoacetamide was added and incubated for 20 min in the dark at room temperature. The filter was centrifuged for 10 min. Wash steps were performed with 100 µL of 8 M urea (3×) and 100 µL of 0.05 M ammonium bicarbonate in H_2O (3×), each step the filter was centrifuged for 10 min. The flow-through container vials were switched out with clean ones. 40 uL of ammonium bicarbonate buffer and 2.5 µL of 0.5 µg/µL trypsin solution were added onto the filter. Trypsin digestion was performed for 16 hours at 37 ^{o}C. Following digestion, 40 µL of 0.1% formic acid was added to the filter, and the peptides were eluded by centrifugation for 10 min at 14,000 × *g*. Protein concentration was determined using a NanoDrop 2000c (Thermo Scientific, Wilmington, DE) reading at 280 nm.

2.2.3 StageTip Anion Fractionantion

A detailed stage tip procedure was used and is described by Gokce and co-workers.[22] Briefly, a 200 µL pipette tip (Eppendorf, Hauppauge, NY) was packed with six anion disks from an Empore C18 extraction disk (3M, St. Paul, MN). The peptides that were obtained from the FASP procedure were diluted to ~200 µL by adding 100 µL of Britton Robinson buffer pH 11 and were loaded on top of the anion stage tip. Peptides were eluted into a micro centrifuge tubes by centrifugation. Additional peptides were eluted from the anion stage tip using Britton Robinson buffers pH 8, 6, 5, 4, and 3 separately. Fractions were dried to completion and stored at -80 ^{o}C until analysis.

2.2.4 Nano-Flow Reversed Phase Chromatography

A Thermo Scientific EASY nLC II (Thermo Scientific, San Jose, CA) was coupled to a quadrupole Orbitrap benchtop mass spectrometer (Q-Exactive, San Jose, CA) using a vented column configuration.[23] The vented column consisted of a 75 µm x 5 cm trap (IntegraFrit, New Objectives, Woburn, MA) coupled to a 75 µm x 15 cm column (PicoFrit, New Objectives, Woburn, MA). Both the trap and column were packed in-house with Magic C18AQ stationary phase (5 µm particle, 200 Å pore, Auburn, CA). Mobile phase A and B were composed of water/acetonitrile/formic acid (98/2/0.2% and 2/98/0.2% respectively). The dried samples were reconstituted to 0.08 µg/µL concentration. 15 µL of the sample was loaded onto the trap at 4.5 µL/min with 100% mobile phase A. Flow was then

diverted onto the column at the start of the gradient: 2% B (0-5 min), 2%-5% B (5-7 min), 5%-40% B (7-208 min), 40%-95% (208-218 min), 95% (218-228 min), 95%-2% (228-230 min), 2% (230-240 min).

Mass spectrometric analysis was performed on a Q-Exactive with optimized global proteomics parameters.[24] To summarize, MS transients in the orbitrap were acquired with 70 k_{FWHM} resolving power at *m/z* = 200. The automatic gain control (AGC) target for MS acquisitions was set to 1E6 with a maximum ion injection time of 30 ms. The scan range was set from 400 to 1600 *m/z*. Microscans were set to 1 for both the MS and MS/MS. Data dependent acquisition was set for 12 MS/MS spectra and the dynamic exclusion was set to 30 sec. The MS/MS resolving power was set to 17.5 k_{FWHM} at *m/z* = 200. MS/MS AGC target was set to 2E5 with a maximum ion injection time of 250 ms. All MS and MS/MS spectra were obtained in profile mode.

2.2.5 Data Analysis

Raw LC-MS/MS data files were processed into peak lists in a .MGF format using Proteome Discoverer. The resulting .MGF files were searched using MASCOT[4] (Matrix Science, Boston, MA) against a concatenated target-reverse mus musculus database (SwissProt, Feb 2013). Carbamidomethyl (C) was set as a fixed modification, and oxidation (M) were set as variable modifications and included a maximum of 2 missed cleavages. Data were also searched with and without deamidation (N and Q) as a variable modification. The precursor ion search

tolerance were searched from 1 to 100 ppm, and the fragment ion tolerance was set to ± 0.02 Da, ± 0.6 Da, ± 0.005 Da. Statistical filtering using a 1% false discovery rate of the identified proteins were accomplished using ProteoIQ.[25-27] Proteins that were identified with only 1 peptide were filtered from the total protein list.

2.3 Results and Discussion

To maximize the number of proteins identified, the mass tolerance for the precursor mass varied from 1 to 100 ppm, while MS/MS data was searched at ± 0.02 Da, 0.6 Da and .005 Da. **Figure 2.1a** depicts the overall trend of increasing proteins identified as the precursor ion mass tolerance reached 20 ppm. Subsequently, the data indicate that searching with ± 0.02 Da resulted in more identified proteins than 0.6 and 0.005 Da. All global proteomics experiments prior to this study were performed on an LTQ-Orbitrap or LTQ-FT system. Typical search criteria in our lab consisted of ± 5 ppm for precursor mass tolerance and ± 0.6 Da MS/MS tolerance when using LTQ detection for MS/MS sequencing. Since all MS and MS/MS spectra are obtained using the orbitrap on the Q Exactive, MS/MS data were searched with a more stringent tolerance of 0.02 Da (20 ppm at 1000 Da).

The observations from **Figure 2.1a** indicate that higher MS precursor mass tolerances results in more identified proteins. Data were further investigated by examining the identified peptides to determine the cause of the increase in identified proteins. **Figure 2.1b** illustrates the frequency of the mass accuracies of all identified

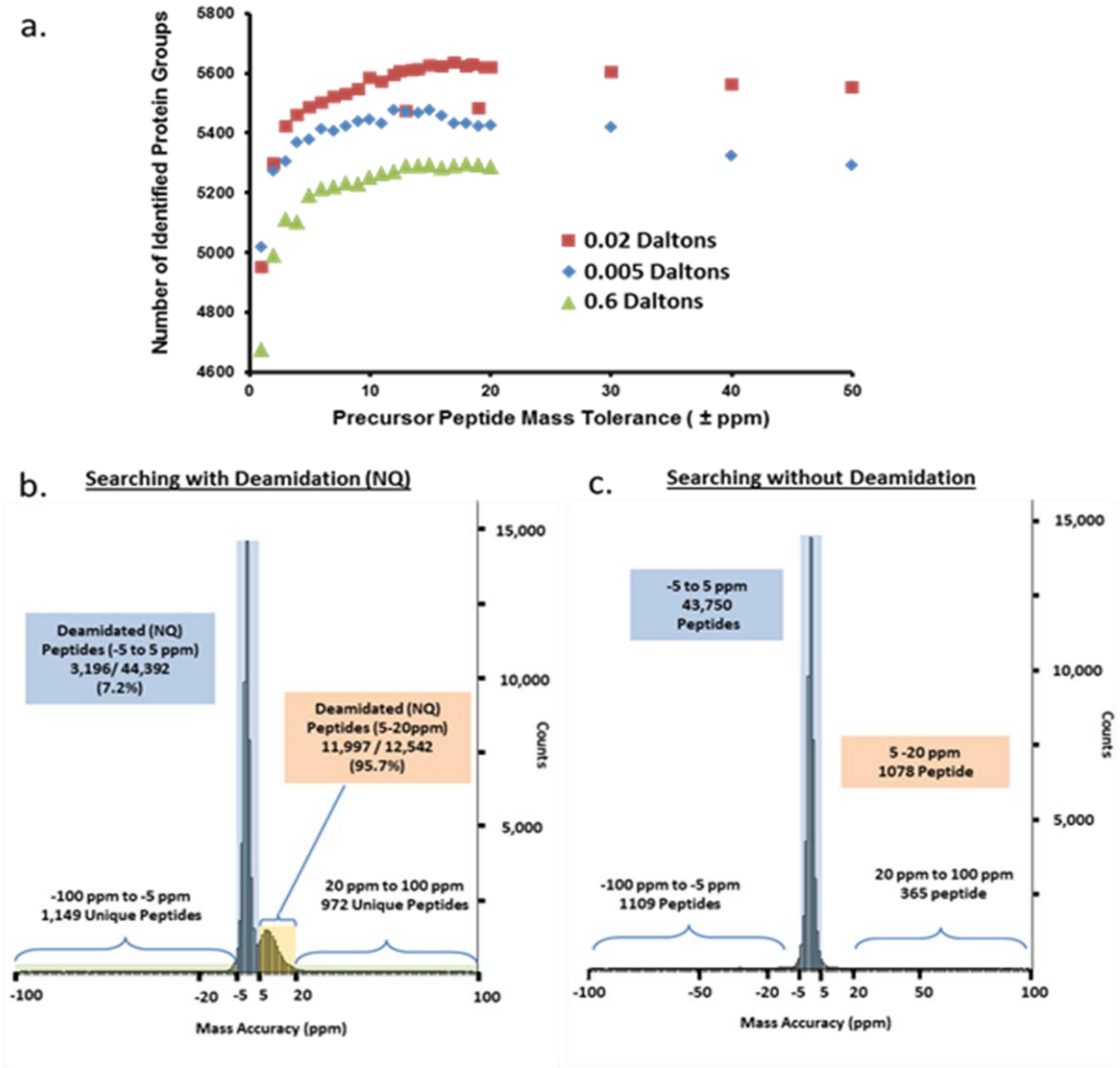


Figure 2.1 (a) A plot of the number of identified proteins with increasing precursor peptide mass tolerance. Data was searched at three separate values for the MS/MS mass tolerances, 0.06 Da, 0.02 Da, and 0.005 Da. 0.02 Da provided the greatest number of proteins identified at all precursor peptide mass tolerances. **(b)** Histogram of the mass accuracy of all peptides that were identified when searching data with Deamidation as a variable modification, ± 100 ppm precursor mass tolerance, and ± 0.02 Da MS/MS mass tolerance. There are two observable distributions, the second distribution which ranges from 5-20 ppm illustrates a large number of deamidated peptides. **(c)** Deamidation was not included in the search parameters, clearly proving that the second distribution is a result from using deamidation as a variable modification.

peptides using ± 100 ppm precursor mass tolerance and 0.02 Da as the MS/MS tolerance. **Figure 2.1b** distinctly reveals there are two separate distributions from the 59,055 total peptides identified. One distribution of peptides has a precursor mass measurement accuracy (MMA) that falls between -5 and 5 ppm, and another between 5 to 20 ppm. The latter distribution had experimental measured masses that were systematically above the theoretical mass resulting in a higher MMA according to Equation 1.

$$MMA = \frac{Experimental - Theoretical}{Theoretical} \times 10^6 \qquad \textbf{Equation 2.1}$$

It was initially thought that space-charge effects within the orbitrap may have been the cause for the overwhelming amount of peptides with a MMA greater than 5 ppm. Space-charge effects result from measuring too many ions in the orbitrap causing a systematic bias in mass measurments. The consequence of having too many ions in the orbitrap is ion-ion interactions, producing a shift in lower observed axial frequency (ω). The decrease in the observed axial frequency will result in an increase in observed *m/z* as shown in **Equation 2.2**. The increase in *m/z* will consequently result in the positive increase in mass error, similarly seen in **Figure 2.1b.** Space-charge effects are often compensated by using an AGC that limits the number of ions entering the orbtrap. To determine if the AGC was operating correctly, a set of deamidated from the -5 to 5 ppm and 5 to 20 ppm range were chosen for further investigation to determine if the AGC was overpopulating the orbitrap. In several cases, the ionization time (IT) was maximized suggesting that the

AGC limit was not reached. This suggests that the AGC was working properly by not allowing too many ions in the orbitrap thus preventing space-charge effects.

$$\omega = \sqrt{\frac{q}{m}k} \qquad \textbf{Equation 2.2}$$

Upon further investigation of the 44,392 peptides between -5 and 5 ppm, 7.2% of them were deamidated (N,Q). The percent of deamidated peptides between 5 to 20 ppm was calculated to be much higher at 95.7% deamidated (N,Q). **Figure 2.1c** is the same data but searched without deamidation (N,Q) as a variable modification. Clearly the second distribution in the 5 – 20 ppm range has noticeably disappeared. Similar distributions were identified in whole cell yeast lysates by Venable and coworkers.[28] They attribute a second distribution to deamidation of asparagine and glutamine residues which are identified by the + 0.98 Da mass shift. Analyses of the yeast lysate were performed on a LTQ mass spectrometer measuring their mass error in Daltons. Distributions similar to this were also recognized in other datasets obtained by our laboratory (data not shown).

To further investigate the reason why there was a larger percentage of deamidated peptides found in the 5 -20 ppm range, several peptides with mass measurement accuracy (MMA) between 5 – 20 ppm were examined closely. **Figure 2.2a.** depicts the identification of a deamidated (left) and amidated (right) form of the same peptide within the same chromatographic run. This peptide was selected because it was a deamidated peptide that has a mass accuracy of greater than 5 ppm and the corresponding amidated form was also identified in the same

chromatogram. A precursor spectrum is shown on the left hand side in which the

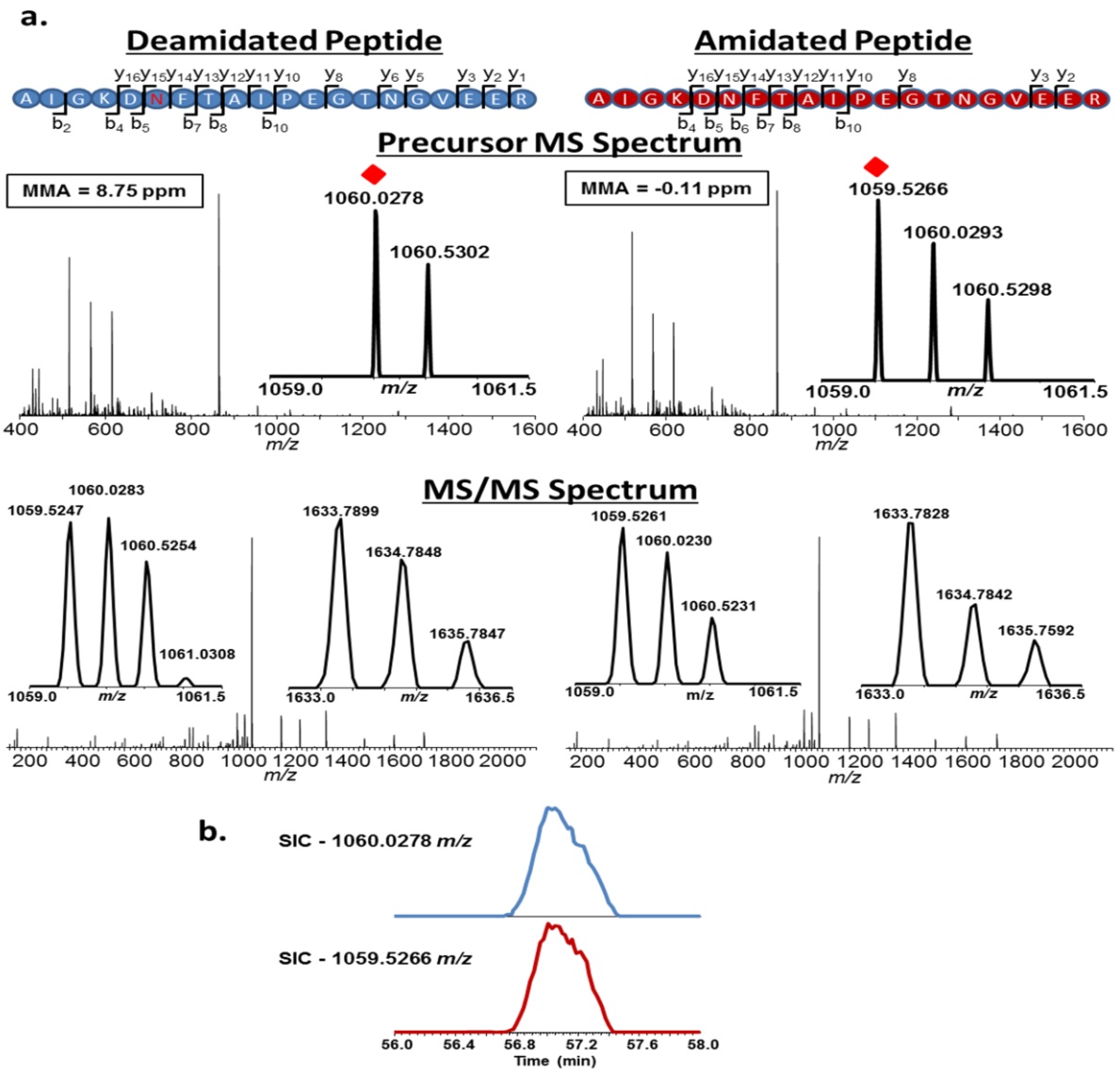


Figure 2.2 **(a)** A deamidated and amidated form of the AIGKDNFTAPEGTNGVEER peptide were identified from the MASCOT results. On the left side, the monoisotopic peak of 1060.02783 was established to be the deamidated form due to the mass accuracy of 8.75 ppm. From the MS/MS spectrum below, it is alluded that the deamidated peptide is in fact misassigned. The right illustrates the identification of the amidated peptide. And below is the MS/MS spectra that is comparable to that of the deamidated peptide. **(b)** Are the selected ion chromatograms for each peptide.

1060.0278 *m/z* was identified and selected for MS/MS. The (♦) above the peak is the *m/z* that was chosen for MS/MS. Below is the corresponding MS/MS spectrum for the 1060.0278 *m/z*. MASCOT identified the 1060.0278 *m/z* as the deamidated form for the peptide sequence shown in the **Figure 2.2a**. Investigation of the MS/MS spectra (below) indicates that there are residual precursor ions detected. An isotopic cluster with a monoisotopic peak at 1059.52466 *m/z* is clearly observed. This monoisotopic peak would correspond to the amidated form of the peptide. Also enhanced in the MS/MS spectrum is that of the y_{15} fragment (1633.790 *m/z*) of the peptide, which also illustrates that the fragment would be generated from the amidated form of the peptide. The overwhelming evidence from the MS/MS spectrum indicates that the peptide should be the amidated form. The mass accuracy of the deamidated precursor ion was calculated by MASCOT to be 8.75 ppm. The inability to detect the monoisotopic peak can be attributed to the fact the precursor spectrum obtained was at the beginning of the elution profile where the signal intensity is low. Based on the residual precursor ion mass from the MS spectrum, the peptide should be the deamidated form. However, the product ions observed in the MS/MS spectrum suggests the amidated form is present. Further, the mass accuracy of the monoisotopic peak for an amidated peptide was -1.95 ppm which falls within the expected mass measurement accuracy range for orbitrap analyzers.

The right column in **Figure 2.2a** confirms the amidated peptide from both the MS and MS/MS spectra used for the identification, the precursor mass accuracy was

calculated to be -0.11 ppm. In this example, the monoisotopic peak that was detected and selected for MS/MS (♦) corresponds to that of the unmodified form. The intact precursor ion agrees with the unmodified form. The mass accuracy of the precursor obtained from the MS/MS spectra is calculated to be -0.57 ppm. It can be concluded that the deamidated peptide was inaccurately assigned, and the 1060.0278 peak is truly the M+1 peak of the amidated form. **Figure 2.2b** are the selected ion chromatograms obtained for the monoisotopic peak of the amidated peptide (1059.5266) and the monoisotopic peak (1060.0278) of the "deamidated" peptide which show a lack of any retention time shift. Both chromatograms are similar; yet researchers have shown that deamidated and amidated peptides have different elution profiles.[10,17,29] This suggests that the "deamidated" and amidated peptide is actually the same peptide.

Figure 2.3 is another example of the same peptide identified as both deamidated and amidated. **Figure 2.3a** corresponds to the identification of the deamidated form of the peptide and the precursor mass spectrum. Enhanced is the isotopic cluster in which the 1082.03308 (♦) peak was chosen as the monoisotopic peak and selected for MS/MS. Since the 1082.03308 peak was chosen as the "monoisotopic mass" during data acquisition, it was then associated as the deamidated peptide resulting in a 10.9 ppm mass error because this was within the specified mass error tolerance of 100 ppm when searching the data against the protein database. Based on the isotopic distribution, the 1081.53149 should have been assigned as the monoisotopic peak. Similar to the previous example, the

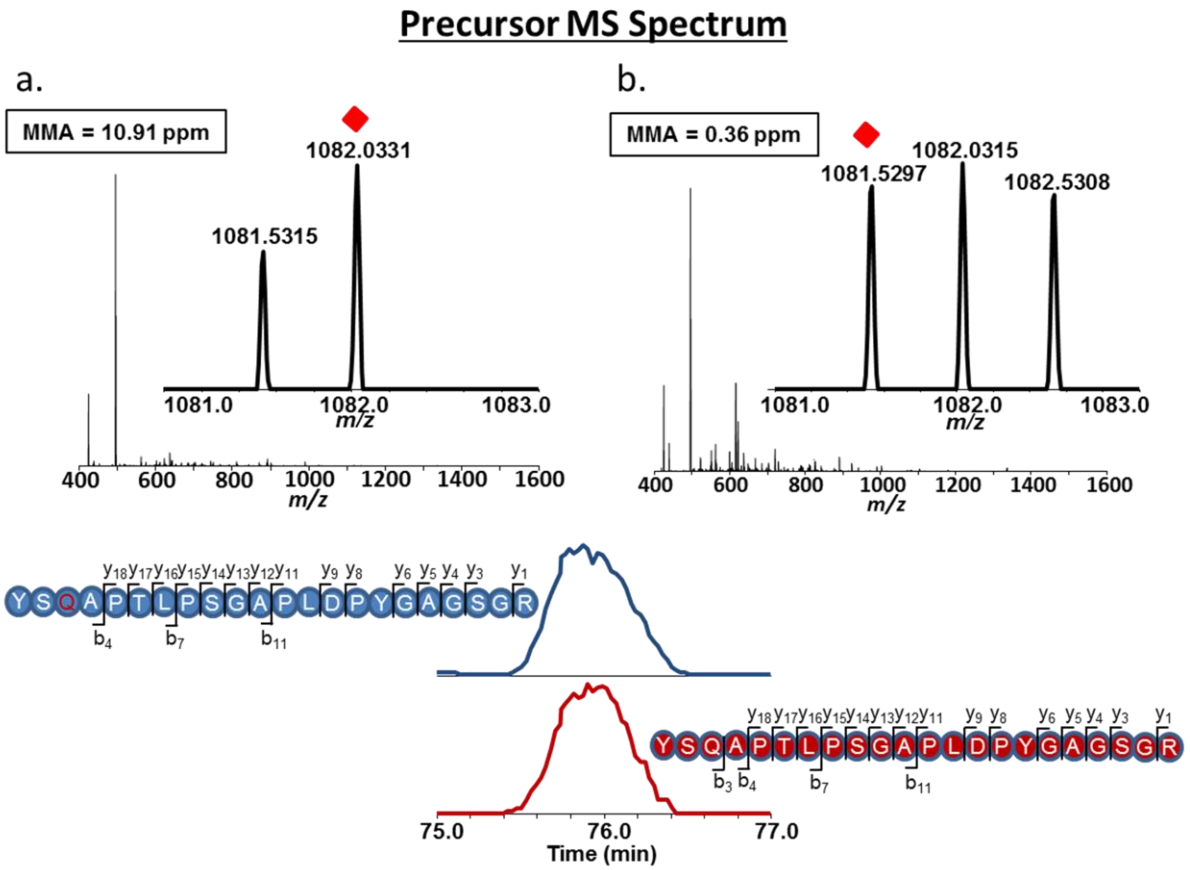


Figure 2.3 (a) Another example of a deamidated peptide that has a mass measurement accuracy of 10.91 ppm which results in a misidentification. The software chose the wrong peak (♦), in doing so it results in the identification of a deamidated peptide. **(b)** The correct amidated peptide was identified since the correct monoisotopic peak was used. (Bottom) Shows the selected ion chromatogram for both the deamidated and amidated peptide.

MS/MS spectrum (not shown) had fragment ions associated with the amidated version. Additionally, there are no fragments ions that would suggest that the deamidated form is present. **Figure 2.3b** is of the precursor that was used to identify the amidated form of the peptide. The mass accuracy calculated for the monoisotopic peak of the cluster above is 0.36 ppm. In the enhanced view of the

isotopic cluster, the monoisotopic peak 1081.52966 was the correct peak to be selected for MS/MS. Fragmentation spectrum indicates fragments corresponding to the amidated form. And similar to the previous example, the elution profiles (bottom) are overlapping. Since the wrong precursor peak was chosen as the monoisotopic peak and a wide mass accuracy tolerance was used, MASCOT was able to assign it as the deamidated peptide despite the fact that the fragmentation spectrum will have all amidated fragments. However, since the M+1 peaks of all the amidated fragments will correlate to the theoretical monoisotopic deamidated fragment peaks, falsely identified deamidated peptides can result if using a relatively wide mass error tolerance. Taken together, the MS, MS/MS spectra and the elution profile of the peptide conclude the peptide as amidated in the above examples.

Deamidated peptides that fell in the range of -5 to 5 ppm where also inspected. **Figure 2.4** illustrates a deamidated peptide correctly identified from MASCOT. The precursor spectrum is shown on the left with an enhanced view of the isotopic cluster for the peptide. From the isotopic cluster it was determined that the 652.29657 peak was the monoisotopic peak. The monoisotopic peak corresponded to that of the deamidated peptide (SMWSVNGDSISK) with a mass measurement accuracy of -2.04 ppm. MS/MS was performed on the 652.29657 peak (♦) and the corresponding MS/MS spectrum is shown on the right. The inset of the MS/MS spectrum is the isotopic cluster corresponding to the y_7 fragment ion. Based on the isotopic cluster of the y_7 fragment ion, it can be determined that this is a deamidated form of the peptide. The mass accuracy of the fragment ion monoisotopic peak is

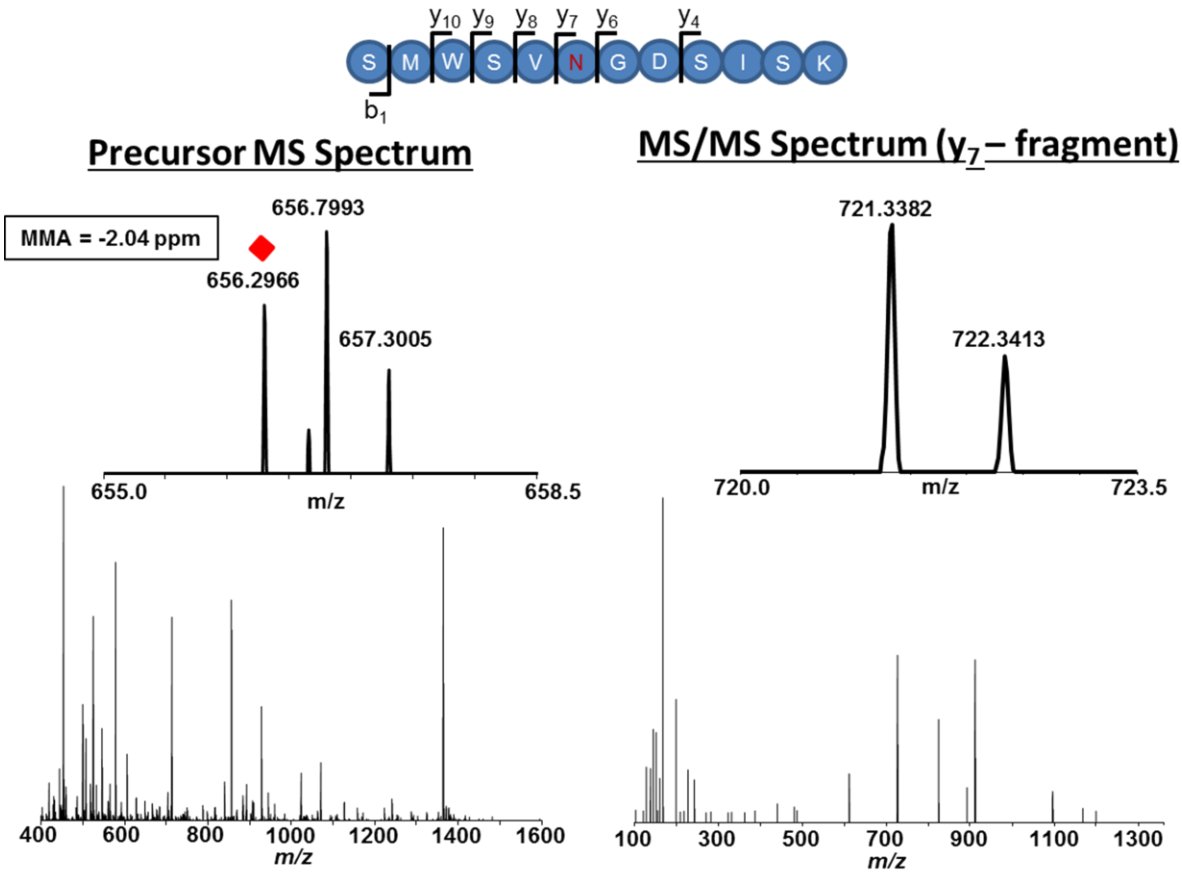


Figure 2.4 The correct identification of a deamidated peptide. The correct monoisotopic peak was chosen leading to a -2.04 mass error. The peptide was then confirmed from the MS/MS spectrum (right). The y_7 fragment ion confirms that it is a deamidated peptide.

2.67 ppm. Since the amidated counterpart was not detected, elution profiles could not be compared. Evidence obtained from the MS and MS/MS data as well as a mass measurement error <5 ppm increases the confidence that a deamidated peptide was accurately identified.

It has been reported that deamidated and amidated peptides will elute at different times depending on the type of chromatography used.[17,29] An example of a peptide that was found to be both deamidated and amidated within the ± 5 ppm range were examined. **Figure 2.5** illustrates the selected ion chromatograms of the identified deamidated and amidated peptide DISTNYYASQKK. The top shows that the deamidated peptide elutes almost 3 minutes prior to that of the amidated form. The mass errors for both peptides were within ± 3 ppm. The combination of unique elution times and high mass measurement accuracies of both forms of the peptide, and supporting evidence from MS/MS spectra, it can be assumed that the deamidated peptide was correctly assigned.

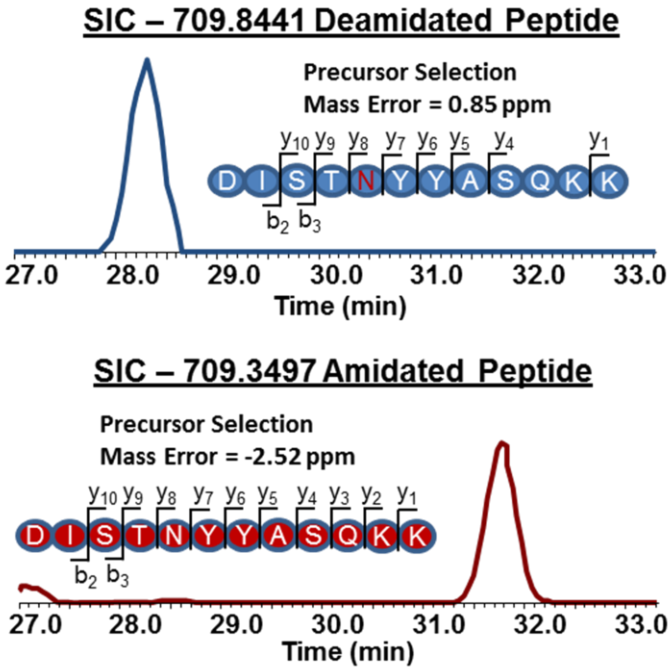


Figure 2.5 Selection ion chromatograms for both the deamidated (top) and amidated (bottom) peptide. Different elution profiles indicate that both are identified correctly.

The accurate identification of the monoisotopic peak obtained in the precursor spectrum is inherently important to the identification deamidated peptides. Proteome Discover was initially used for peak picking procedure for producing MGF files needed for MASCOT searching. As a result, several other peak picking programs were also tested such as MASCOT Distiller and Mass Matrix to evaluate if these programs corrected misidentified precursor peak values. MGF files generated from these programs were searched in MASCOT using ± 100 ppm precursor mass tolerance. The datasets showed similar systematic mass errors for peptides ranging from 5 – 20 ppm (data not shown). However, it is likely that the selection of the monoisotopic peak is a result of data dependent acquisition (DDA) software. Acquiring in a DDA mode, a precursor spectrum is taken populating a candidate peak list of *m/z* values to be used for the following MS/MS spectra. In our study 12 sequential *m/z* values following the precursor spectrum were targeted for MS/MS. Once a candidate was chosen for MS/MS it was no longer a candidate for MS/MS for 30 seconds, which is often the elution time of a peptide. A consequence of using DDA mode, often a candidate is picked during the beginning of its elution profile.It then becomes a possibility to have an incomplete isotopic distribution present in the precursor spectrum shown in **Figure 2.3**. This results in the misidentification of the monoisotopic peak.

To determine if searching with or without deamidation significantly affects the number of proteins identified, the same data was searched without deamidation included. The number of total proteins identified using search parameters with and

Table 2.1 Comparison of the number of proteins identified when searching with (left) and without (right) deamidation (NQ) as a variable modification. The total number of proteins identified is shown when using either ± 5, 20 and 100 ppm as a precursor mass tolerance. All data was searched with ± 0.02 Da MS/MS tolerance.

Searching With Deamidation (N and Q) Variable Modification		Searching Without Deamidation (N and Q) Variable Modification	
Search Parameters	Number of Proteins Identified	Search Parameters	Number of Proteins Identified
±5 ppm Precursor MS Tolerance	4822	±5 ppm Precursor MS Tolerance	4933
±20 ppm Precursor MS Tolerance	5049	±20 ppm Precursor MS Tolerance	4907
±100 ppm Precursor MS Tolerance	4937	±100 ppm Precursor MS Tolerance	4832

without deamidation as a variable modification is presented in **Table 2.1**, the MS tolerances are shown for 5, 20, and 100 ppm. At 5 ppm the total number of proteins identified without deamidation was 4933, including deamidation in the search parameter resulted in 4822 identified proteins. When variable modifications are searched, this further complicates the searching algorithm by increasing the potential m/z space that peptides can occupy. The search space is enlarged within the database because both modified and unmodified forms can potentially exist. Deamidation commonly occurs on 2 different residues which further enlarges the search space. However, when deamidation is included in the search parameters the number of identified proteins is greater using 20 ppm and 100 ppm as the MS tolerance. This is best explained by the misidentified deamidated peptides that fall largely in the 5 to 20 ppm range.

Label free quantitative proteomic techniques involve either the use of area under the curve[30] and/or spectral counting[2,31]. The primary method used in our

laboratory for relative quantitation is spectral counting. To determine if searching without deamidation has an effect on quantitative global proteomics, we compared spectral counts from searches with and without deamidation. Average normalized spectral counts of proteins found (using 5 ppm MS tolerance) with and without deamidation were compared. No significant differences found in normalized spectral counts (NSpC) between proteins searched with or without deamidation. Searching data without deamidation will not ultimately affect the relative quantification of proteins in global proteomic datasets.

2.4 Conclusions

Deamidation is a variable modification parameter often searched because it is a nonenzymatic modification and can occur during sample preparation under basic conditions. Tryptic digestion protocols often involve a pH range between 7-8 which can induce the deamidation of asparagine and glutamine sites. Instrumentation with low resolving power and low mass measurement accuracy can be problematic in identifying the mass shift of 0.98 Da caused by deamidation. For accurate determination of deamidated sites, mass tolerance parameters must be set within 5 ppm for the precursor ion and high resolving power instruments must be used to acquire data. Using wider mass tolerances can lead to the misassignment of these peptides. Omitting deamidation as a variable modification did not significantly impact global quantification data and led to an increase in protein identifications when searching within ± 5 ppm MS window and a 0.02 Da MS/MS tolerance.

2.5 References

(1) Cravatt, B. F.; Simon, G. M.; Yates, J. R., 3rd. The biological impact of mass-spectrometry-based proteomics. *Nature* **2007**, *450*(7172), 991-1000.

(2) Eng, J. K.; Mccormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J Am Soc Mass Spectr* **1994**, *5*(11), 976-989.

(3) Yates, J. R.; Eng, J. K.; Mccormack, A. L.; Schieltz, D. Method to Correlate Tandem Mass-Spectra of Modified Peptides to Amino-Acid-Sequences in the Protein Database. *Anal Chem* **1995**, *67*(8), 1426-1436.

(4) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*(18), 3551-3567.

(5) Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*(9), 1466-1467.

(6) Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **2011**, *10*(9), M111 011015.

(7) Kelstrup, C. D.; Young, C.; Lavallee, R.; Nielsen, M. L.; Olsen, J. V. Optimized Fast and Sensitive Acquisition Methods for Shotgun Proteomics on a Quadrupole Orbitrap Mass Spectrometer. *J Proteome Res* **2012**.

(8) Jung, H. J.; Purvine, S. O.; Kim, H.; Petyuk, V. A.; Hyung, S. W.; Monroe, M. E.; Mun, D. G.; Kim, K. C.; Park, J. M.; Kim, S. J.; Tolic, N.; Slysz, G. W.; Moore, R. J.; Zhao, R.; Adkins, J. N.; Anderson, G. A.; Lee, H.; Camp, D. G., 2nd; Yu, M. H.; Smith, R. D.; Lee, S. W. Integrated post-experiment monoisotopic mass refinement: an integrated approach to accurately assign monoisotopic precursor masses to tandem mass spectrometric data. *Anal Chem* **2010**, *82*(20), 8510-8.

(9) Hsieh, E. J.; Hoopmann, M. R.; MacLean, B.; MacCoss, M. J. Comparison of database search strategies for high precursor mass accuracy MS/MS data. *J Proteome Res* **2010**, *9*(2), 1138-43.

(10) Hao, P.; Ren, Y.; Alpert, A. J.; Sze, S. K. Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Mol Cell Proteomics* **2011**, *10*(10), O111 009381.

(11) Shuford, C. M.; Li, Q. Z.; Sun, Y. H.; Chen, H. C.; Wang, J.; Shi, R.; Sederoff, R. R.; Chiang, V. L.; Muddiman, D. C. Comprehensive Quantification of Monolignol-Pathway Enzymes in Populus trichocarpa by Protein Cleavage Isotope Dilution Mass Spectrometry. *J Proteome Res* **2012**, *11*(6), 3390-3404.

(12) Ren, D.; Pipes, G. D.; Liu, D. J.; Shih, L. Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V. An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Anal Biochem* **2009**, *392*(1), 12-21.

(13) Hong, J.; Schoneich, C. The metal-catalyzed oxidation of methionine in peptides by Fenton systems involves two consecutive one-electron oxidation processes. *Free Radic Biol Med* **2001**, *31*(11), 1432-41.

(14) Morand, K.; Talbo, G.; Mann, M. Oxidation of peptides during electrospray ionization. *Rapid Commun Mass Spectrom* **1993**, *7*(8), 738-43.

(15) Robinson, N. E.; Robinson, A. B. Molecular clocks. *Proc Natl Acad Sci U S A* **2001**, *98*(3), 944-9.

(16) Robinson, N. E.; Robinson, A. B. Amide molecular clocks in drosophila proteins: potential regulators of aging and other processes. *Mech Ageing Dev* **2004**, *125*(4), 259-67.

(17) Dasari, S.; Wilmarth, P. A.; Rustvold, D. L.; Riviere, M. A.; Nagalla, S. R.; David, L. L. Reliable detection of deamidated peptides from lens crystallin proteins using changes in reversed-phase elution times and parent ion masses. *J Proteome Res* **2007**, *6*(9), 3819-26.

(18) Robinson, N. E.; Robinson, A. B. Deamidation of human proteins. *Proc Natl Acad Sci U S A* **2001**, *98*(22), 12409-13.

(19) Li, X. J.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B. Use of O-18 labels to monitor deamidation during protein and peptide sample processing. *J Am Soc Mass Spectr* **2008**, *19*(6), 855-864.

(20) Krokhin, O. V.; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G. Deamidation of -Asn-Gly- sequences during sample preparation for

proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem* **2006**, *78*(18), 6645-50.

(21) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **2009**, *6*(5), 359-U60.

(22) Gokce, E.; Franck, W. L.; Oh, Y.; Dean, R. A.; Muddiman, D. C. In-Depth Analysis of the Magnaporthe oryzae Conidial Proteome. *J Proteome Res* **2012**, *11*(12), 5827-5835.

(23) Andrews, G. L.; Shuford, C. M.; Burnett, J. C., Jr.; Hawkridge, A. M.; Muddiman, D. C. Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, *877*(10), 948-54.

(24) Randall, S.; Cardasis, H.; Muddiman, D. Factorial Experimental Designs Elucidate Significant Variables Affecting Data Acquisition on a Quadrupole Orbitrap Mass Spectrometer. *J Am Soc Mass Spectr* **2013**1-12.

(25) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **2002**, *74*(20), 5383-5392.

(26) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **2003**, *75*(17), 4646-4658.

(27) Weatherly, D. B.; Atwood, J. A.; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R. A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Molecular & Cellular Proteomics* **2005**, *4*(6), 762-772.

(28) Venable, J. D.; Xu, T.; Cociorva, D.; Yates, J. R., 3rd. Cross-correlation algorithm for calculation of peptide molecular weight from tandem mass spectra. *Anal Chem* **2006**, *78*(6), 1921-9.

(29) Krokhin, O. V.; Ying, S.; Cortens, J. P.; Ghosh, D.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. Use of peptide retention time prediction for protein identification by off-line reversed-phase HPLC-MALDI MS/MS. *Anal Chem* **2006**, *78*(17), 6265-6269.

(30) Wiener, M. C.; Sachs, J. R.; Deyanova, E. G.; Yates, N. A. Differential Mass Spectrometry: A Label-Free LC−MS Method for Finding Significant Differences in Complex Peptide and Protein Mixtures. *Anal Chem* **2004**, *76*(20), 6085-6096.

(31) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **2005**, *4*(10), 1487-502.

CHAPTER 3

Ovostatin 2 (OVOS2) is Up-regulated in Chicken and Human Ovarian Cancer

The following work was reprinted from the recently submitted manuscript: Angelito I. Nepomuceno, Huanjie Shao, Kai Jing, David C. Muddiman, James N. Petitte, Michael O. Idowu, Xianjun Fang, and Adam M. Hawkridge. *Proteomics-Clinical Applications.* Submitted*:* July 7, 2013.

3.1 Introduction

Ovarian cancer (OVC) is among the most lethal gynecological cancers in the Western World due to a combination of ineffective early-stage detection methods and late-stage treatment strategies [1]. More than 22,000 women were diagnosed with OVC in 2012 [2], approximately 70% of which present with advanced stages of the disease (Stage III and IV) when surgical intervention and chemotherapeutic treatment are least effective. Early diagnosis is the single most important determinant for survival with 5-year mortality rates less than 10% for women with Stage I OVC. Thus, significant effort has been focused on identifying early-stage markers for OVC including trans-vaginal ultrasound and blood-based proteins (e.g., CA-125 and HE-4) [3]. Although these markers have clinical value for managing OVC, they do not possess the predictive power necessary to be a useful population-based screen [1].

Biomarker discovery in OVC has proven very challenging due to significant tumor heterogeneity [4], the paucity of early-stage biospecimens (i.e., primary tissues and plasma), and few natural animal models that faithfully recapitulate the pathophysiology of the human disease [5-7]. The domestic chicken is a unique

experimental model that develops OVC spontaneously [8,9] with significant morphological [10] and molecular similarities [11-22] to humans. The prevalence of spontaneous OVC in the chicken can exceed 35% depending on the birds' age, genetic background (i.e., strain), and number of eggs produced [9] the latter being consistent with the incessant ovulation theory [23]. Important molecular-level similarities include CA-125 expression [14], frequency of p53 mutations [12], E-cadherin up-regulation [21], and gene expression patterns [22]. There are several advantages to studying spontaneous OVC in the chicken such as a tightly controlled genetic background that is maintained in the agricultural industry [16], a predictable OVC onset window, and control over environment and diet. These are important advantages in biomarker discovery with additional strengths being the ability to draw large amounts of blood longitudinally [24] prior to and after the onset of OVC and collect matched tissue samples from the ovary, oviduct, and neighboring tissues in the abdominal cavity. Longitudinal sampling facilitates the determination of both intra- and inter-individual variability (i.e., personalized- vs. population-based reference ranges), the former of which has been shown to be more sensitive to OVC detection using CA-125 [25-29].

We recently established a biorepository of chicken biospecimens that included longitudinal plasma samples from approximately 250 birds starting at 2.5 years of age over a one-year period combined with matched ovary and oviduct tissue samples [11]. The OVC prevalence for the birds in this biorepository was approximately 10%. In an initial proteomics study, we selected longitudinal plasma

samples from two birds (Healthy and Late-stage OVC) and carried out quantitative mass spectrometry-based proteomics to identify potential biomarkers for OVC [11]. A critical finding in this plasma-based study was the identification of a predicted form of ovomacroglobulin (i.e. predicted form of ovostatin and hereafter referred to as OVOS2) that increased over time with the onset and progression of ovarian cancer. OVOS2 is similar to chicken ovostatin (hereafter referred to as OVOS1) which is expressed primarily in the oviduct and highly abundant in egg-white [30-36]. In a recent report by Lim *et al.* [37], elevated levels of OVOS1 mRNA in chicken ovarian tumors were observed which is interesting because OVOS1 is expressed primarily in the healthy oviducts yet not in healthy ovaries. This finding builds on an earlier report by Johnson and coworkers [18] that showed high levels of ovoalbumin, an oviduct-specific protein, in serous ovarian tumors of the hen. Collectively, these data point to emerging yet undefined roles of ovostatins in ovarian cancer of the chicken and provide unique candidates for human OVC biomarkers.

Herein we report an in-depth proteomic analysis of matched plasma, ovary, and oviduct tissue samples from healthy, early-stage OVC, and late-stage OVC birds from the previously established biorepository [11]. The proteomics dataset showed elevated levels of OVOS2 in the matched plasma, ovary, and oviduct tissues for all 3 late-stage birds yet was not detected in healthy or early-stage OVC birds. Following these results, we BLAST-searched the chicken OVOS2 protein sequence (F1NEW8) against the human UniProt database and identified ovostatin homolog 2 (Q6IE36) as the top hit (45% sequence identity). We then measured the

gene and protein expression of human OVOS2 in OVC cell lines compared with cultures of normal ovarian epithelial (NOE) cells using RT-qPCR and Western blot analysis, respectively. OVOS2 was significantly up-regulated at both the mRNA and protein levels in all 11 OVC cell lines compared to the control NOE cells. Finally, OVC specimens from 5 primary human ovarian adenocarcinomas were immunohistochemically stained for expression of OVOS2. Strong OVOS2 staining was detected in malignant ovarian epithelial cells of all 5 OVC specimens whereas the surrounding non-cancerous tissues were negative or only faintly stained. Collectively, these data provide the first evidence for a conserved, yet undefined, role for OVOS2 in ovarian cancer of the chicken and human.

3.2 Experimental

3.2.1 Animal Care, Biospecimen Collection, and Pathology

All animals were cared for in accordance with North Carolina State University IACUC regulations. Details of the sample collection, sample storage, and bird mortality are described in Hawkridge *et al* [11]. Briefly, approximately 2 ml of blood was drawn from the birds every 3 months for 1-year starting at 2.5 years of age followed by centrifugation at 3,000 × *g* for 3 min at 10°C and plasma collection. Birds were euthanized at the final blood draw followed by ovary and oviduct tissue collection. Tissues were preserved for both pathological (formalin fixation) and proteome (LN_2 snap frozen) characterization. 5 µm thick tissue slides were stained

with hematoxylin and eosin (H&E) and assessed by a board-certified veterinary pathologist.

3.2.2 Chicken Tissue Lysate Preparation and SDS-PAGE.

Frozen ovary and oviduct tissues were removed from -80°C storage, weighed (frozen), and then placed on ice. Lysis buffer was added at a ratio of 5 µL:1 mg of tissue. The lysis buffer consisted of 50 mM Tris (Sigma Aldrich, St. Louis, MO, USA), 8 M Urea (Sigma Aldrich), 2 M Thiourea (Sigma Aldrich), 10 mM EDTA (Sigma Aldrich), 10 mM DTT (Sigma Aldrich), 0.001% sodium azide (Sigma Aldrich) and was adjusted to a pH of 7.8 using hydrochloric acid (Sigma Aldrich). The sample was then homogenized using an OMNI TIP Homogenizing Kit (Omni International) for 1 min. Sodium dodecyl sulfate (SDS, Sigma Aldrich) was then added to make up a final concentration of 2% (w/v) and then placed onto the Genie Disrupter (Scientific Industries) for 1 minute followed by 5 minutes of incubation period on ice. This step was then repeated twice more. The sample was then centrifuged for 30 minutes at 14,000 rpm. The soluble fraction was removed from the centrifuge tube without disrupting the cellular debris pellet or the top lipid layer. The extracts were then stored at -80^{o}C until immediately prior to 1D SDS-PAGE.

The ovary and oviduct lysates were allowed to thaw to room temperature and briefly vortexed. The lysate sample was then diluted with 50 mM Tris-HCl pH 6.8 in a 1:5 (*v:v*) ratio. This mixture was then combined with a freshly prepared 2-mercapthethanol/Laemmli (Bio-Rad, Hercules, CA, USA) loading buffer in a 1:1 (v:v)

ratio. The resulting solution was vortexed and then heated at 95^{o}C for 5 minutes. A total of 30 µL of solution was loaded onto the Bio-Rad Criterion Tris-HCl precast 1D gel (12.5% Tris-HCl, 12+2 well format) and were run at a constant 200 V for approximately 50 minutes. The gel was then stained using a Coomassie G-250 (Bio-Rad) followed by destaining and imaging.

3.2.3 Plasma Preparation and SDS-PAGE

Total protein concentrations for plasma samples were determined using a Bradford assay (Coomassie Plus^{TM}) and aliquots of each sample were diluted with 15 µl 50 mMTris-HCl (pH 6.8). The sample was then combined with freshly prepared 2-mercaptoethanol/Laemmli loading buffer in a 1:1 (*v:v*) ratio. Solutions were briefly vortexed and then heated at 95^{o}C for 5 min. The total protein that was loaded onto a 12.5% Tris-HCl precast gel (Bio-Rad Criterion 12+2 well format) ranged from 20-32 µg. The gel was run at a constant 200 V for approximately 50 minutes. The gel was then stained using a Coomassie G-250 (Bio-Rad) followed by destaining and imaging.

3.2.4 In-gel Digestion

The gels were processed in a Laminar flow hood by cutting 36 even gel bands per lane (2 mm x 7 mm) using a grid cutter (The Gel Company, San Francisco, CA) and then transferring the 12 × 3 neighboring bands into individual vials for in-gel digestion [38]. 150 µL of a 50:50 mixture of 100 mM ammonium

bicarbonate (Sigma Aldrich) and acetonitrile (ACN) (Burdick and Jackson, Muskegon, USA) was added to each fraction and incubated for 30 minutes at room temperature. The solution was removed, discarded, and then 200 µL of ACN was added to each vial for 20 minutes to dehydrate the gels. The ACN was removed and replace with 150 µL of 10 mM dithiothreitol (DTT) followed by a 30 minute incubation at 56°C. The DTT solution was removed and the gel bands were again dehydrated with 200 µL of ACN. The ACN was removed and replaced with 55 mM iodoacetamide (Sigma Aldrich) followed by a 20 minute incubation at room temperature in the dark. The solution was removed and then replaced with 150 µL of 5 µg/mL of trypsin (Promega, Madison, WI, USA) followed by overnight incubation at 37°C. The digestion was quenched with a 1:2 mixture of 5% formic acid:ACN at 37^0C for 30 minutes. Each fraction was then dehydrated using a 2:1 mixture of ammonium bicarbonate and ACN. The tryptic peptide solution was removed and reduced to dryness in a Speedvac.

3.2.5 LC-MS/MS

A nanoLC-2D (Eksigent Technologies, Dublin, CA) was coupled to a LTQ-FT-XL (Thermo Scientific, San Jose, CA) using a vented column configuration [39]. The vented column consisted of a 75 µm x 5 cm trap (IntegraFrit: New Objectives, Woburn, MA) coupled to a 75 µm x 15 cm column (PicoFrit: New Objectives, Woburn, MA). Both the trap and analytical column were self-packed with Magic C18AQ stationary phase (5 µm particle, 200 Å pore, Auburn CA). Mobile phase A

and B were composed of water/acetonitrile/formic acid (98%/2%/0.2% and 2%/98%/0.2% respectively). The Speedvac-dried fractions were reconstituted in 65 µL of mobile phase A. 10 µL of sample was loaded onto the trap at 1.5 µL/min with 2% B. After washing the sample, the valve diverted the flow onto the column and the following gradient was applied at 350 nL/min: 2% B (0-5 min), 2-60% B (5-65 min), 40-90% B (65-67 min), 90% B (67-77 min), 90%-2% B (77-78 min), 98% (78-90 min). Data acquisition was initiated 5 min after the start of the nLC gradient. The sample analysis order for plasma and ovary tissues was randomized and run in triplicate with a blank injection every thirteenth LC-MS/MS run. The oviduct tissues were single LC-MS/MS runs.

Mass spectrometric analysis was performed on a 7T LTQ-FT Ultra (ThermoFisher) with a pulse sequence consisting of a broadband acquisition in profile mode followed by 8 data dependent MS/MS events in the ion trap. The LTQ-FT was mass-calibrated with a target AGC limit of 1×10^6 every 40 LC-MS/MS runs (~2.5 days). The full scan was collected at 400-1600 *m/z* with a resolving power was set at $100{,}000_{FWHM}$ at *m/z* = 400. The dynamic exclusion time was set to 3 min.

3.2.6 Data Analysis

LC-MS/MS data was searched against a concatenated target-reverse chicken database (IPI ver.3.74) with *Homo sapiens* keratin, keratin related proteins, and porcine trypsin included. Mascot Distiller (ver. 2.3.2, Matrix Science, Boston, MA) was used to generate peaks lists, and then Mascot Daemon to perform the

searches. Carbamidomethyl (C) was set as a fixed modification and deamidation (NQ), oxidation (M), and carbamyl (K and N-term) were set as variable modifications. Additional search parameters included maximum of 2 missed cleavages, peptide tolerance of ± 5 ppm, and MS/MS tolerance of ±0.6 Da. Protein grouping, statistical filtering (1% false discovery rate), and quantification of Mascot DAT files were accomplished using ProteoIQ (NuSep, Athens, GA, ver 2.3.01), a spectral counting label-free software package that uses a combination of Peptide/Protein Prophet [40,41], and PROVALT [42].

In this study, a spectral count (SpC) represents a confidently identified tryptic peptide and a Normalized SpC (NSpC) is calculated with two levels of normalization. The first normalizes the spectral counts for the three replicates within a sample and the second normalizes spectral counts between samples for a specific class (i.e., plasma, ovarian tissue, and oviduct tissue). For example, if a plasma sample (i.e., a single gel lane) gave a total SpC for replicates 1, 2, and 3 of 1000, 900, and 800, respectively; the normalization factor for all proteins (protein groups) in each replicate would be 1000/1000 = 1.00, 1000/900 = 1.11, and 1000/800 = 1.25. Thus, if the SpC for the same protein in replicates 1, 2, and 3 was 15, 13, and 12, respectively; their normalized SpC values would be 1.00×15 = 15.00, 1.11×13 = 14.43, and 1.25×12 = 15.00, respectively. The second and final level of normalization takes the maximum SpC from the replicate of all samples in a specific class (e.g., all plasma sample gel lanes). If the maximum total SpC from a replicate in all plasma samples was 1200, the total SpC's from replicate 1 in the plasma

sample above would be normalized by a factor of 1.20 (1200/1000). The reported NSpC's in this study for the plasma sample above would then be 1.20×15 = 18.00, 1.20×14.43 = 17.32, and 1.20×15 = 18.00.

3.2.7 Human Cell Lines

The sources and culture conditions of the eleven human ovarian cell lines (Caov-3, SKOV3, HEY, ES-2, Dov-13, OVCA420, OVCA429, OVCA432, OVCA433, 2780, and 2780R) and two specimens of normal ovarian epithelial cells in culture (NOE008, NOE72) were described previously [43-45]. The cells in early passages were maintained in culture for less than 4 weeks before isolation of total cellular RNA with TRIzol following the protocol of the manufacture (Invitrogen)

3.2.8 RT-qPCR

Expression levels of OVOS2 mRNA were measured by RT-qPCR. Complementary DNA (cDNA) was synthesized from RNA (1 μg, random primers), using the High-Capacity cDNA Reverse Transcription Kit from Applied Biosystems. The relative levels of OVOS2 were determined by the gene-specific OVOS2 probe, the TaqMan Universal PCR Master Mix and the 7900HT Real-Time PCR System (Applied Biosystems). The results of OVOS2 were normalized to the levels of GAPDH and presented as per mil of GAPDH.

3.2.9 Western Blot Analysis

Whole cell lysates were prepared with cell lysis buffer (Cell Signaling Technologies, MA, USA), separated on SDS-PAGE electrophoresis and transferred to PVDF membrane (Bio-Rad, Hercules, CA). Blots were probed with affinity purified polyclonal antibodies against OVOS2 (Abgent) or GAPDH (Cell Signaling) following the protocols of manufacturers. Immunocomplexes were visualized with an enhanced chemiluminescence detection kit from Amersham.

3.2.10 Immunohistochemistry

The paraffin-embedded ovarian cancer blocks were preexisting specimens obtained from the VCU Tissue & Data Acquisition & Analysis Core. These samples had patient consent for use in scientific research but were not collected specifically for the present work. The use for OVOS2 study was approved by the Virginia Commonwealth University Human Subjects Committee. Immunohistochemical staining of paraffin-embedded sections was conducted using a standard procedure as described previously [46].

3.3 Results and Discussion

3.3.1 Quantitative Proteomics Analysis of Ovarian Cancer in the Chicken

The pathologies and tumor locations for the 9 birds in this study are shown in **Table 3.1**. Birds 1-3 were assessed 'Healthy' with no visible neoplastic lesions present on any internal organs. Birds 4-6 were 'Early Stage OVC' with tumors

Table 3.1 Pathological classification of the birds used in this study and tumor location.

Bird Identification #	Pathology	Neoplastic Lesions
1	Healthy	No visable neoplastice lesions
2	Healthy	No visable neoplastice lesions
3	Healthy	No visable neoplastice lesions
4	Early Stage OVC	Ovary
5	Early Stage OVC	Ovary
6	Early Stage OVC	Ovary
7	Late Stage OVC	Ovary, Oviduct, Intestine/Pancrease
8	Late Stage OVC	Ovary, Intestine/Pancrease, Liver
9	Late Stage OVC/OVD	Ovary, Oviduct, Intestine/Pancrease

confined solely to the ovary and Birds 7 and 8 were 'Late Stage OVC' with tumors present on the ovary and neighboring tissues (Bird 7 oviductal tumor was benign). Bird 9 is classified as 'Late Stage OVC/OVD' because tumors were present on the ovary and neighboring tissues including the oviduct (OVD). In cases where tumors are present on both the ovary and oviduct (which are common in the chicken), it is difficult to confidently assign the tissue origin of the tumor.

The matched plasma, ovary, and oviduct tissue proteomes of the 9 birds in **Table 3.1** were prepared by in-gel digestion followed by nLC-LTQ-FT mass spectrometry (GeLC). The 1D gels for plasma, ovary lysates, and oviduct lysates are shown in **Figure 3.1A, B, and C**, respectively. The size and location of the gel bands for each gel lane are given on the right side of each gel image and the total numbers of unique proteins identified in each sample are listed below each respective lane. A summary for the plasma proteome dataset is shown in a Venn diagram for all 9 birds in **Figure 3.1A.** The plasma analysis resulted in the

identification of 248 unique proteins (i.e., protein groups), 227 of which were shared by all 9 birds. A total of 2680 proteins were identified GeLC of the ovary lysates and

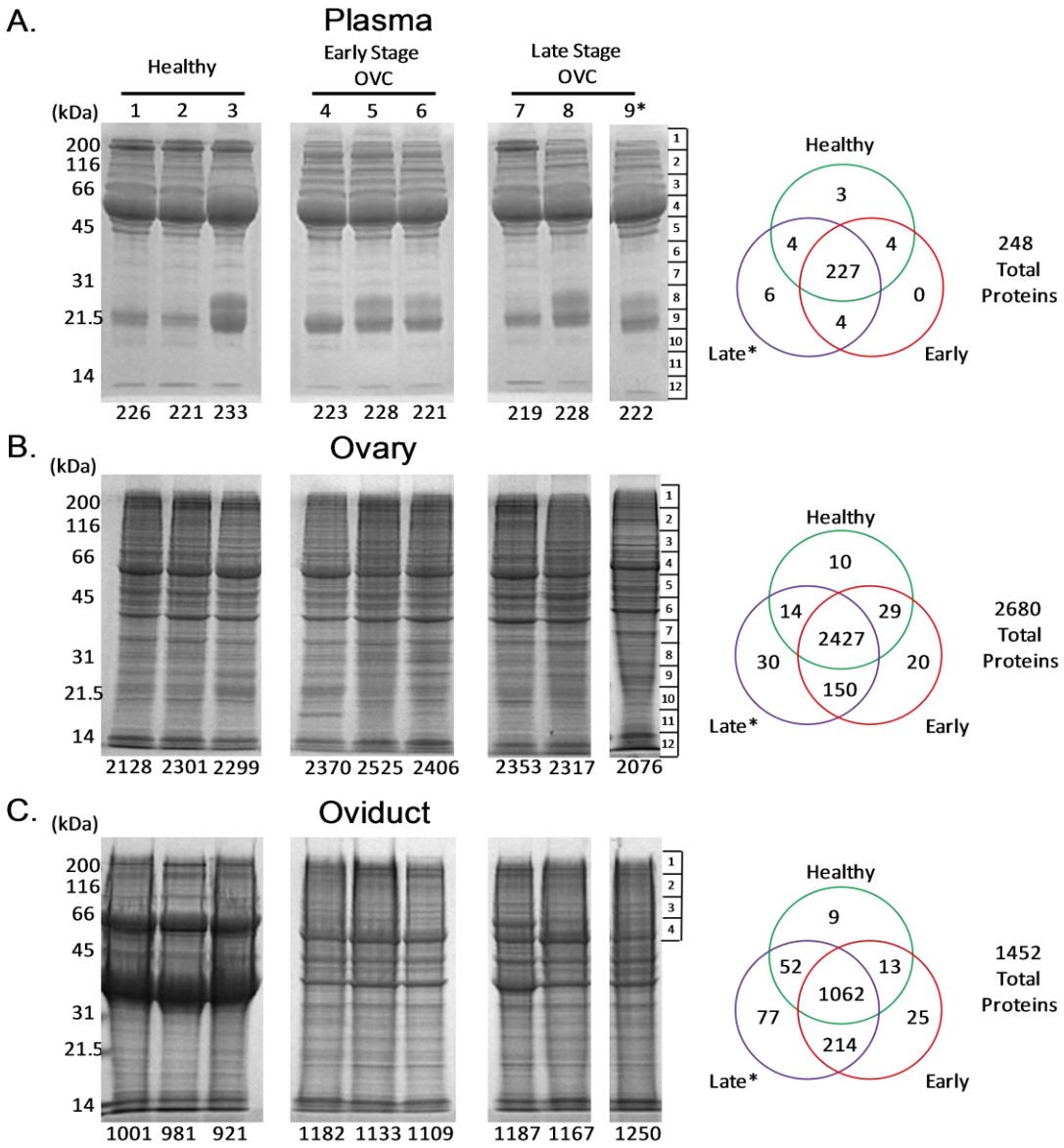


Figure 3.1 One-Dimensional gels of nine bird samples that originated from plasma (A), ovary tissue lysates (B) and oviduct lysate (C) of each respective bird. The Venn diagram to the right of each gel indicated the overlapping proteins between Healthy (green), Early- (red), and Late--stage (purple) birds. The total number of protein groups identified in each lane is given below each gel. (*) denotes that inclusion of bird #9 as a Late-stage EOC bird.

the proteome distribution is shown in the **Figure 3.1B** Venn diagram. Out of the 2680 unique proteins identified in the ovaries, 2427 of which were shared amongst all 9 birds. GeLC of the >50kDa fraction of the oviduct lysates indentified a total of 1452 unique proteins, 1062 of which were shared by all 9 birds (**Figure 3.1C**). The >50kDa oviduct fractions were analyzed to establish the presence or absence of OVOS1 and OVOS2 which are both >150 kDa.

Normalized spectral counting (NSpC) [47-49] was used to measure quantitative changes in the GeLC proteomics dataset. Thus, relative changes in protein abundance are measured as a function of the number of confidently identified tryptic peptides. A subset of proteins and their expression levels is provided in **Figure 3.2** for discussion. Proteins were grouped into either 'Macroglobulins', 'Reproduction/Ovulation', or 'Inflammation' as a function of Bird # and tissue type. The average Healthy NSpC values are listed adjacent to each heat map expression row for reference and asterisks are used to denote imputed data.

The data in **Figure 3.2** for the Macroglobulin proteins shows that OVOS2 levels are significantly elevated in all 3 cases of advanced OVC. A2M would not be expected to correlate because it is primarily synthesized in the liver. OVOS1 was not detected in any of the plasma or ovary tissue samples in contrast with the mRNA results by Lim *et al*. [37] but it was highly abundant in the Healthy oviduct tissues (Avg NSpC = 181.7). The OVOS2 levels for the Late Stage OVC plasma, ovary, and oviduct samples were over 2.5 fold higher relative to the Health birds (**Figure 2)**. Low OVOS2 NSpC values were observed in the plasma and ovaries for most

Healthy and Early-stage OVC birds but no OVOS2 peptides were detected in the oviducts for these 6 birds.

Several Macroglobulin (or Macroglobulin-like) proteins in the chicken were detected in this study (**Figure 3.2)** and are considered functional homologues to the human I39 family. The MEROPS I39 protease inhibitor family, which includes the abundant plasma protein alpha-2 macroglobulin, exhibit broad substrate specificity for all four classes or proteases [50]. Ovostatin (OVOS1; P20740) is an abundant protein in egg white [32,33] that has been measured at the protein level with 42%

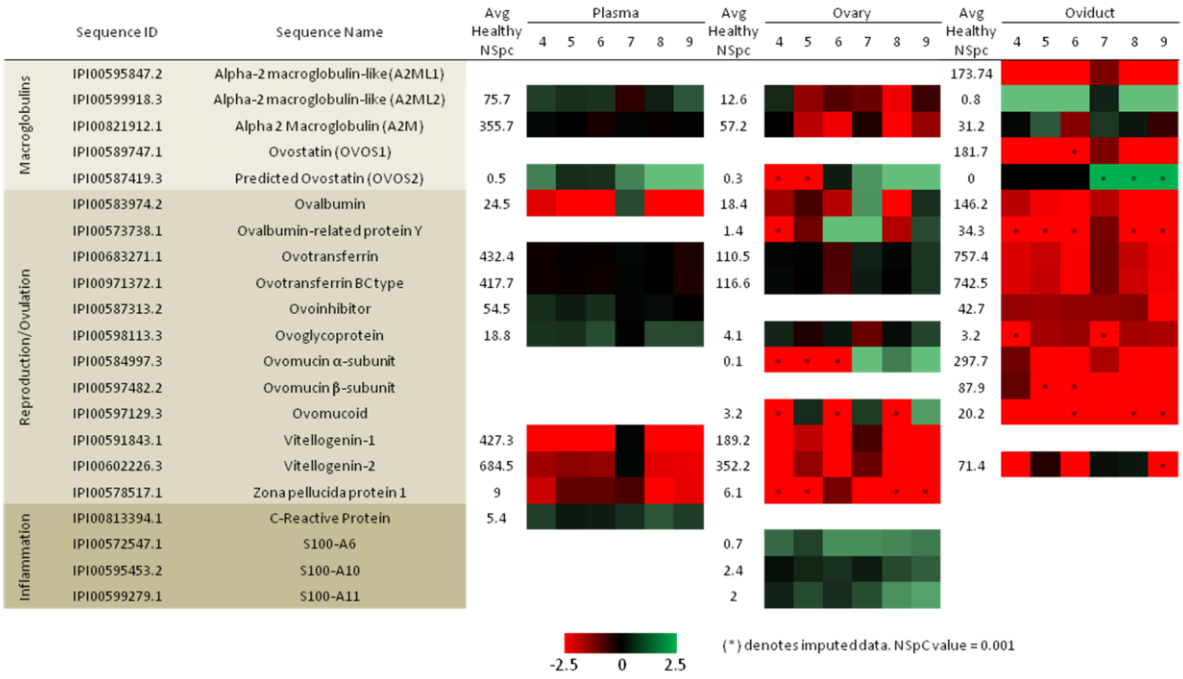


Figure 3.2 Heat map of the $\log_2$ expression values for proteins in three main groups (Macroglobulins, Reproduction/Egg Production, and Inflammation) as a function of tissue type (plasma, ovary, and oviduct) and phenotype. Birds 4-6 are Early-stage OVC and birds 7-9 are Late-stage OVC. The ratio cut off was set at -2.5 and 2.5 fold change. Squares with asterisks (*) denotes imputed data (NSpC = 0.001 was used for calculations).

sequence coverage by mass spectrometry-based proteomics [35]. OVOS1 is expressed primarily in the oviduct [33] and functional studies have confirmed it has similar substrate specificity to the I39 protease inhibitor family [30-33]. The up-regulation of OVOS1 mRNA in chicken ovarian cancer was recently reported by Lim *et al.* [37] using a combination of RT-qPCR and in-situ hybridization of fixed tumor tissues. The authors used alpha-2-macroglobulin (A2M) and ovostatin interchangeably but their PCR primers appeared to be specific for OVOS1. The data showed that for certain ovarian tumor histopathological subtypes, OVOS1 was over-expressed relative to healthy ovarian tissues. This report builds on earlier observations by Johnson and co-workers who reported the presence of ovoalbumin, an oviduct-specific protein, in chicken ovarian tumors suggesting it may be a model for oviductal (i.e., fallopian tubes) involvement in human serous ovarian cancer [18]. There is increasing evidence [51-54] that supports oviductal involvement in serous OVC which conflicts with the established hypothesis that OVC originates from the ovarian surface epithelium (OSE) [55].

OVOS2 (F1NEW8) was a predicted 165 kDa protein in the IPI Chicken database based on EST data until our original proteomics study of ovarian cancer in the chicken [11] detected it at the protein level with 17% sequence coverage in the final plasma sample (5th timepoint) of a Late-stage OVC bird (Bird #8 in this study). In the present study with a higher level of protein fractionation, we were able to detect OVOS2 in the same sample with 37% sequence coverage **Figure 3.3**. Furthermore, we have developed a SRM-based protein cleavage isotope dilution

OVOS2 (IPI00587419.3/UniProt F1NEW8)

MAYEEVLKTNNGCQYVLLVPTVVRSNSPQTACVQFQSLSEPLSLSVVLEYSNIQTTLFEE
FVTKNDYFTCCEFKVPPATSDPLAFISFSAKGNTVNLTERRSVAIENVHNTLFIQTDKPI
YKPGQKVMFRVVTLDSQFRPVQETYPRIIIKDPEQNQIFQWLDVSSMNGIIQLSFPLIEE
PILGSYHIVVEKKSGDKEHEYFTVEEYGKQALQYVLLVPTVVRSNSPQTACVQFQSLSEP
LSLSVVLEYSNIQTTLFEEFVTKNDYFTCCEFKVPPATSDPLAFISFSAKGNTVNLTERR
SVAIENVHNTLFIQTDKPIYKPGQKVMFRVVTLDSQFRPVQETYPRIIIKDPEQNQIFQW
LDVSSMNGIIQLSFPLIEEPILGSYHIVVEKKSGDKEHEYFTVEEYVLPKFEVTTSMPRR
ISFFDEEIRVNVCALYTYGQPVQGSARINVCQRHFYNPQCQQSQKPRCEAVIGLLENNGC
LSTVVSTKTFQLYRSYARMYASFNIETIVTENGTGIQMKNYDYVAVSQENDRVMFRNMDQ
YYRRGIPYFGEITVTNADGKPVPSRVVVLEVNGEYQANYTTDENGTAAFSLDTSNFFNPT
VKLRATQAPDDCADFFMWRNDHESEALFFVRRFYSRTNSFVRIEPVEEKLRCGQQRMINI
HYVLSRKGYRNATHTDFYYVVMAKGKIVLSGQKQVRIAHAPWGTFAITLDVTEKLTPSAR
LLLYTVHPDGEIVADSSWIHSDVCFKNKLQLEFSEKQAYPGSKINIHLEAAANSYCALRA
VDQSVFLLQPERELSAESVYYRLHLSDLYGYYYNGLNLQDDPPEECTPVKTTFFDGLYYE
PVNVSHDGDVYRIFREMGLKVFTNSTLRKPVLCNEDKLDWEENRIYFEHGASGGSAFGKE
FSKITAAGVVNTVRKYFPETWIWDLVHTDSTGEANIFYTVPDTITEWKASAFCLQDDAGF
GISSPVSLTAFQPFFVDLALPYSVIRGEKINLIANIFNYLDKCIQISAILAESSDYKAEV
LSPEGNTARVCANERKTYIWAVSPLSLGEVKFTITAEAKLNTKGAKNSTPPEEESIRTDT
LTQTLLVEPEGIKKELTQSSLICTKGTTVSEPVLLSLPRNVVQGSARVYFSVIGDILGTA
LRNMENLLHMPYGCGEQNMALFTPNIYVLDYLNKTGQLTEEIRVKSTGYLTTGYQKQLSY
KHQDGSYSSFGTRDKEGNVWLTAFVYKSFAQARRYIYVDENVQSQTLIWLARKQKSDGCF
ENAESHFNNALKGGEEGEYSLTAYIVAALLEAGHSVQHPVVWNGMNCLETAFSNGVHNLY
NHALFAYVYGLADKQERYQYFLEKLDKRATRDGGSVYWQRENKPPAEHFPAFYSRAPSAE
IEMTSYVLLALLNKAELTPDDLSYISRIVYWLVKQQNPYGGFSSSQDTVVAIQALAQYGY
LTFSKESHNTVKVNFMEIPKKAFQVNDENRFLLQQTSLPTVLGNYSVEVYGTGCVYMQTT
LKYNIHLPKKAAGFFLSVEPANVSCTSNFPLKFDLVFSASYTGNRNVSNMAIIDVKMLSG
FISDRSSLKKLQYQASVVDHVDIKNDHVFFYLQKLSQKEVSFSFSVEQSLPVSDIKPAPV
HIYDYYET

Figure 3.3 Chicken OVOS2 protein sequence with the LC-MS/MS detected tryptic peptides shown in red. Sequence coverage = 37%.

assay with 3 stable isotope labeled OVOS2 tryptic peptides which collectively gives us a high degree of confidence that we have detected this gene at the protein level [56]. We are unaware of any prior reports that detected chicken OVOS2 at the protein level and therefore are not aware of any functional analysis that confirms OVOS2 has protease inhibitor activity or that the biological function and tissue specificity is comparable to OVOS1. Herein we tentatively assign chicken OVOS2 as an I39

protease inhibitor based on the 50% sequence identity to chicken OVOS1 and the predicted function from the UniProt database.

The quantitative data for the proteins related to Reproduction and Ovulation are included to explain the broad down-regulation of these proteins in the Early- and Late-stage birds. Egg production can have a profound effect on the protein expression levels in plasma, ovary, and oviductal tissues in the bird. All Healthy birds were producing eggs at the time of necropsy with the strongest molecular-level indicator being high plasma levels of Vitellogenin-1 and -2, both lipid transport proteins synthesized in the liver. With the exception of Bird #7 which was also producing eggs at necropsy, the Early- and Late-stage birds were out-of-production with significant down-regulation of Vitellogenin-1 and -2 in plasma. The expression levels of OVOS1 and A2ML1 levels in the oviduct were nearly identical for the healthy and OVC birds with a strong correlation to egg-production. Furthermore, both OVOS1 and A2ML1 have been measured with mass spectrometry-based proteomics at high levels in the chicken egg white [35].

Ovomucin α-subunit (also referred to as MUC-5B, UniProt Q98UI9), an abundant protein in egg white [57], is up-regulated in the Late-stage OVC tumor tissues, down-regulated in oviduct tissues, and not detected in any of the plasma samples. Chicken MUC-5B was BLAST-searched against the Human UniProt database and resulted in a 45% sequence identity to MUC-5AC (P98088), a gel-forming protein found in gastric and respiratory tract epithelia [58]. Dis-regulation of MUC-5AC has observed in several cancers [59-61] including ovarian neoplasms [62,63].

Thus, the up-regulation of MUC-5B in the Late-stage OVC tissues of the chicken agrees with earlier OVC studies in humans.

The inflammation response to the onset and progression of cancer has been well established [64-66]. C-reactive protein (CRP) is a 23kDa acute-phase protein synthesized in the liver, circulates in blood as a homopentamer, and can increase in concentration by as much as 1000-fold in response to inflammation [67]. CRP levels in plasma were found to correlate with the Stage of OVC in humans [68,69]. Higher concentrations of CRP in the plasma of chickens with OVC were observed in the **Figure 3.2** NSpC data and correlate with Healthy, Early-Stage, and Late-Stage OVC birds. Box-whisker plots of the NSpC CRP levels as a function of the three phenotypes are shown in **Figure 3.4** with the statistically significant differences ($p < 0.05$) for Healthy, Early-stage, and Late-stage birds. The S100 class proteins also show increased NSpC's in ovarian tumor tissues relative to healthy ovaries (**Figure 3.2**). S100 proteins are ~12 kDa that bind calcium in a dimeric structure and are involved in several types of inflammation responses. S100A6 levels in plasma were recently shown to predict tumor burden in a xenograft mouse model of ovarian cancer [70]. OVOS2 may also be involved in the inflammation response in the tumor microenvironment. There are reports showing that human A2M is involved directly and indirectly in the inflammation response for several diseases illustrating a complex role beyond proteolytic inhibition [71]. Therefore, we performed further studies to examine the role of OVOS2 in the context of human OVC.

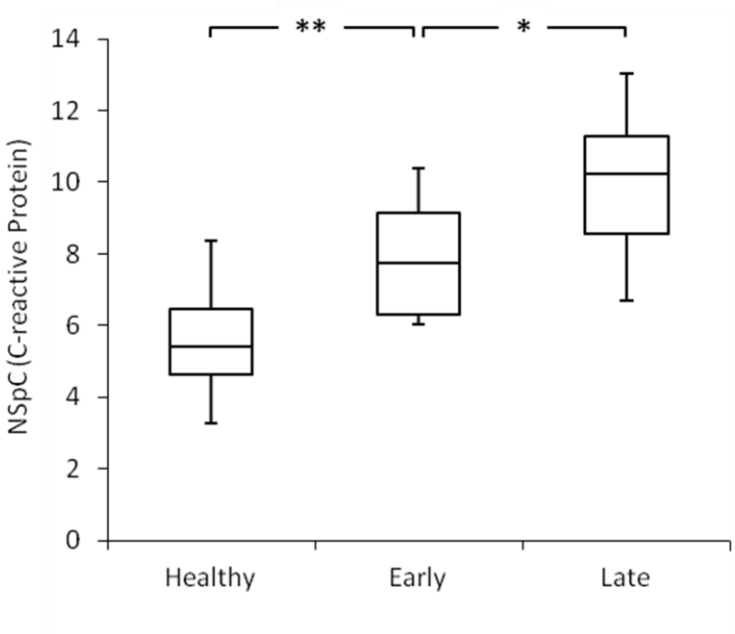


Figure 3.4 Heat map of the $\log_2$ expression values for proteins in three main groups (Macroglobulins, Reproduction/Egg Production, and Inflammation) as a function of tissue type (plasma, ovary, and oviduct) and phenotype. Birds 4-6 are Early-stage OVC and birds 7-9 are Late-stage OVC. The ratio cut off was set at -2.5 and 2.5 fold change. Squares with asterisks (*) denotes imputed data (NSpC = 0.001 was used for calculations).

3.3.2 Up-regulation of OVOS2 in Human Ovarian Cancer

To assess the presence of OVOS2 in human OVC, we measured OVOS2 mRNA and protein levels in eleven human OVC cell lines and two normal ovarian epithelial (NOE) cell cultures by RT-qPCR and Western blot analysis, respectively. The qPCR results are shown in **Figure 3.5A** with the OVOS2 mRNA levels normalized to Ct values of GAPDH. Importantly, the qPCR product for OVOS2 mRNA analysis spanned an exon to prevent the detection of genomic DNA (Hs00416380_m1, Applied Biosystems). OVOS2 mRNA levels in all 11 OVC cell lines were significantly higher ($p<0.05$) than in NOE cells. OVOS2 mRNA levels in NOE008

and NOE72 were statistically similar but the mean OVOS2/GAPDH for NOE72 was slightly higher (0.062 vs. 0.061). Fold increases in expression of OVOS2 over the NOE72 cells ranged from 3- (2780) to 35-fold (OVCA432). In agreement with overexpression of OVOS2 mRNA in OVC cell lines, Western blot analysis confirmed the presence of high levels of OVOS2 protein of 100−150 kDa in all OVC cell lines but not in NOE cells (**Figure 3.5B**).

To confirm the expression of OVOS2 in primary human tissues, we performed immunohistochemical staining on 5 Stage III, high-grade serous ovarian carcinomas, the most common human ovarian cancer. The paraffin-embedded sections were stained with the same anti-OVOS2 antibody that was used for the Western blot analysis in **Figure 3.5C**. In all 5 OVC tumor specimens, malignant epithelial cells were more intensely stained compared to the surrounding non-cancerous tissues. **Figure 3.5C** shows representative images of OVOS2 staining in 4 OVC patient tissues.

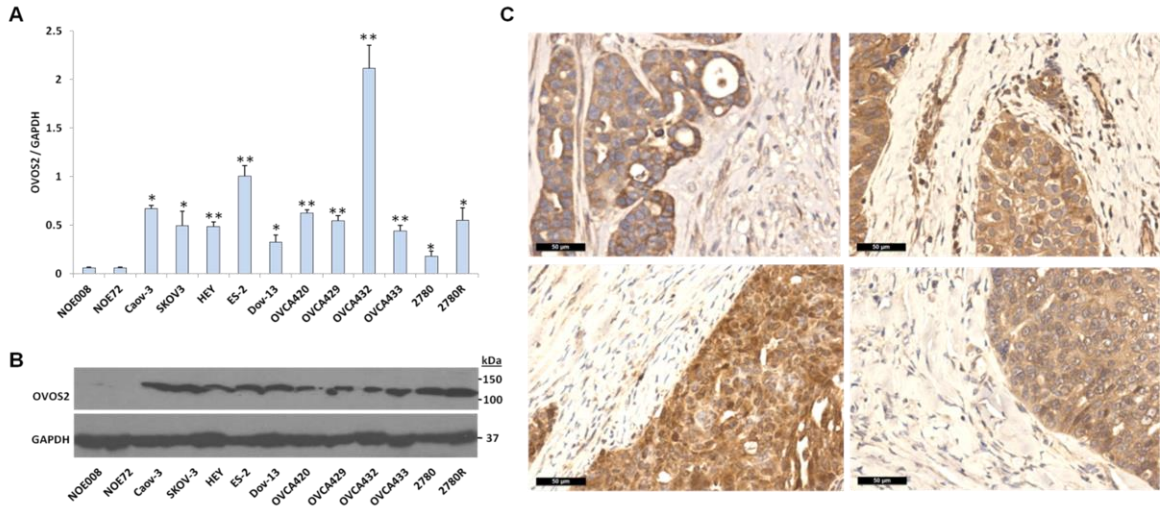
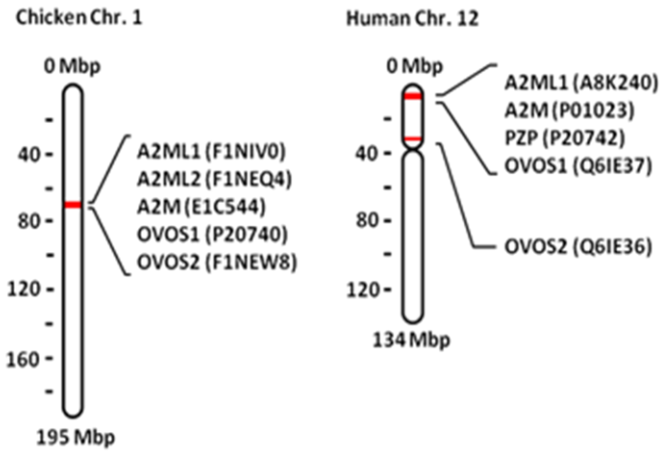


Figure 3.5 RT-qPCR (A) Western blot (B) analysis of OVOS2 expression in human ovarian cancer cell lines and normal ovarian epithelial cells. The OVOS2 mRNA results were normalized with GAPDH as internal control and presented as per mil of *Ct* values of NADPH. The statistical significances between each of ovarian cancer cell lines (Caov-3, SKOV3, HEY, ES-2, Dov-13, OVCA420, OVCA429, OVCA432, OVCA433, 2780, and 2780R) and NOE72 cells were determined with the Welch's t-test. * p-value < 0.05; ** p-value < 0.01. (C) IHC staining of OVOS2 protein in human ovarian cancer tissues. Shown are positively stained malignant epithelial cells relative to the neighboring non-cancerous tissues from patients with Stage III, high grade serous ovarian adenocarcinomas. Bar = 50 μm

3.3.3 OVOS2 in Chicken and Human Ovarian Cancer

The functional annotation, tissue distribution, or disease-specific association of human ovostatins has not been explored to the best of our knowledge. The chicken OVOS2 gene resides on the petite region of Chromosome 1 clustered with OVOS1, A2M, A2ML1, and A2ML1 (**Figure 3.6**). There is conserved synteny [72] between this region of Chromosome 1 and the petite region of Human Chromosome 12 as illustrated in **Figure 3.6** and shown in greater detail by Scherer *et. al.*.[72] All of

these proteins belong to the MEROPS I39 family of protease inhibitors with the ranked matches to chicken OVOS2 (F1NEW8) in humans shown in the **Figure 3.6** table inset. The interplay between proteases and their endogenous inhibitors have been actively studied in relation to cancer for many years. The dysregulation of several protease inhibitors have been observed in ovarian cancer including serpins (i.e., SERPINA1/alpha 1-antitrypsin[73], SERPINE1/plasminogen activator inhibitor-1 [74], SERPINB2/plasminogen activator inhibitor-2 [74,75], SERPINA5/plasminogen activator inhibitor-3 [76], SERPINB5/maspin [77]), tissue inhibitors of metalloproteases (TIMP) [78,79], secretory leukocyte protease inhibitor (SLPI) [80,81], and now OVOS2 in this study. The conventional view of cancer progression would suggest that up-regulation of extracellular protease inhibitors like OVOS2 are beneficial endogenous responses to limiting the proteolyisis of extracellular matrix proteins which leads to metastasis. However, functional studies have shown that many endogenous protease inhibitors can actually promote cancer cell growth and metastasis through complex interactions or as indirect products from key regulators (e.g., colony stimulating factor 1 [74,75]). Thus, the up-regulation of OVOS2 in OVC adds to a growing list of dysregulated protease inhibitors in ovarian cancer and their complex roles in cancer.



BLAST Results for Chicken OVOS2 (F1NEW8)

Human Gene	Accession (UniProt)	BLAST Identity (% AA match)
OVOS2	Q6IE36	45%
A2M	P01023	43%
PZP	P20742	42%
A2ML1	A8K240	37%
OVOS1	Q6IE36	36%

Figure 3.6 Map of conserved synteny for Macroglobulins on chicken Chromosome 1 and human Chromosome 12. The top scoring BLAST results for chicken OVOS2 against the human UniProt database are shown in the inset table.

3.4 Conclusion

An in-depth label-free proteomics study of matched plasma, ovary, and oviduct tissues in the hen resulted in the quantification of over 2680 unique proteins representing the most comprehensive proteomics study of ovarian cancer in the chicken to date. Proteins related to reproduction (e.g., Vitellogenin-1 and -2) and inflammation (e.g., CRP, S100 proteins) were identified, the latter correlated with the stages of OVC in both plasma and tumor tissues. OVOS2, a newly discovered protein in our laboratory [11], was detected at elevated levels in the plasma, ovaries, and oviduct of all Late-Stage OVC birds in this study. Because OVOS2 is secreted, found in plasma, and has biological relevance to cancer with a potential link to extracellular tissue remodeling, it is a prime candidate for translational studies in

human OVC. Human OVOS2 expression levels at both mRNA and protein levels were significantly up-regulated in human OVC cell lines compared to normal ovarian epithelial cells. We analyzed OVOS2 protein expression in human OVC tissues by immunohistochemical staining. Malignant epithelial cells in all 5 OVC tissues were highly positive for OVOS2 while surrounding non-cancerous tissues were negative or only faintly stained. Collectively, these results provide convincing evidence that the up-regulation of OVOS2 is conserved in both chicken and human OVC with the potential to serve as a clinical biomarker for the detection and management of human OVC.

3.5 References

(1) Bast, R. C.; Hennessy, B.; Mills, G. B. The biology of ovarian cancer: new opportunities for translation. *Nat. Rev. Cancer* **2009**, *9*(6), 415-428.

(2) Cancer Facts & Figures 2012; American Cancer Society: Atlanta, 2012.

(3) Williams, T. I.; Toups, K. L.; Saggese, D. A.; Kalli, K. R.; Cliby, W. A.; Muddiman, D. C. Epithelial ovarian cancer: disease etiology, treatment, detection, and investigational gene, metabolite, and protein biomarkers. *J Proteome Res* **2007**, *6*(8), 2936-2962.

(4) Cho, K. R.; Shih Ie, M. Ovarian cancer. *Annu. Rev. Pathol.* **2009**, *4*287-313.

(5) Vanderhyden, B. C.; Shaw, T. J.; Ethier, J. F. Animal models of ovarian cancer. *Reprod. Biol. Endocrinol.* **2003**, *1*67.

(6) Garson, K.; Shaw, T. J.; Clark, K. V.; Yao, D. S.; Vanderhyden, B. C. Models of ovarian cancer - are we there yet? *Mol Cell Endocrinol* **2005**, *239*(1-2), 15-26.

(7) Johnson, P. A.; Giles, J. R. The hen as a model of ovarian cancer. *Nat. Rev. Cancer* **2013**, *13*(6), 432-436.

(8) Wilson, J. E. Adeno-carcinomata in hens kept in a constant environment. *Poult Sci* **1958**, *37*1253.

(9) Fredrickson, T. N. Ovarian Tumors of the Hen. *Environ. Health Perspect.* **1987**, *73*35-51.

(10) Barua, A.; Bitterman, P.; Abramowicz, J. S.; Dirks, A. L.; Bahr, J. M.; Hales, D. B.; Bradaric, M. J.; Edassery, S. L.; Rotmensch, J.; Luborsky, J. L. Histopathology of ovarian tumors in laying hens: a preclinical model of human ovarian cancer. *Int. J. Gynecol. Cancer* **2009**, *19*(4), 531-539.

(11) Hawkridge, A. M.; Wysocky, R. B.; Petitte, J. N.; Anderson, K. E.; Mozdziak, P. E.; Fletcher, O. J.; Horowitz, J. M.; Muddiman, D. C. Measuring the intra-individual variability of the plasma proteome in the chicken model of spontaneous ovarian adenocarcinoma. *Anal Bioanal Chem* **2010**, *398*(2), 737-49.

(12) Hakim, A. A.; Barry, C. P.; Barnes, H. J.; Anderson, K. E.; Petitte, J.; Whitaker, R.; Lancaster, J. M.; Wenham, R. M.; Carver, D. K.; Turbov, J.;

Berchuck, A.; Kopelovich, L.; Rodriguez, G. C. Ovarian adenocarcinomas in the laying hen and women share similar alterations in p53, ras, and HER-2/neu. *Cancer Prev. Res.* **2009**, *2*(2), 114-121.

(13) Stammer, K.; Edassery, S. L.; Barua, A.; Bitterman, P.; Bahr, J. M.; Hales, D. B.; Luborsky, J. L. Selenium-binding protein 1 expression in ovaries and ovarian tumors in the laying hen, a spontaneous model of human ovarian cancer. *Gynecol. Oncol.* **2008**, *109*(1), 115-121.

(14) Jackson, E.; Anderson, K.; Ashwell, C.; Petitte, J.; Mozdziak, P. E. CA125 expression in spontaneous ovarian adenocarcinomas from laying hens. *Gynecol. Oncol.* **2007**, *104*(1), 192-198.

(15) Urick, M. E.; Johnson, P. A. Cyclooxygenase 1 and 2 mRNA and protein expression in the Gallus domesticus model of ovarian cancer. *Gynecol. Oncol.* **2006**, *103*(2), 673-678.

(16) Johnson, P. A.; Giles, J. R. Use of genetic strains of chickens in studies of ovarian cancer. *Poult Sci* **2006**, *85*(2), 246-250.

(17) Giles, J. R.; Olson, L. M.; Johnson, P. A. Characterization of ovarian surface epithelial cells from the hen: a unique model for ovarian cancer. *Exp. Biol. Med.* **2006**, *231*(11), 1718-1725.

(18) Giles, J. R.; Shivaprasad, H. L.; Johnson, P. A. Ovarian tumor expression of an oviductal protein in the hen: a model for human serous ovarian adenocarcinoma. *Gynecol. Oncol.* **2004**, *95*(3), 530-533.

(19) Rodriguez-Burford, C.; Barnes, M. N.; Berry, W.; Partridge, E. E.; Grizzle, W. E. Immunohistochemical expression of molecular markers in an avian model: a potential model for preclinical evaluation of agents for ovarian cancer chemoprevention. *Gynecol. Oncol.* **2001**, *81*(3), 373-379.

(20) Dixon, R. B.; Bereman, M. S.; Petitte, J. N.; Hawkridge, A. M.; Muddiman, D. C. One-Year Plasma N-linked Glycome Intra-individual and Inter-individual Variability in the Chicken Model of Spontaneous Ovarian Adenocarcinoma. *Int. J. Mass Spectrom. Ion Proc.* **2011**, *305*(2-3), 79-86.

(21) Ansenberger, K.; Zhuge, Y.; Lagman, J. A.; Richards, C.; Barua, A.; Bahr, J. M.; Hales, D. B. E-cadherin expression in ovarian cancer in the laying hen, Gallus domesticus, compared to human ovarian cancer. *Gynecologic oncology* **2009**, *113*(3), 362-369.

(22) Gonzalez Bosquet, J.; Peedicayil, A.; Maguire, J.; Chien, J.; Rodriguez, G. C.; Whitaker, R.; Petitte, J. N.; Anderson, K. E.; Barnes, H. J.; Shridhar, V.; Cliby, W. A. Comparison of gene expression patterns between avian and human ovarian cancers. *Gynecologic oncology* **2011**, *120*(2), 256-264.

(23) Fathalla, M. F. Incessant ovulation--a factor in ovarian neoplasia? *Lancet* **1971**, *2*(7716), 163.

(24) Hawkridge, A. M.; Muddiman, D. C. Mass Spectrometry–Based Biomarker Discovery: Toward a Global Proteome Index of Individuality. *Ann. Rev. Anal. Chem.* **2009**, *2*(1), 265-277.

(25) Skates, S.; Troiano, R.; Knapp, R. C. Longitudinal CA125 detection of sporadic papillary serous carcinoma of the peritoneum. *Int. J. Gynecol. Cancer* **2003**, *13*(5), 693-696.

(26) Skates, S. J.; Menon, U.; MacDonald, N.; Rosenthal, A. N.; Oram, D. H.; Knapp, R. C.; Jacobs, I. J. Calculation of the risk of ovarian cancer from serial CA-125 values for preclinical detection in postmenopausal women. *J Clin Oncol* **2003**, *21*(10 Suppl), 206s-210s.

(27) Tuxen, M. K.; Soletormos, G.; Petersen, P. H.; Schioler, V.; Dombernowsky, P. Assessment of biological variation and analytical imprecision of CA 125, CEA, and TPA in relation to monitoring of ovarian cancer. *Gynecologic oncology* **1999**, *74*(1), 12-22.

(28) Tuxen, M. K.; Soletormos, G.; Rustin, G. J.; Nelstrop, A. E.; Dombernowsky, P. Biological variation and analytical imprecision of CA 125 in patients with ovarian cancer. *Scand. J. Clin. Lab. Invest.* **2000**, *60*(8), 713-721.

(29) Skates, S. J.; Xu, F. J.; Yu, Y. H.; Sjovall, K.; Einhorn, N.; Chang, Y.; Bast, R. C., Jr.; Knapp, R. C. Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers. *Cancer* **1995**, *76*(10), 2004-2010.

(30) Nielsen, K. L.; Sottrup-Jensen, L. Evidence from sequence analysis that hen egg-white ovomacroglobulin (ovostatin) is devoid of an internal beta-Cys-gamma-Glu thiol ester. *Biochim Biophys Acta* **1993**, *1162*(1-2), 230-232.

(31) Nielsen, K. L.; Sottrup-Jensen, L.; Nagase, H.; Thogersen, H. C.; Etzerodt, M. Amino acid sequence of hen ovomacroglobulin (ovostatin) deduced from cloned cDNA. *DNA Seq.* **1994**, *5*(2), 111-119.

(32) Hideaki Nagase, E. D. H. J. Ovostatin: A novel Proteinase Inhibitor from Chicken Egg White. *J Biol Chem* **1983**, *258*(12), 7490-7498.

(33) Nagase, H.; Harris, E. D.; Woessner, J. F.; Brew, K. Ovostatin - A novel proteinase-inhibitor from chicken egg-white 1. purification, physicochemical properties, and tissue distribution of ovostatin. *J. Biol. Chem.* **1983**, *258*(12), 7481-7489.

(34) Miller, H. T., Feeney, R.E. The Physical and Chemical Properties of an Immunologically Cross-Reactive Protein from Avian Egg Whites. *Biochemistry* **1966**, *5*(3), 952-958.

(35) Mann, K. The chicken egg white proteome. *Proteomics* **2007**, *7*(19), 3558-68.

(36) Mann, K.; Mann, M. The chicken egg yolk plasma and granule proteomes. *Proteomics* **2008**, *8*(1), 178-91.

(37) Lim, W.; Jeong, W.; Kim, J. H.; Lee, J. Y.; Kim, J.; Bazer, F. W.; Han, J. Y.; Song, G. Differential expression of alpha 2 macroglobulin in response to dietylstilbestrol and in ovarian carcinomas in chickens. *Reproductive Biology and Endocrinology* **2011**, *9*137-146.

(38) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins from Silver-Staining Polyacrylamide Gels. *Anal Chem* **1996**, *68*850-858.

(39) Andrews, G. L.; Shuford, C. M.; Burnett, J. C., Jr.; Hawkridge, A. M.; Muddiman, D. C. Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2009**, *877*(10), 948-954.

(40) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Analytical Chemistry* **2002**, *74*5383-5392.

(41) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Analytical Chemistry* **2003**, *75*4646-4658.

(42) Weatherly, D. B.; Atwood, J. A., 3rd; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R. A Heuristic method for assigning a false-discovery rate for

protein identifications from Mascot database search results. *Molecular & Cellular Proteomics* **2005**, *4*(6), 762-772.

(43) Yu, Y.; Xu, F.; Peng, H.; Fang, X.; Zhao, S.; Li, Y.; Cuevas, B.; Kuo, W. L.; Gray, J. W.; Siciliano, M.; Mills, G. B.; Bast, R. C., Jr. NOEY2 (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas. *Proc Natl Acad Sci U S A* **1999**, *96*(1), 214-9.

(44) Fang, X.; Yu, S.; Bast, R. C.; Liu, S.; Xu, H. J.; Hu, S. X.; LaPushin, R.; Claret, F. X.; Aggarwal, B. B.; Lu, Y.; Mills, G. B. Mechanisms for lysophosphatidic acid-induced cytokine production in ovarian cancer cells. *The Journal of Biological Chemistry* **2004**, *279*(10), 9653-9661.

(45) Wu, J.; Mukherjee, A.; Lebman, D. A.; Fang, X. Gene expression of the lysophosphatidic acid receptor 1 is a target of transforming growth factor beta. *Oncogene* **2012**, *Advanced on-line publication*.

(46) Teves, M. E.; Zhang, Z.; Costanzo, R. M.; Henderson, S. C.; Corwin, F. D.; Zweit, J.; Sundaresan, G.; Subler, M.; Salloum, F. N.; Rubin, B. K.; Strauss, J. F., 3rd. Sperm-associated antigen-17 gene is essential for motile cilia function and neonatal survival. *Am J Respir Cell Mol Biol* **2013**, *48*(6), 765-772.

(47) Liu, H. B.; Sadygov, R. G.; Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **2004**, *76*(14), 4193-4201.

(48) Wiener, M. C.; Sachs, J. R.; Deyanova, E. G.; Yates, N. A. Differential mass spectrometry: A label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal Chem* **2004**, *76*(20), 6085-6096.

(49) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **2005**, *4*(9), 1265-72.

(50) Schaller, J.; Gerber, S. A.; Kampfer, U.; Lejon, S.; Trachsel, C. *Human Blood Plasma Proteins: Structure and Function*; John Wiley & Sons: West Sussex, England, 2008, p 526.

(51) Karst, A. M.; Levanon, K.; Drapkin, R. Modeling high-grade serous ovarian carcinogenesis from the fallopian tube. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*(18), 7547-7552.

(52) Kim, J.; Coffey, D. M.; Creighton, C. J.; Yu, Z.; Hawkins, S. M.; Matzuk, M. M. High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*(10), 3921-3926.

(53) Dubeau, L. The cell of origin of ovarian epithelial tumors and the ovarian surface epithelium dogma: does the emperor have no clothes? *Gynecologic oncology* **1999**, *72*(3), 437-442.

(54) Dubeau, L. The cell of origin of ovarian epithelial tumours. *Lancet Oncol* **2008**, *9*(12), 1191-1197.

(55) Auersperg, N.; Wong, A. S.; Choi, K. C.; Kang, S. K.; Leung, P. C. Ovarian surface epithelium: biology, endocrinology, and pathology. *Endocr. Rev.* **2001**, *22*(2), 255-288.

(56) Andrews Kingon, G. L.; Petitte, J. N.; Muddiman, D. C.; Hawkridge, A. M. Multi-peptide nLC-PC-IDMS-SRM-based Assay for the Quantification of Biomarkers in the Chicken Ovarian Cancer Model. *Methods* **2013**in press.

(57) Robinson, D. S.; Monsey, J. B. Studies on the composition of egg-white ovomucin. *Biochem J* **1971**, *121*(3), 537-47.

(58) Reid, C. J.; Harris, A. Developmental expression of mucin genes in the human gastrointestinal system. *Gut* **1998**, *42*(2), 220-226.

(59) Berger, J. T.; Voynow, J. A.; Peters, K. W.; Rose, M. C. Respiratory carcinoma cell lines. MUC genes and glycoconjugates. *Am J Respir Cell Mol Biol* **1999**, *20*(3), 500-510.

(60) Sylvester, P. A.; Myerscough, N.; Warren, B. F.; Carlstedt, I.; Corfield, A. P.; Durdey, P.; Thomas, M. G. Differential expression of the chromosome 11 mucin genes in colorectal cancer. *J. Pathol.* **2001**, *195*(3), 327-335.

(61) Tsukashita, S.; Kushima, R.; Bamba, M.; Sugihara, H.; Hattori, T. MUC gene expression and histogenesis of adenocarcinoma of the stomach. *Int. J. Cancer* **2001**, *94*(2), 166-70.

(62) Stewart, C. J.; Tsukamoto, T.; Cooke, B.; Leung, Y. C.; Hammond, I. G. Ovarian mucinous tumour arising in mature cystic teratoma and associated with pseudomyxoma peritonei: report of two cases and comparison with ovarian involvement by low-grade appendiceal mucinous tumour. *Pathology* **2006**, *38*(6), 534-538.

(63) Fazzari, C.; Fedele, F.; Pizzi, G.; Crisafulli, C.; Parisi, A.; Caruso, R. A. Krukenberg tumour of the ovary: a case report with light microscopy, immunohistochemistry and electron microscopy study. *Anticancer Res.* **2008**, *28*(2B), 1417-1420.

(64) Dube, D. H.; Bertozzi, C. R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **2005**, *4*(6), 477-488.

(65) Mantovani, A.; Allavena, P.; Sica, A.; Balkwill, F. Cancer-related inflammation. *Nature* **2008**, *454*(7203), 436-444.

(66) Coussens, L. M.; Werb, Z. Inflammation and cancer. *Nature* **2002**, *420*(6917), 860-867.

(67) Black, S.; Kushner, I.; Samols, D. C-reactive Protein. *J Biol Chem* **2004**, *279*(47), 48487-48490.

(68) Ogata, Y.; Heppelmann, C. J.; Charlesworth, M. C.; Madden, B. J.; Miller, M. N.; Kalli, K. R.; Cilby, W. A.; Bergen, H. R., 3rd; Saggese, D. A.; Muddiman, D. C. Elevated levels of phosphorylated fibrinogen-alpha-isoforms and differential expression of other post-translationally modified proteins in the plasma of ovarian cancer patients. *J Proteome Res* **2006**, *5*(12), 3318-25.

(69) Williams, D. K.; Muddiman, D. C. Absolute quantification of C-reactive protein in human plasma derived from patients with epithelial ovarian cancer utilizing protein cleavage isotope dilution mass spectrometry. *J Proteome Res* **2009**, *8*(2), 1085-1090.

(70) Wei, B. R.; Hoover, S. B.; Ross, M. M.; Zhou, W.; Meani, F.; Edwards, J. B.; Spehalski, E. I.; Risinger, J. I.; Alvord, W. G.; Quinones, O. A.; Belluco, C.; Martella, L.; Campagnutta, E.; Ravaggi, A.; Dai, R. M.; Goldsmith, P. K.; Woolard, K. D.; Pecorelli, S.; Liotta, L. A.; Petricoin, E. F.; Simpson, R. M. Serum S100A6 concentration predicts peritoneal tumor burden in mice with epithelial ovarian cancer and is associated with advanced stage in patients. *PLoS One* **2009**, *4*(10), e7670.

(71) Rehman, A. A.; Ahsan, H.; Khan, F. H. Alpha-2-Macroglobulin: a physiological guardian. *J Cell Physiol* **2013**, *228*(8), 1665-1675.

(72) Scherer, S. E.; Muzny, D. M.; Buhay, C. J.; Chen, R.; Cree, A.; Ding, Y.; Dugan-Rocha, S.; Gill, R.; Gunaratne, P.; Harris, R. A.; Hawes, A. C.; Hernandez, J.; Hodgson, A. V.; Hume, J.; Jackson, A.; Khan, Z. M.; Kovar-Smith, C.; Lewis, L. R.; Lozado, R. J.; Metzker, M. L.; Milosavljevic, A.; Miner,

G. R.; Montgomery, K. T.; Morgan, M. B.; Nazareth, L. V.; Scott, G.; Sodergren, E.; Song, X. Z.; Steffen, D.; Lovering, R. C.; Wheeler, D. A.; Worley, K. C.; Yuan, Y.; Zhang, Z.; Adams, C. Q.; Ansari-Lari, M. A.; Ayele, M.; Brown, M. J.; Chen, G.; Chen, Z.; Clerc-Blankenburg, K. P.; Davis, C.; Delgado, O.; Dinh, H. H.; Draper, H.; Gonzalez-Garay, M. L.; Havlak, P.; Jackson, L. R.; Jacob, L. S.; Kelly, S. H.; Li, L.; Li, Z.; Liu, J.; Liu, W.; Lu, J.; Maheshwari, M.; Nguyen, B. V.; Okwuonu, G. O.; Pasternak, S.; Perez, L. M.; Plopper, F. J.; Santibanez, J.; Shen, H.; Tabor, P. E.; Verduzco, D.; Waldron, L.; Wang, Q.; Williams, G. A.; Zhang, J.; Zhou, J.; Allen, C. C.; Amin, A. G.; Anyalebechi, V.; Bailey, M.; Barbaria, J. A.; Bimage, K. E.; Bryant, N. P.; Burch, P. E.; Burkett, C. E.; Burrell, K. L.; Calderon, E.; Cardenas, V.; Carter, K.; Casias, K.; Cavazos, I.; Cavazos, S. R.; Ceasar, H.; Chacko, J.; Chan, S. N.; Chavez, D.; Christopoulos, C.; Chu, J.; Cockrell, R.; Cox, C. D.; Dang, M.; Dathorne, S. R.; David, R.; Davis, C. M.; Davy-Carroll, L.; Deshazo, D. R.; Donlin, J. E.; D'Souza, L.; Eaves, K. A.; Egan, A.; Emery-Cohen, A. J.; Escotto, M.; Flagg, N.; Forbes, L. D.; Gabisi, A. M.; Garza, M.; Hamilton, C.; Henderson, N.; Hernandez, O.; Hines, S.; Hogues, M. E.; Huang, M.; Idlebird, D. G.; Johnson, R.; Jolivet, A.; Jones, S.; Kagan, R.; King, L. M.; Leal, B.; Lebow, H.; Lee, S.; LeVan, J. M.; Lewis, L. C.; London, P.; Lorensuhewa, L. M.; Loulseged, H.; Lovett, D. A.; Lucier, A.; Lucier, R. L.; Ma, J.; Madu, R. C.; Mapua, P.; Martindale, A. D.; Martinez, E.; Massey, E.; Mawhiney, S.; Meador, M. G.; Mendez, S.; Mercado, C.; Mercado, I. C.; Merritt, C. E.; Miner, Z. L.; Minja, E.; Mitchell, T.; Mohabbat, F.; Mohabbat, K.; Montgomery, B.; Moore, N.; Morris, S.; Munidasa, M.; Ngo, R. N.; Nguyen, N. B.; Nickerson, E.; Nwaokelemeh, O. O.; Nwokenkwo, S.; Obregon, M.; Oguh, M.; Oragunye, N.; Oviedo, R. J.; Parish, B. J.; Parker, D. N.; Parrish, J.; Parks, K. L.; Paul, H. A.; Payton, B. A.; Perez, A.; Perrin, W.; Pickens, A.; Primus, E. L.; Pu, L. L.; Puazo, M.; Quiles, M. M.; Quiroz, J. B.; Rabata, D.; Reeves, K.; Ruiz, S. J.; Shao, H.; Sisson, I.; Sonaike, T.; Sorelle, R. P.; Sutton, A. E.; Svatek, A. F.; Svetz, L. A.; Tamerisa, K. S.; Taylor, T. R.; Teague, B.; Thomas, N.; Thorn, R. D.; Trejos, Z. Y.; Trevino, B. K.; Ukegbu, O. N.; Urban, J. B.; Vasquez, L. I.; Vera, V. A.; Villasana, D. M.; Wang, L.; Ward-Moore, S.; Warren, J. T.; Wei, X.; White, F.; Williamson, A. L.; Wleczyk, R.; Wooden, H. S.; Wooden, S. H.; Yen, J.; Yoon, L.; Yoon, V.; Zorrilla, S. E.; Nelson, D.; Kucherlapati, R.; Weinstock, G.; Gibbs, R. A.; Baylor College of Medicine Human Genome Sequencing Center Sequence Production, T. The finished DNA sequence of human chromosome 12. *Nature* **2006**, *440*(7082), 346-351.

(73) Normandin, K.; Peant, B.; Le Page, C.; de Ladurantaye, M.; Ouellet, V.; Tonin, P. N.; Provencher, D. M.; Mes-Masson, A. M. Protease inhibitor SERPINA1 expression in epithelial ovarian cancer. *Clin Exp Metastasis* **2010**, *27*(1), 55-69.

(74) Chambers, S. K.; Gertz, R. E., Jr.; Ivins, C. M.; Kacinski, B. M. The significance of urokinase- type plasminogen activator, its inhibitors, and its receptor in ascites of patients with epithelial ovarian cancer. *Cancer* **1995**, *75*(7), 1627-1633.

(75) Chambers, S. K.; Ivins, C. M.; Carcangiu, M. L. Expression of plasminogen activator inhibitor-2 in epithelial ovarian cancer: a favorable prognostic factor related to the actions of CSF-1. *Int. J. Cancer* **1997**, *74*(6), 571-575.

(76) Sieben, N. L.; Oosting, J.; Flanagan, A. M.; Prat, J.; Roemen, G. M.; Kolkman-Uljee, S. M.; van Eijk, R.; Cornelisse, C. J.; Fleuren, G. J.; van Engeland, M. Differential gene expression in ovarian tumors reveals Dusp 4 and Serpina 5 as key regulators for benign behavior of serous borderline tumors. *J Clin Oncol* **2005**, *23*(29), 7257-7264.

(77) Bauerschlag, D. O.; Habermann, M.; Weimer, J.; Meinhold-Heerlein, I.; Hilpert, F.; Weigel, M.; Bauer, M.; Mundhenke, C.; Jonat, W.; Maass, N.; Schem, C. Heterogeneous expression of serine protease inhibitor maspin in ovarian cancer. *Anticancer Res.* **2010**, *30*(7), 2739-2744.

(78) Okamoto, T.; Niu, R.; Yamada, S. Increased expression of tissue inhibitor of metalloproteinase-2 in clear cell carcinoma of the ovary. *Mol Hum Reprod* **2003**, *9*(10), 569-575.

(79) Ripley, D.; Tunuguntla, R.; Susi, L.; Chegini, N. Expression of matrix metalloproteinase-26 and tissue inhibitors of metalloproteinase-3 and -4 in normal ovary and ovarian carcinoma. *Int. J. Gynecol. Cancer* **2006**, *16*(5), 1794-1800.

(80) Devoogdt, N.; Rasool, N.; Hoskins, E.; Simpkins, F.; Tchabo, N.; Kohn, E. C. Overexpression of protease inhibitor-dead secretory leukocyte protease inhibitor causes more aggressive ovarian cancer in vitro and in vivo. *Cancer Sci* **2009**, *100*(3), 434-440.

(81) Rasool, N.; LaRochelle, W.; Zhong, H.; Ara, G.; Cohen, J.; Kohn, E. C. Secretory leukocyte protease inhibitor antagonizes paclitaxel in ovarian cancer cells. *Clin Cancer Res* **2010**, *16*(2), 600-609.

CHAPTER 4

Global Proteomic Analysis of Functional Compartments in Immature Avian Follicles Using Laser Microdissection Coupled to LC-MS/MS

The following work was reprinted from the recently submitted manuscript:
Nepomuceno, A.I.; Muddiman, D.D.; Petitte, J.N. *Journal of Biological Reproduction*

4.1 Introduction

The avian ovary is unique among vertebrates in that it contains hundreds of follicles but has about 5 large yolky follicles arranged in a hierarchy based upon size from about 9mm to 40mm in diameter. The largest follicle is designated as the F1 follicle which is the most mature and is destined for the next ovulation. A reproductively active domestic hen ovulates roughly every day, and an active ovary generally contains a succession of preovulatory follicles (F2-F5) that mature to the next size to take the place of the newly released F1 follicle. A new F5 follicle is recruited from a pool of 7-9 small yellow follicles (SYF) that range 5 - 9 mm in diameter. Below SYF lies a pool of large white follicles (LWF) 2-5 mm in size that contain no apparent yellow yolk. The smallest and most immature of the avian ovarian follicles are the numerous small white follicles (SWF) that are less than 2 mm in diameter. Maturation of avian ovarian follicles has largely focused upon the process of recruitment from the SYF to the preovulatory follicles and the maturation of the F1. Much of this has centered on changes in steroidogenesis associated with the transition from SYF and maturation of the F1 follicle. Granulosa cells and theca cells of the follicular wall are the main sources for the production of steroid hormones.[1-3] Two hypotheses exist to explain the interplay between the two cell

types for the production of steroids. The first suggests that the granulosa cells of preovulatory follicles produce progestins that are, in turn, used as a substrate by the theca cells for the production of androgen and estrogen.[1] The other proposes that theca cells can synthesize progestins, androgen, and estrogen independently of granulosa cells.[4] A key component to the steroidogenic pathway is the synthesis of the steroidogenic acute regulatory protein (StAR). Recently, it has been shown that prehierarchal follicles (6-8 mm) fail to express mRNA of the StAR protein.[5,6] However, it is not well understood how these steroidogenic pathways function, if at all, in the small white follicles.

Very little research has been done to understand the biology of the smallest follicles of the avian ovary and most work has consisted in removing the follicles from the stroma of the ovary for the culture of whole follicles. We wished to gain access to the smallest follicles; and, if possible, apply global proteomics to the various compartments. Laser microdissection LMD) was used to isolate the following specific compartments: follicular wall (granulosa/theca cells), white yolk, and ovarian stromal tissue. A global proteome profile of each concentrated compartment was analyzed using LC-MS/MS. In the process we were able to identify 2,889 total proteins among the three compartments. This is the first report of the identification of 1,984 proteins identified specifically from the white yolk of SWF. Previous reports have demonstrated LC-MS/MS techniques used to identify 119 yellow egg yolk proteins.[7] This mass spectrometric proteomics based approach has proved to be complementary to that of targeted mRNA studies in follicular development. This

extensive global proteomic data can be extended towards studying other biological systems.

4.2 Experimental

4.2.1 Animal Care And Biospecimen Collection

Chicken ovary was obtained from a 22 month old healthy Bovans White commercial layer chicken. The hen was obtained at 4 months of age from the North Carolina Department of Agriculture and Consumer Services, Piedmont Research facility in Salisbury, NC. The hen was transported to the North Carolina State University Lake Wheeler Poultry Science Facility in Raleigh, NC. The hen was managed in accordance with the Institute for Laboratory Animal Research Guide with all the husbandry practices being approved by and under the oversight of North Carolina State University Institutional Animal Care and Use Committee. Birds were housed in individual cages under 16L:8D photoperiod and fed an 18% protein layer diet. The health and egg-laying status of the birds were monitored daily and only reproductively active hens were chosen for the study.

The chicken ovary was then removed and cleaned with phosphate buffered saline, pH 8.0. The ovary was cut and portioned into cryomold intermediate holders (Miles Inc., Elkhart, IN) frozen in optimal cutting temperature (OCT) compound (Leica Microsystems, Buffalo, IL). Each mold was placed into isopentane (Sigma Aldrich, St. Louis, MO) bath surrounded by liquid nitrogen. Serial slices of 10 µm thick pieces were cut using a CM1950 (Leica Microsystems, Buffalo, IL). Sections

used for laser microdissection were placed onto PEN membrane slides (Leica Microsystems, Buffalo, IL). Sections used for Hemotoxylin and Eosin staining were placed onto microscope slides (VWR, Radnor, PA).

4.2.2 Laser MIcrodissection

Each PEN membrane slide was prepared individually for laser microdissection. The protocol is as follows. A starting rinse was accomplished using 100% ethanol (Anhydrous KOPTEC USP, VWR, Randor PA) for 2 minutes. Followed with 1 minute rinses in 95% and 75% ethanol. For 30 seconds the slide was placed in nuclease free distilled water (Molecular Devices, Sunnyvale, CA). 100 µL of Paradise®^{Plus} staining solution (Molecular Devices) was used to cover the entire section for approximately 30 seconds. The excess stain was gently removed from the slide prior to washing with 75% then 95% ethanol for 30 seconds. An additional wash was performed in 100% ethanol for 1 minute. The slide was then placed into xylene (Sigma Aldrich, St. Louis, MO) for approximate 5 minutes for complete dehydration of the tissue section.

Nine follicles were targeted; the white yolk, follicular wall cells, and stromal tissue portions were excised utilizing a laser microdissection (LMD 7000, Leica Microsystems). Diameters of each follicle were extrapolated using an algorithm processed in MATLAB (MathWorks, Natick, MA). Calculated follicle sizes ranged from 270 µm to 1600 µm. A fourth sample consisting of random regions of the ovary were also sampled. Proteins samples were extracted using lysis buffer consisting of

30 µL of 100 mM Tris HCl pH 7.5, 100 mM DTT, and 4% SDS. The solution was then heated at 95 ^{0}C for 5 minutes. For separation of the intact PEN membrane from the solution, the samples were then centrifuged for 20 min at 14,000 g.

4.2.3 Filter Aided Sample Preparation (FASP)

The samples were digested using a modified filter aided sample preparation (FASP) method[8,9] with minor changes. Briefly, the disulfide bonds were reduced by adding 3 µL of 50 mM dithiothreitol (DTT) to each sample and incubated for 30 min at 56 ^{0}C. The sample was then mixed with 200 µL of 8 M urea solution which was made with 100 mM Tris HCl pH 8.0. The solution was then transfered onto a Vivacon 500 30 kDa MW cutoff filter (Vivacon Products, Littleton, MA). To prevent carbamylation caused by the presence of urea a refrigerated centrifuge (Eppendorf, Hauppauge, NY) was use and the samples were centrifuged for 15 mins at 21 ^{0}C. Another wash of 200 µL of 100 mM Tris HCl pH 8.0 was repeated and the flow through was then discarded. The alkylation step was accomplished when 100 µL of 50 mM iodoacetamide was added and incubated for 20 min in the dark at room temperature. Wash steps then ensued with 100 µL 8 M urea and 100 µL of 100 mM Tris HCl pH 7.0 with each wash repeated 3 times. At each step the filters were centrifuged for 10 mins at 14,000 g. Trypsin digestion occurred when a total of 0.25 µg of trypsin in 40 µL of 100 mM Tris HCl pH 7.0 was added onto the filter. Allowing for 16 hours of digestion at 37 ^{0}C, the peptides were eluted through the filter by centrifugation for 10 min at 14,000 g. An additional 40 µL of 100 mM Tris HCl pH 7.0

was added and then spun down for 10 min at 14,000 g. The sample was then reduced to dryness.

4.2.4 Nano-LC-MS/MS

A thermo scientific Easy nLC II (Thermo Scientific, San Jose, CA) was coupled to a quadrupole Orbitrap benchtop mass spectrometer (Q-Exactive) using a vented column configuration.[10] The column and trap were packed in-house with Magic C18AQ stationary phase (5 µm particle, 200 Å pore, Auburn, CA). Each sample was reconstituted in 25 µL mobile phase A which was composed of water/acetonitrile/formic acid (98/2/0.2%). 5 µL was then loaded onto the 75 µm × 5 cm trap (IntegraFrit, New Objectives, Woburn, MA) at a flow rate of 3.5 µL/min using 100% mobile phase A. Mobile phase B comprised of water/acetonitrile/formic acid (2/98/0.2%). Flow was diverted onto the column the which began the start of the gradient: 2% B (0-5 min), 2%-5% B (5-7 min), 5%-40% B (7-208 min), 40%-95% (208-218 min), 95% (218-228 min), 95%-2% (228-230 min), 2% (230-240 min).

Global proteomics parameters were utilized on a Q-Exactive platform.[11] Full scan MS transients were acquired in the orbitrap with a 70 k_{FWHM} resolving power at *m/z* = 200. The automatic gain control (AGC) target for MS acquisition was set to 1E6 with a maximum ion injection of 30 ms. All full scan spectra were acquired with a *m/z* range of 400 to 1,600. MS/MS acquisition was acquired in data dependent acquisition mode where the maximum MS/MS events were set to 12 and the dynamic exclusion was set to 30 sec. The resolving power for all MS/MS spectra

was set to 17.5 k_{fwhm} at *m/z* = 200. With a maximum ion injection time of 250 ms for an AGC target of 2E5.

4.2.5 Data Analysis

Proteome Discoverer was utilized for converting all RAW LC-MS/MS chromatogram files into an MGF format. The resulting MGF format was then searched using MASCOT (Matrix Science, Boston, MA)[12] against a concatenated target-reverse *Gallus gallus* reference database (NCBI, Nov 2013). The data was also searched with a fixed carbamidomethyl (C) modification, while oxidation (M) and deamidation (NQ) were set to variable modifications and a maximum of 2 missed cleavages. The precursor ion search tolerance was set to 5 ppm and the fragment ion tolerance was set to ± 0.02 Da. Statistical filtering using a 1% false discovery rate for identifying protein was performed using ProteoIQ.[13-15]

Quantification of proteins was accomplished by calculating SpCs and NSAF values. A SpCs is calculated for each protein and that number coincides with the number of MS/MS spectra that result in the identification of its tryptic fragments. Proteins were then compared between samples by calculating the NSAF values. NSAF values are calculated by taking the SpC for a given protein and dividing it by its length (L) to give a spectral abundance factor (SAF). The SAF for all the proteins within a run are summed to generate a Normalized SAF (NSAF) that is used to compare the relative abundance of proteins between samples.

4.3 Results

4.3.1 Protein Identification

LMD was utilized for the collection of tissue compartments obtained from a reproductively active hen. **Figure 1** illustrates an H&E stained ovary tissue outlining the three types of tissue compartments collected separately. First, white yolk components from follicles less than 1.6 mm with no presence of nuclei were accumulated. Once the white yolk samples were removed and collected, their complementing follicular walls were then excised. Next the adjacent stromal tissues

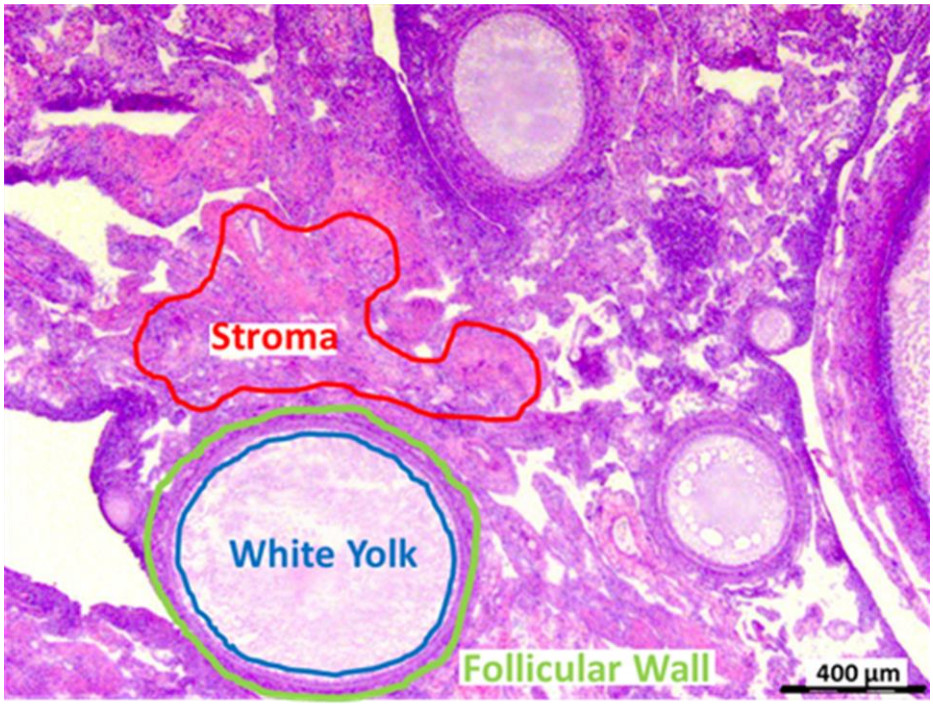


Figure 4.1 H&E stained slide of actively reproductive hen ovary. Outlined are representative areas of white yolk, follicular wall and stromal cells excised and collected for protein extraction.

from each of the follicles were collected for proteomic analysis. It is typical to consume the entire tissue when conducting a global proteomic profile of tissues. To represent an entire tissue sample, a “Pooled” sample was collected from random regions of the tissue. This included portions of the white yolk, follicular wall, and stromal tissue.

Figure 2a is a Venn diagram which displays the distribution of identified proteins from each tissue compartment. A total of 2,889 proteins were identified from the three individual sample compartments. A majority of proteins identified resulted from the ovarian stroma and the follicular wall cells, each with 2456 and 2470 proteins identified, respectively. The white yolk resulted in the identification of 1,984 proteins. This is the first comprehensive reported protein list of white yolk proteins from SWF. A total of 2903 proteins were identified in the white yolk, stromal cells, and follicular wall compartments as well as the “Pooled” sampled (**Figure 2B**). 99.5% of the total proteins were identified from summing the proteins from the three compartments. While the “Pooled” sample resulted in only 2507 proteins identified (86.4%). An increase in proteome coverage was gained through the concentration of tissue compartments from the use of LMD.

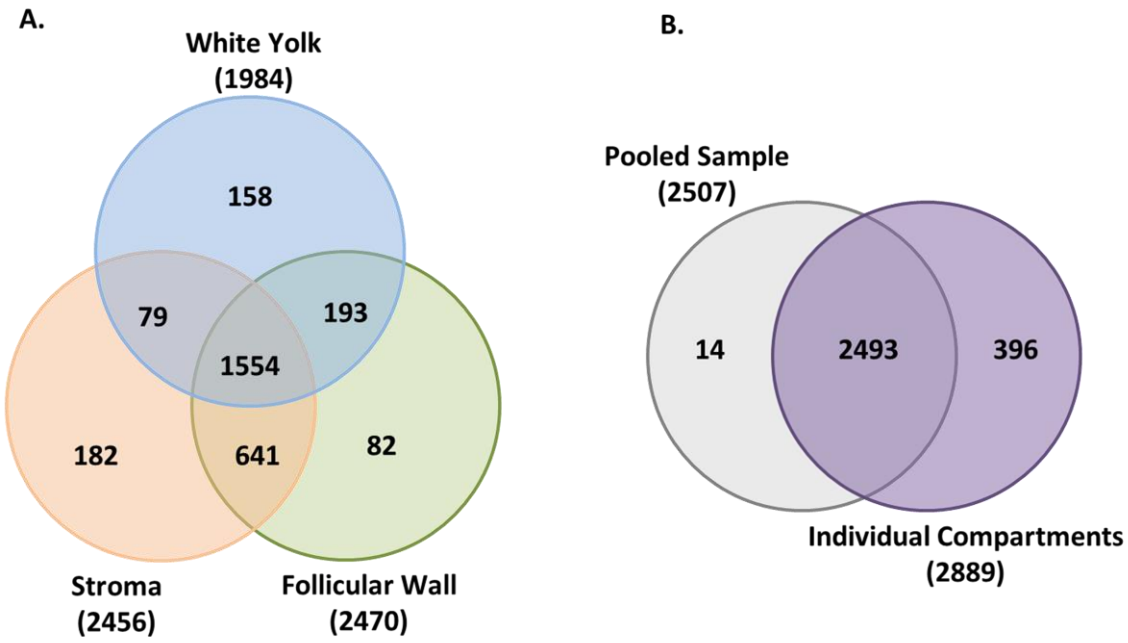


Figure 4.2 (A) The number of proteins that were in the white yolk, stromal cells and follicular wall. A total of 1984, 2456 and 2470 were identified in each cell type respectively. (B) Random areas of the ovary containing all three cell types were excised and analyzed. The "Control" sampled resulted in a total of 2507 proteins. 2493 proteins overlapped with the 2889 proteins that were identified when concentrating individual cell types.

An added advantage of discovery-based proteomics methods is the ability to search data for post-translational modifications. All Proteins were searched with phosphorylation as a variable post-translation modification. A total of 91 proteins contained phosphorylated peptides. A total of 4 peptides were identified as singly or doubly phosphorylated in both Vitellogenin I and II. Interestingly, Epidermal Growth Factor Receptor (EGFR) protein had one peptide [SLFATQSQKTKIIQNR] identified with multiple phosphorylation sites. A total of four phosphorylation sites on EFGR

were recognized in both the stromal tissue and follicular cells. For the stromal tissue, the Serine 1 and 7 positions on the EGFR peptide were phosphorylated as well as the Threonine 10. However phosphorylation sites in the follicular wall samples differed by with the Threonine 5 sites as phosphorylated and not the Serine 1 site.

Global proteomic data can be daunting to analyze due to the large number of proteins identified. Statistical analysis was implemented to consolidate a list of proteins for further investigation. Proteins with greater than a two-fold difference in concentration and were statistically significant using a p-value cut-off of 0.05 were chosen for further investigation. Relative protein abundances between each compartment are displayed in **Figure 3** through the use of volcano plots. Proteins observed in the orange shaded regions of the volcano plots were calculated as greater than 2-fold change and statistically significant, and were chosen for further analysis**.** The DAVID algorithm program was used to obtain KEGG pathway analysis, cellular component and biological processes algorithms of the selected proteins.[16] **Table 1** lists the KEGG pathways identified from the proteins observed from the individual compartments. Characterization of white yolk proteins was of great interest due to the large discrepancy of proteins identified between white yolk proteins and yellow egg yolk proteins. The top two GO annotations for cellular component classification resulted in “Cytoplasm” and “Cytoplamic part” with p-value scores of 5.6E-28 and 4.0E-18 respectively. The top scoring Biological Processes for the white yolk proteins resulted as “generation of precursor metabolites and energy” with a p-value of 5.8E-8.

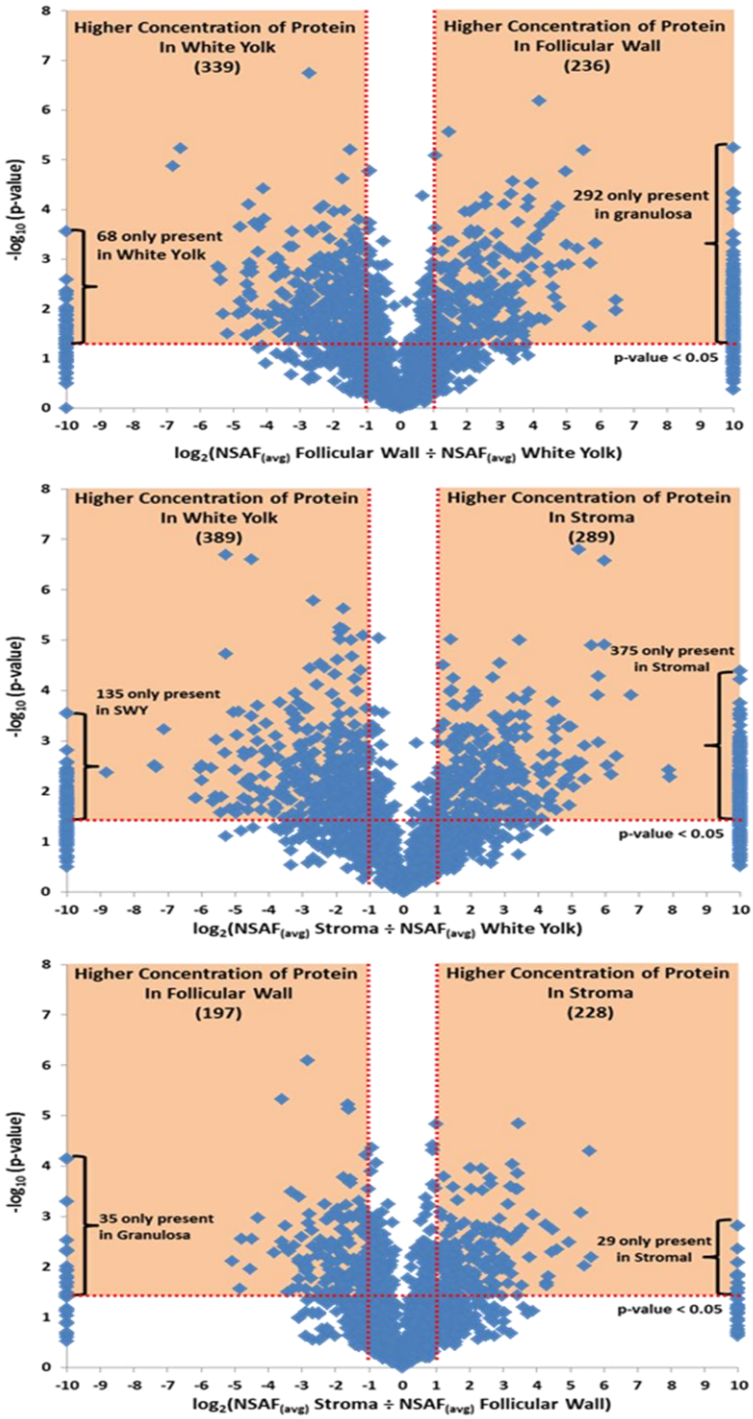


Figure 4.3 Volcano plots graphing $-\log_{10}$(p-value) versus $\log_2$ expressions. Highlighted in orange are proteins that greater than 2-fold higher concentration and statistically significant using a p-value of 0.05.

Table 4.1 Identified KEGG pathways based on Tissue Compartments

	Stroma	Follicular Wall	White Yolk
Focal Adhesion	X	X	
Ribsome	X	X	
ECM-receptor Interaction	X	X	
Spliceosome	X	X	
Tight Junction	X	X	
Adherens Junction	X	X	
Proteasome	X		
Cardiac mudscle contraction	X		
Vascular Smooth Muscle Contraction	X		
Drug Metabolism	X		
Regulation of Actin cytoskeleton	X		
Glutathione metabolism	X		
Glycolysis / glueconeogenesis		X	X
arginine and proline metabolism		X	X
citrate cycle (TCA cycle)		X	X
lysine degradation		X	
limonene and pinene degradation		X	
histidine metabolism		X	
pyruvate metabolism			X
oxidative phosphorylation			X
lysosome			X
RNA degradation			X
Endocytosis			X
Pentose phophate pathway			X

4.3.2 Targeted Proteins in Follicular Development

Many genes have been targeted in follicles with respect to their association with follicular activation and development.[5,17-19] **Table 2** presents a short list of Gene names and their corresponding Protein accession numbers that were previously examined in the development of SWF.[5,18,20,21] A few of the proteins were not detectable using LC-MS/MS techniques and can be attributed to low concentrations. Another series of proteins of great interest are the steroidogenic pathway enzymes. **Figure 4** illustrates the major steroidogenic pathways associated with follicular maturation in the domestic hen. **Table 3.** is of all the steroidogenic pathway proteins that were identified in their respective compartments. Relative quantification of each

protein present is based on their Spectral Count (Avg-SpC) and Normalized Spectal Abundance Factor (NSAF). The highest concentrations of proteins were identified in the follicular wall cells with some trace proteins detected in the stromal tissue. However, StAR protein was not detected due to the undifferentiated state of the SWF follicles.

Table 4.2 Protein Detection of Commonly used Follicular Developmental Genes

Gene Name	mRNA Present	Protein Accession Number - Protein Name	Protein Detected
CGA (FSH - Follicle Stimulating Hormone)	+	>gi|487441431|ref|NP_001264950.1| glycoprotein hormones alpha chain precursor	NA
FSHB (FSH - Follicle Stimulating Hormone)	+	>gi|259414634|gb|ACW82409.1| follicle stimulating hormone, beta polypeptide	NA
FSHR (Follicle Stimulating Hormone Receptor)	+	>gi|219663002|gb|ACL30987.1| follicle stimulating hormone receptor	NA
BMP6 (Bone Morphogenetic Protein 6)	+	>gi|513170824|ref|XP_418956.4| PREDICTED: bone morphogenetic protein 6	NA
SMAD 2 (Mothers Against Decapentaplegic Homolog 2)	+	>gi|17384013|emb|CAC85407.1| MADH2 protein	+
GATA4	+	>gi|1169844|sp|P43691.1|Transcription factor GATA-4	+
AMH (Anti-Mullerian Hormone)	+	>gi|1432158|gb|AAB04022.1| anti-Mullerian hormone	+
EGFR (Epidermal Growth Factor Receptor)	+	gi|119223|sp|P13387.1|EGFR_CHICK| Epidermal growth factor receptor	+
MAPK1 (Mitogen-Activated Protein Kinase, ERK2)	+	>gi|45383812|ref|NP_989481.1| mitogen-activated protein kinase 1	+
FOXL2	+	gi|332646982|gb|AEE80502.1|Transcription Factor FOXL2	+

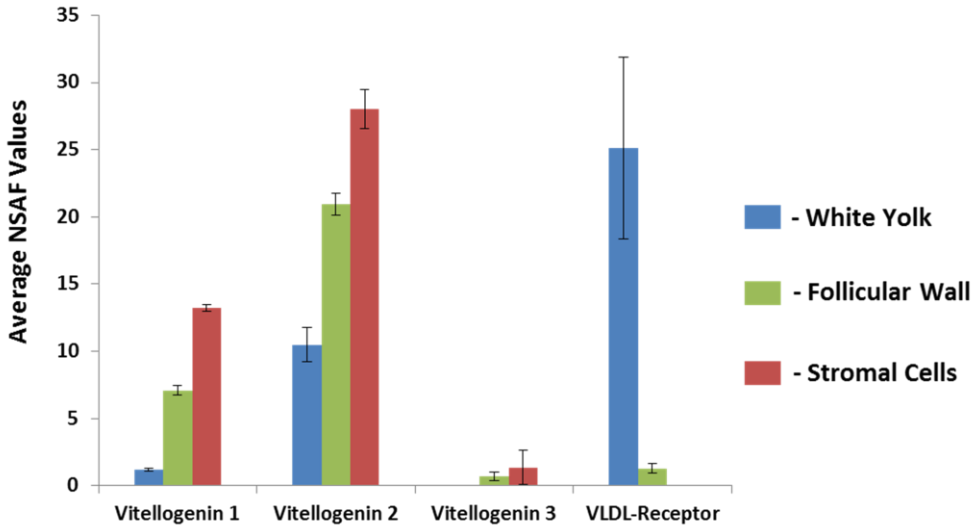


Figure 4.4 Volcano plots graphing $-\log_{10}$(p-value) versus $\log_2$ expressions. Highlighted in orange are proteins that greater than 2-fold higher concentration and statistically significant using a p-value of 0.05.

Table 4.3 Quantification of Compartmentalized Steroidogenic Proteins

					White Yolk		Follicular Wall		Stromal Cells	
	Sequence Id NCBI -Accession Number	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Avg SpC	Avg NSAF (x10,000)	Avg SpC	Avg NSAF (x10,000)	Avg SpC	Avg NSAF (x10,000)
Steroidogenic Pathway	gi|1906771|dbj|BAA18920.1|	cytochrome P450scc	508	58.1	0.0	0.0	7.0	2.6	0.3	0.1
	gi|1019356|dbj|BAA07819.1|	3β-HSD-II	377	42.8	0.0	0.0	7.7	3.8	0.7	0.3
	gi|1944049|dbj|BAA19567.1|	17β-HSD-I	302	32.8	0.0	0.0	1.0	0.6	0.0	0.0
	gi|211170|gb|AAA48598.1|	aromatase	495	56.7	0.0	0.0	16.7	6.3	2.0	0.8
	gi|117282|sp|P12394.1|CP17A_CHICK	P450c17	508	56.9	0.0	0.0	25.3	9.4	2.3	0.8

As follicles are recruited and enter the preovulatory hierarchy, they increase in size in large part through the accumulation vitellogenin. Since the SWF are in an undifferentiated state and are not actively sequestering yellow yolk, vitellogenins levels should be relatively low. **Figure 5** shows the relative abundances of three types of vitellogenin and very low density lipoprotein receptor (VLDL-receptor) found in each tissue compartment. Samples were taken from reproductively active hens meaning that vitellogenin will circulate in the blood stream at relative high concentrations. As shown in **Figure 5**, vitellogenins are present at higher concentrations in the stroma, followed by the follicular wall cells. Follicles in an undifferentiated state do not need necessarily need vitellogenins to transport nutrients into the yolk. Thus the reason for the small amounts of vitellogenin detected within the yolk. VLDL-receptor is shown to be primarily localized within the white yolk. Once the follicle enters follicular differentiation, the VLDL-receptors will migrate towards the follicular wall allowing for the endocytosis of vitellogenin into the yolk.

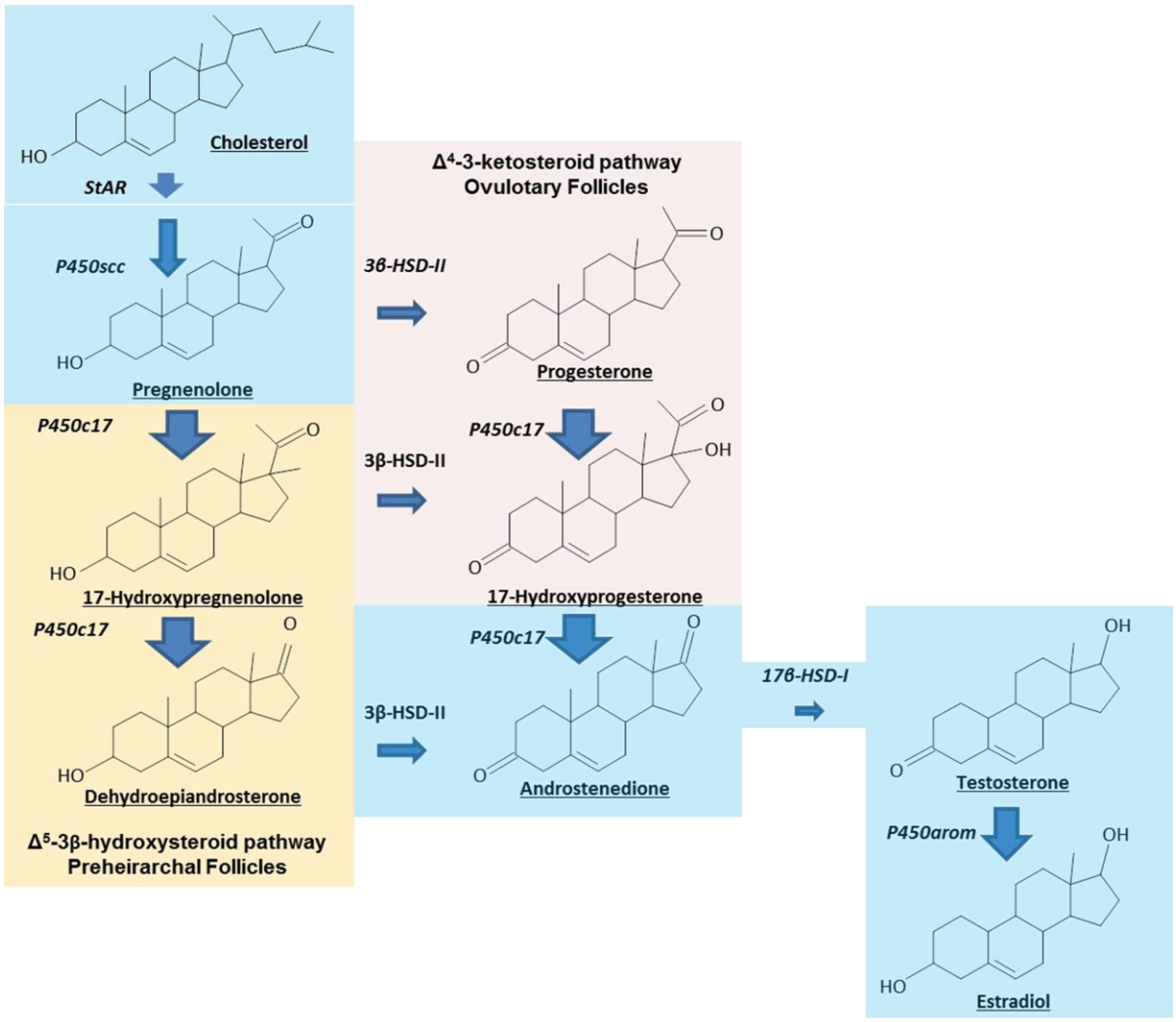


Figure 4.5 The bar graph depicts the Average Normalized Spectral Abundance Factor (NSAF) values of the Vitellogenins and VLDL-receptor with respect to their localization in the ovary. It can be seen that the VLDL-receptor is located in the white yolk, while Vitellogenins remain mostly outside the yolk since the follicles have not undergone vitellogenesis.

4.4 Discussion

4.4.1 Global Protein Profile

A global proteomic mass spectrometry based approached is demonstrated by the identification of proteins extracted from compartments obtained from chicken follicles ranging from 0.3 – 1.6 mm. Tissue compartments were harvested through the utilization of LMD. Global proteomics profiles were provided for each compartment analyzed using LC-MS/MS. The white yolk proteome resulted in 1984 proteins, while the stroma and follicular wall resulted in 2456 and 2470 respectively. Prior to this study, only proteomes of egg yolk plasma, granule and whites were established.[7,22-24] Researchers have identified up to 158 proteins in the egg white portion[24] while only 119 proteins were identified in the chicken egg yolk[7]. This is the first report of proteins identified from yolk cells that were isolated from SWF. The explanation for the large discrepancy in total proteins is not yet understood. It can be hypothesized that the highly abundant proteins such as vitellogenins, albumins and alpha-2-macroglobulins that are deposited during the development of the egg and can mask the detection of low concentration proteins. The deposition of added proteins occurs during follicular differentiation, a series of white and yellow yolk spheres appear to surround the latebra.[25] The latebra as well as the Pander's nucleus is made up of white yolk spheres. While the yellow yolk spheres concentrically surround the latebra in an alternating light and dark yolk sphere pattern. Yellow spheres constitute a majority of the yolk. It can be proposed that the

proteome identified in SWY is identical to the white yolk which constitutes the latebra.

In the present study, 50, 53, and 53 mitochondrial proteins were identified in white yolk, follicular wall and stroma compartments. The identification of mitochondrial proteins within the yolk samples agrees with Bellairs *et. al.* which demonstrated electron microscopy of small follicles in the chicken ovary revealed numerous mitochondrial structures lying at the periphery of the yolk adjacent to the granulosa cells of the follicular wall.[26] The most statistically significant biological processes was “generation of precursor metabolites and energy”, this is partially due to the number of mitochondrial proteins identified. The mitochondria are considered the power source of the cell due to the production of adenosine triphosphate (ATP) that takes place.

The DAVID Bioinformatics Database was used to categorize proteins found in their respective compartments based on biological processes and cellular component. **Figure 6** illustrates the clustering of proteins based on biological processes. As shown in **Figure 2**, 1554 proteins were proteins were identified in all three compartments. In **Figure 6**, the highlighted lavender pie pieces are some of the biological process shared throughout all three compartments. However, a few processes that were present in the follicular wall and stromal cell proteins were cellular Macromolecule metabolic process, Negative regulation of biological process, and Negative regulation of cellular process. These biological processes are important for regulation of cell death and atresia. **Figure 7** illustrates the

classification of proteins based on their cellular component. There were cellular components identified only in the follicular wall and stromal cells. Many of these proteins are expected to originate from the nucleus. The absence of proteins originating from the nucleus supports that no nuclei were sampled from the white yolk of the follicles chosen.

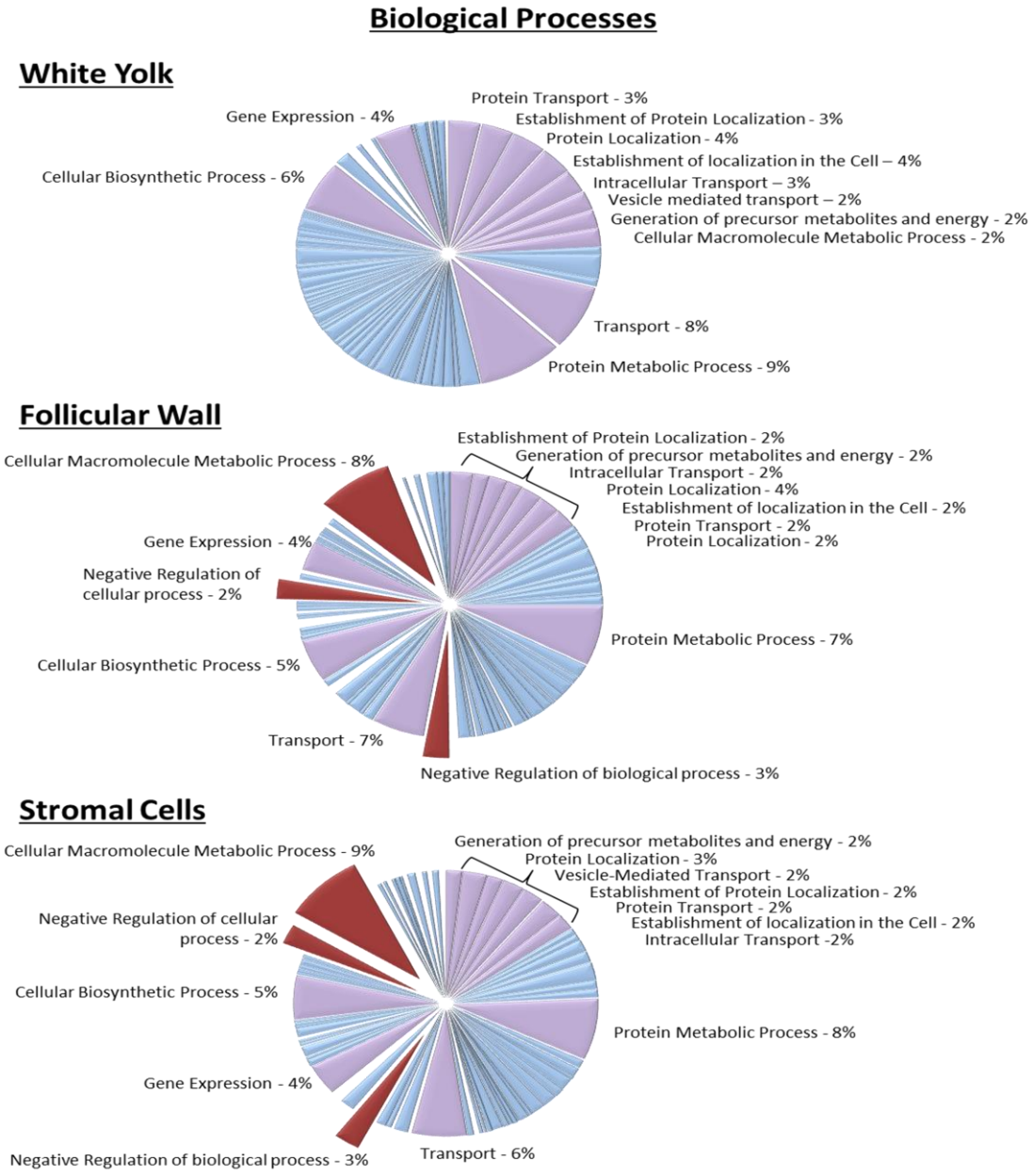


Figure 4.6 Biological processes of the proteins identified from the individual compartments using DAVID algorithm. Major biological processes that differed from the follicular wall and stroma to that of the white yolk are highlighted in maroon.

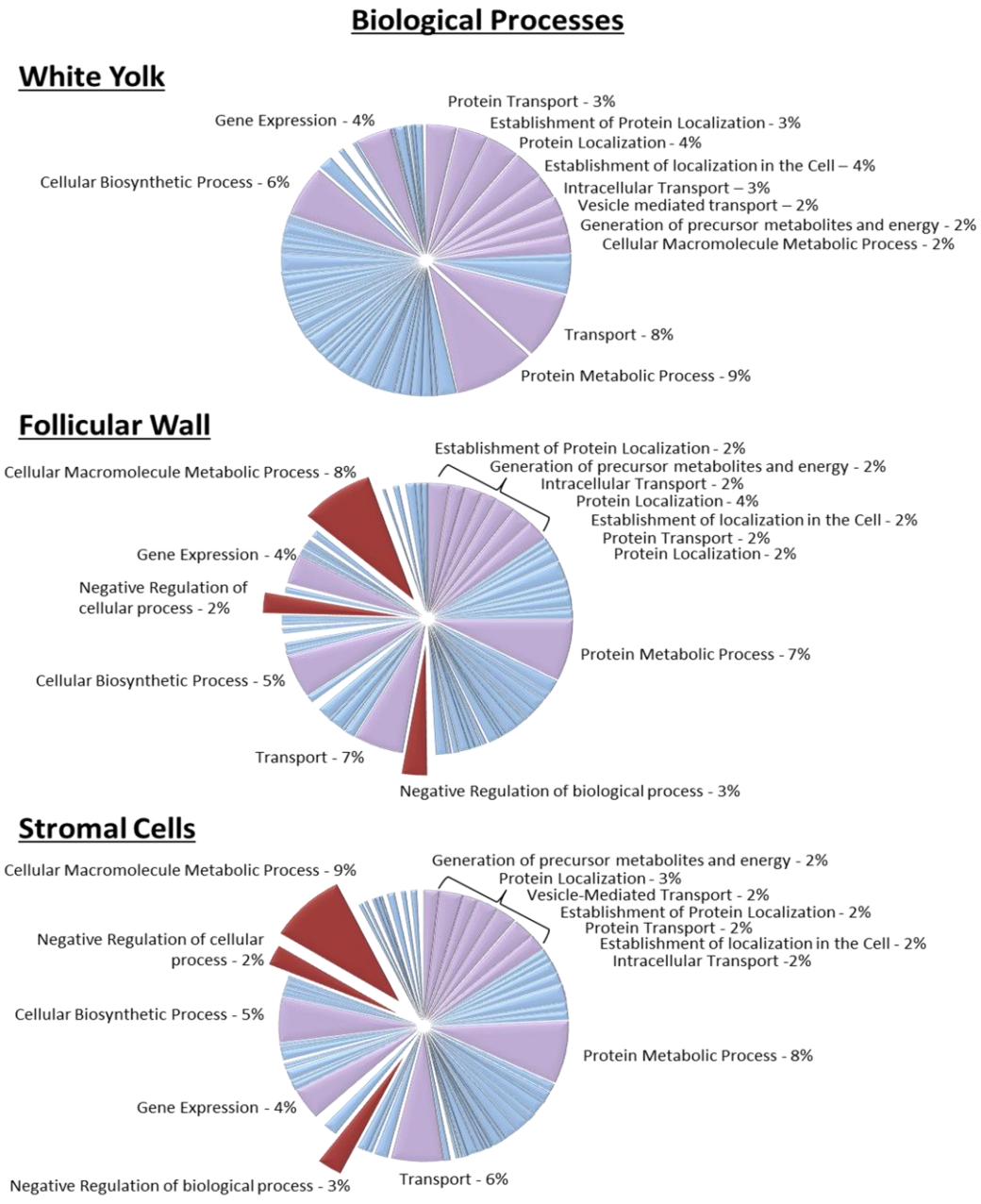


Figure 4.7 Cellular components of proteins identified from the individual compartments. Expressed in the follicular wall and stroma were proteins identified to be nuclei related.

4.4.2 Protein Compartmentalization

Compartmentalization of proteins are demonstrated through cellular separation provided by using LMD, a prime example is shown by the detection of Vitellogenin I,II, III. **Figure 5** is a bar graph demonstrating Average Normalized Spectral Abundance Factors (Avg-NSAF) of Vitellogenin 1, 2, 3 and VLDL receptor in the white yolk, granulosa cells and stroma obtained from SWF. Concentrations of vitellogenin are greatest in the stromal tissue. Vitellogenins are yolk precursor proteins produced by the liver and circulate within the bloodstream till the follicle enters a stage of vitellogenesis in which endocytosis of vitellogenins transporting them into the yolk. These proteins are essential for egg production in order to provide the nutrients to the follicles at defined developmental stages. SWF were obtained from healthy egg-producing hens, levels in which vitellogenins are high, especially in the stroma and granulosa cells. While concentrations of vitellogenins were minimal within the white yolk of undifferentiated follicles, it is not uncommon for Vitellogenin 2 to be present in the yolk even at early stages. Vitellogenin 2 has been shown to enter the yolk proceeding Vitellogenin 1, and 3 which has been reported by Williams and co-workers in striped bass ovaries.[27] The proportional abundances of Vitellogenins 1, 2, 3 in this study are similar to what was found in egg yolk plasma and granule presented by Mann *et. al..*[7] The opposite trend of was observed with very low density lipoprotein (VLDL) receptor protein. VLDL receptor mediates a key step in the reproductive efforts of the hen, in the absence of VLDL receptor, oocytes are unable to enter the rapid growth phase of the follicle.[28] These receptor proteins

are maintained in the cytoplasm of the follicle till transport of lipoproteins (Vitellogenins, apovitellenin) are needed within the oocyte. They will then assemble towards the follicular wall allowing for endocytosis of essential lipoproteins.

A shift in steroidogenesis during follicular maturation and development is a hallmark event.[18] Targeting the identification of steroidogenic enzymes becomes very important. **Figure 4** is a schematic of the steroidogenic pathway, labeled are the metabolites and enzymes associated at each step of the pathway. All enzymes were identified in the pathway except for the StAR protein. **Table 3** shows the Average Normalized Spectral Counts (AvgNSpC) and the Normalized Spectral Abundance Factors (NSAFs) for the proteins identified in the steroidogenic pathway. The proteins identified are predominantly found in the follicular wall cells while minimal amounts were found in the stromal tissue. Steroidogenisis can occur in stromal tissue containing cortical follicles, producing estrogen.[18] Our findings of not observing StAR protein corresponds with transcript analysis Johnson *et. al.* which notes that undifferentiated cells from prehierarchal follicles will fail to express StAR, meanwhile differentiated granulosa cells from the preovulatory follicles display an up-regulation of StAR leading to the production of progesterone.[6] The small white undifferentiated follicles are known to be a major source of dehydroepiandrosterone (DHEA), estradiol, and androsetendione[18] and are produced through the Δ^5-3β-hydroxysteroid pathway(**Figure 5**). Knowing that undifferentiated follicles go through the Δ^5-3β-hydroxysteroid pathway and relatively high amounts of P450c17 present, pregnenolone is converted to DHEA. While in turn the DHEA is converted to

androstenedione. Relatively lower amounts of the 17β-HSD-I protein, it can be suggested that not all the androstenedione is converted to testosterone. However, with higher amounts of aromatase present in can be conceived that all the testosterone is converted to estradiol. In the early stages of follicular development, studies have demonstrated estradiol as a key component in preventing follicular atresia and the oxidative stress pathways since it can influence the expression of antioxidants.[29,30] The relative amounts of each steroid cannot be quantified based on solely the amounts of proteins present. The relative levels of enzymes correlates well with the production of DHEA, androsenedione, and estradiol produced by gonadotropin-stimulated cultured small white follicles.[18]

Along with steroidogenic gene mRNA that are studied in the follicular development, the transcript levels of many other genes are studied.[5] **Table 2** is of a few gene mRNAs typically studied in follicular development, on the right side of the table are the corresponding translated proteins that were identified or not detected in our global proteome data. Anit-Mullerian Hormone (AMH) an important protein studied by researchers to because it's known to be involved in the regression of the Mullerian ducts in birds. In male birds the Mullerian ducts regress completely, while in females, only the right ducts regress allowing the left oviduct to develop. The left ovary is protected from the AMH due to the presence of estrogen.[31,32] Johnson and co-workers have used quantitative real time PCR to study Anit-Mullerian hormone (AMH) expression in the granulosa cells of various follicles sizes.[33] Their study concluded that granulosa cells obtained from follicles of 1 mm in size had expressed

the largest amounts of AMH over the larger follicles examined. Observations made from the global proteomics data indicated that AMH was present and only detected in the follicular wall cells. Other ovary developing associated proteins identified were Retinoic Acid Receptor-alpha isoform 1, and retinaldehyde dehydrogenases (RALDH1, RALDH2, RALDH3). Retinoic acid (RA) is studied because it is involved in numerous cell functions such as cell growth, differentiation and is used in binding to nuclear receptors RAR and RXR to regulate gene expression.[34,35] These cellular processes occur by binding to nuclear receptors RAR and RXR which in turns regulates gene expression. However, RA is controlled by the expression of RA-synthesizing retinaldehyde dehydrogenase (RALDH1, RALDH2, RALDH3).[36] Unlike AMH which is primarily found in the follicular wall cells, the RALDH family of enzymes was found mainly in the stromal and white yolk cells. Smith *et. al.* observed RALDH2 mRNA is expressed in the ovarian cortex.[37] They concluded expression of RALDH2 and CYP26B1 potentially leads to the RA synthesis and degradation suggesting a conserved role for retinoic acid in vertebrate meiosis.

Follicular Stimulating Hormone (FSH) and Follicular Stimulating Hormone Receptor (FSHR) are valuable biomarkers used to study the ovarian follicular maturation. Woods and co-workers have noted follicles can express Follicular-Stimulating Hormone receptor (FSHR) mRNA as early as 1-2 mm follicles.[38] However, the highest levels occur in prehierarchal follicles (6-8 mm) and newly selected preovulatory follicles (9-12 mm). FSHR was not detected in the current, possibly due to low concentrations in follicles between .3 – 1.6 mm. Woods *et. al.*

studied three distinct stages of follicular maturation, undifferentiated granulosa cells obtained from preheirarchal follicles, granulosa cells recently selected follicles (9-12 mm), and differentiated cells from the preovulatory follicles. Their study of undifferentiated granulosa cells demonstrates the activation of either EGFR receptor mediated MAPK or PKC signaling suppresses FSH-induced StAR protein expression and progesterone production. In our study, the presence of EGFR, PKC, and MAPK proteins and the absence of StAR and FSHR protein agrees with mRNA data of undifferentiated granulosa cells presented by Woods and co-workers.

In a global proteome study a large number of proteins were identified. Sorting through the data is a daunting task. Several articles have sited many important genes that are expressed in the ovary. These genes were then target for protein identification. Several important transcription factors identified were; GATA-4, SMAD2, and FOXL2. The GATA4 and FOXL2 were only detected in the follicular wall cells. Unlike SMAD2 which was found at slightly higher concentration in the stroma than in the follicular wall. More recently studies have involved BMP6 as its role in follicular development. Ocón-Grove and co-workers demonstrate that a concentration of BMP6 mRNA is greatest in small follicle (1 – 2 mm) and that levels decrease with the maturation of the follicle.[21] Unfortunately, BMP6 was not detected in the proteomics data. BMP6 was undetectable due to its low level of circulation in the system. Researchers have employed several enrichment steps to detect BMP proteins using mass spectrometry.[39]

4.5 Conclusion

Results from this study shows that LMD can be successfully be employed for global proteomics on follicles that are less than 2 mm in size. Many of the proteins that were identified are currently being examined by the mRNA complement for follicular development. However, the measured mRNA levels will not necessarily correlate to the protein abundances.[40] Abundances between mRNA and protein levels can differ greatly due to post-translation modification, translational and protein degradation regulations. An added feature of proteomics studies allows for post-translation modifications to be studied. Though this study examined only follicles of < 2 mm, it can be amended to other stages of development. This method was to assess its ability to utilize laser microdissection of small amounts of tissue and obtaining ample proteomics data for further studies.

4.6 References

(1) Huang, E. S.; Kao, K. J.; Nalbandov, A. V. Synthesis of sex steroids by cellular components of chicken follicles. *Biol Reprod* **1979**, *20*(3), 454-61.

(2) Onagbesan, O. M.; Peddie, M. J. Steroid secretion by ovarian cells of the Japanese quail (Coturnix coturnix japonica). *Gen Comp Endocrinol* **1988**, *72*(2), 272-81.

(3) Wang, S. C.; Bahr, J. M. Estradiol secretion by theca cells of the domestic hen during the ovulatory cycle. *Biol Reprod* **1983**, *28*(3), 618-24.

(4) NITTA, H.; OSAWA, Y.; BAHR, J. M. Multiple Steroidogenic Cell Populations in the Thecal Layer of Preovulatory Follicles of the Chicken Ovary. *Endocrinology* **1991**, *129*(4), 2033-2040.

(5) Diaz, F. J.; Anthony, K.; Halfhill, A. N. Early avian follicular development is characterized by changes in transcripts involved in steroidogenesis, paracrine signaling and transcription. *Mol Reprod Dev* **2011**, *78*(3), 212-23.

(6) Johnson, A. L.; Solovieva, E. V.; Bridgham, J. T. Relationship between steroidogenic acute regulatory protein expression and progesterone production in hen granulosa cells during follicle development. *Biol Reprod* **2002**, *67*(4), 1313-1320.

(7) Mann, K.; Mann, M. The chicken egg yolk plasma and granule proteomes. *Proteomics* **2008**, *8*(1), 178-91.

(8) Gokce, E.; Franck, W. L.; Oh, Y.; Dean, R. A.; Muddiman, D. C. In-Depth Analysis of the Magnaporthe oryzae Conidial Proteome. *J Proteome Res* **2012**, *11*(12), 5827-5835.

(9) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **2009**, *6*(5), 359-U60.

(10) Andrews, G. L.; Shuford, C. M.; Burnett, J. C., Jr.; Hawkridge, A. M.; Muddiman, D. C. Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, *877*(10), 948-54.

(11) Randall, S.; Cardasis, H.; Muddiman, D. Factorial Experimental Designs Elucidate Significant Variables Affecting Data Acquisition on a Quadrupole Orbitrap Mass Spectrometer. *J Am Soc Mass Spectr* **2013**1-12.

(12) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*(18), 3551-3567.

(13) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **2002**, *74*(20), 5383-5392.

(14) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **2003**, *75*(17), 4646-4658.

(15) Weatherly, D. B.; Atwood, J. A.; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R. A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Molecular & Cellular Proteomics* **2005**, *4*(6), 762-772.

(16) Dennis Jr, G.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **2003**, *4*(5), P3.

(17) Johnson, P. A.; Kent, T. R.; Urick, M. E.; Trevino, L. S.; Giles, J. R. Expression of anti-Mullerian hormone in hens selected for different ovulation rates. *Reproduction* **2009**, *137*(5), 857-63.

(18) Robinson, F. E.; Etches, R. J. Ovarian steroidogenesis during follicular maturation in the domestic fowl (Gallus domesticus). *Biol Reprod* **1986**, *35*(5), 1096-105.

(19) Woods, D. C.; Haugen, M. J.; Johnson, A. L. Actions of epidermal growth factor receptor/mitogen-activated protein kinase and protein kinase C signaling in granulosa cells from Gallus gallus are dependent upon stage of differentiation. *Biol Reprod* **2007**, *77*(1), 61-70.

(20) Johnson, A. L.; Woods, D. C. Dynamics of avian ovarian follicle development: Cellular mechanisms of granulosa cell differentiation. *Gen Comp Endocrinol* **2009**, *163*(1-2), 12-17.

(21) Ocon-Grove, O. M.; Poole, D. H.; Johnson, A. L. Bone morphogenetic protein 6 promotes FSH receptor and anti-Mullerian hormone mRNA expression in granulosa cells from hen prehierarchal follicles. *Reproduction* **2012**, *143*(6), 825-33.

(22) D'Ambrosio, C.; Arena, S.; Scaloni, A.; Guerrier, L.; Boschetti, E.; Mendieta, M. E.; Citterio, A.; Righetti, P. G. Exploring the chicken egg white proteome with combinatorial peptide ligand libraries. *J Proteome Res* **2008**, *7*(8), 3461-3474.

(23) Mann, K. The chicken egg white proteome. *Proteomics* **2007**, *7*(19), 3558-68.

(24) Mann, K.; Mann, M. In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos. *Proteome Sci* **2011**, *9*(1), 7.

(25) Romanoff, A. L. R., A.J.; J. Wiley & Sons: New York, 1949.

(26) Bellairs, R. The relationship between oocyte and follicle in the hen's ovary shown by electron microscopy. *J. Embryol. Exp. Morph.* **1965**, *13*(2), 215-233.

(27) Williams, V. N.; Reading, B. J.; Amano, H.; Hiramatsu, N.; Schilling, J.; Salger, S. A.; Islam Williams, T.; Gross, K.; Sullivan, C. V. Proportional accumulation of yolk proteins derived from multiple vitellogenins is precisely regulated during vitellogenesis in striped bass (Morone saxatilis). *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology* **2014**n/a-n/a.

(28) Nimpf, J.; Radosavljevic, M. J.; Schneider, W. J. Oocytes from the Mutant Restricted Ovulator Hen Lack Receptor for Very Low-Density Lipoprotein. *J. Biol. Chem.* **1989**, *264*(3), 1393-1398.

(29) Murata, H.; Ihara, Y.; Nakamura, H.; Yodoi, J.; Sumikawa, K.; Kondo, T. Glutaredoxin exerts an antiapoptotic effect by regulating the redox state of Akt. *J Biol Chem* **2003**, *278*(50), 50226-33.

(30) Urata, Y.; Ihara, Y.; Murata, H.; Goto, S.; Koji, T.; Yodoi, J.; Inoue, S.; Kondo, T. 17Beta-estradiol protects against oxidative stress-induced cell death through the glutathione/glutaredoxin-dependent redox regulation of Akt in myocardiac H9c2 cells. *J Biol Chem* **2006**, *281*(19), 13092-102.

(31) Teng, C. S. Differential expression of c-Jun proteins during mullerian duct growth and apoptosis: caspase-related tissue death blocked by diethylstilbestrol. *Cell Tissue Res* **2000**, *302*(3), 377-85.

(32) Teng, C. S. Differential activation of MAPK during Mullerian duct growth and apoptosis: JNK and p38 stimulation by DES blocks tissue death. *Cell Tissue Res* **2001**, *306*(1), 27-34.

(33) Johnson, P. A.; Kent, T. R.; Urick, M. E.; Giles, J. R. Expression and regulation of anti-mullerian hormone in an oviparous species, the hen. *Biol Reprod* **2008**, *78*(1), 13-9.

(34) Niederreither, K.; Subbarayan, V.; Dolle, P.; Chambon, P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat Genet* **1999**, *21*(4), 444-448.

(35) Sakai, Y.; Meno, C.; Fujii, H.; Nishino, J.; Shiratori, H.; Saijoh, Y.; Rossant, J.; Hamada, H. The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. *Gene Dev* **2001**, *15*(2), 213-225.

(36) Ishimaru, Y.; Komatsu, T.; Kasahara, M.; Katoh-Fukui, Y.; Ogawa, H.; Toyama, Y.; Maekawa, M.; Toshimori, K.; Chandraratna, R. A.; Morohashi, K.; Yoshioka, H. Mechanism of asymmetric ovarian development in chick embryos. *Development* **2008**, *135*(4), 677-85.

(37) Smith, C. A.; Roeszler, K. N.; Bowles, J.; Koopman, P.; Sinclair, A. H. Onset of meiosis in the chicken embryo; evidence of a role for retinoic acid. *BMC Dev Biol* **2008**, *8*85.

(38) Woods, D. C.; Johnson, A. L. Regulation of follicle-stimulating hormone-receptor messenger RNA in hen granulosa cells relative to follicle selection. *Biol Reprod* **2005**, *72*(3), 643-650.

(39) Razdorov GV, S. (ed.). The Use of Mass Spectrometry in Characterization of Bone Morphogenetic Proteins from Biological Samples. Tandem Mass Spectrometry - Applications and Principles. 2012.

(40) Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **1999**, *19*(3), 1720-1730.