11-1-2014

# Development and Testing of a Rapid Multiplex Assay for the Identification of Biological Stains

Kevin M. Legg
*University of Denver*

# Development and Testing of a Rapid Multiplex Assay for the Identification of Biological Stains

---

A Dissertation

Presented to

The Faculty of Natural Sciences and Mathematics

University of Denver

---

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

---

By

Kevin M. Legg

November 2014

Advisor: Phillip B. Danielson

Author: Kevin M. Legg
Title: Development and Testing of a Rapid Multiplex Assay for the Identification of Biological Stains
Advisor: Phillip B. Danielson
Degree Date: November 2014

## Abstract

While DNA profiling makes it possible to individualize biological stains, the identification of the stain itself can present forensic serologists with a significant challenge. Current antibody- and enzyme activity-based assays used by forensic practitioners for biological stain identification yield only presumptive results. Positive results with non-target body fluids or cross-reactivity with non-human sources has also been well documented. Some tests can consume unacceptable quantities of precious evidence while for some body fluids (vaginal fluid and menstrual blood) there are simply no available tests at all.

This research presented here aims to develop and rigorously test a fast, accurate, and sensitive multiplex assay for the simultaneous identification of saliva, semen, urine, peripheral blood, menstrual blood, and vaginal secretions. This research is based on the following three research phases:

1. Biomarker Identification – Utilizing multidimensional protein separation technologies, bioinformatics tool, and tandem mass spectrometry a database of fluid-specific candidate markers will be developed for each body fluid.
2. Biomarker Verification – Using the most promising candidates from Phase 1 a targeted multiplex assay will be developed on a Quadruple Time-of-Flight mass spectrometer to verify the specificity of these candidate biomarkers with single

source laboratory samples as well as single and mixed source casework type samples.

3. Prototype Validation – Development and testing of a case working laboratory prototype assay on a triple quadrupole mass spectrometer using single and mixed source casework-type samples.

This work has the potential to significantly improve the accuracy and sensitivity of forensic serological testing. It will provide practitioners with greatly improved tests for saliva and seminal fluid while also enabling the identification of vaginal secretions for which there are currently no accurate tests. The multiplex design will eliminate the need to perform separate tests on an unknown stain. In short, the successful completion and implementation of this research will provide the forensic community with a powerful tool to aid in criminal investigations.

**Acknowledgements**

To my advisor Phil Danielson, thank you for the years of opportunity, guidance, but most of all the freedom to grow this project to where it is now. I will never forget the 100 hour exam, late night and early morning grant writing, and countless discussions over a glass of single malt scotch.

To Nichole, Rick, Roger, and Mike. Thank you for trusting a lowly computer scientist on the instruments back in 2008 as well as the assistance on everything mass spectrometry.

I would like to thank my graduate committee, Joyce Sterling, Jim Fogleman, and finally to Joe Angleson who was willing to participate in my defense committee at the last minute.

Thank you to my family as well. To my Dad who supported me when I wanted to abandon the life of an engineer. To my family who spread across the United States, but especially Judy and Tony here in Denver, for the encouragement along the way.

Lastly, I would like to thank all of my friends, roommates, suitemates, and acquaintances who have willingly donated samples and listened to my ratings and ravings about body fluids for the past six years.

## Table of Contents

Chapter 3: Biomarker Verification



Chapter 4: Prototype Validation

## List of Figures

## List of Tables

# Chapter 1:

# Introduction

# 1 Introduction

The development of DNA typing technologies has transformed the world of forensic science by allowing the routine individualization of biological stains. Despite this tremendous power of this technology, it is not difficult to point to situations where the unambiguous identification of a biological fluid can be very informative and even critical for investigators. Obtaining a confirmatory identification of a biological stain, however, can present a challenge. Currently, there are no dependable tests for vaginal or menstrual fluids. Other fluids (*e.g.,* saliva) have well established assay protocols for detection but theses can provide only presumptive results.

## 1.1 History and Importance of Forensic Serology

Forensic serology can be broadly defined as the discipline concerned with the characterization of biological fluids present on evidentiary material. Currently, the identification of unknown biological stains relies upon the use of microscopic visualization, immunological affinity, chemical reaction and enzymatic activity tests in order to locate and/or identify the source of a stain on a piece of evidentiary material. Once located, stains are processed further in order to generate DNA profiles to identify specific the individual who was the source of the material.

In the context described above, forensic serology is being used as a first-pass screen of evidentiary material prior to and independent of DNA analysis. Traditionally, however, serological assay techniques attempted to combine both stain characterization as well as the individualization of a stain. The most notable example of this being the ABO blood typing system developed by the Austrian physician Karl Landsteiner[1] and

implemented for forensic practice by Leon Lattes in the early 1900's[2]. While Lattes' agglutination method can successfully identify blood, the discriminating power of the technique for individualizing bloodstains is so low that there is a nearly 40% chance that any two randomly selected unrelated individuals will have the same blood type[3]. Furthermore, the agglutination-based test lacked sensitivity and, the result indicating an "AB" type could not be differentiated from that of a failed test result since the identification relies upon the lack of a lack of an agglutination reaction[4]. After decades of research and development, the goal of individualizing stains using serological markers produced only moderate advances. One notable method combined ABO markers along with polymorphic proteins markers such as phosphoglucomutase (PGM) into a single testing system[5, 6]. While the combination of these two markers was an improvement, there is still a 20% chance that two unrelated individuals selected at random will have the exact same type[3].

In recent years, modern DNA typing has allowed for the individualization of biological stains such that the chance of finding an unrelated individual, selected at random from the general population, having the same DNA type as that of a suspect in a criminal can be as low as one out of billions to one out of quadrillions[7]. Because the discrimination power of DNA is so high, compared to historical serological typing methods, individualization by serological typing has largely been abandoned. As such, serological testing in a modern forensic laboratory is considered to be strictly a screening device for locating biological material which will be subsequently used for DNA profiling. Nonetheless, serological screening still has important value as an investigative

tool in many forensic cases. The largest reason for this being that DNA is unable to provide insight as to the tissue source of a DNA profile. Specifically, DNA alone cannot indicate whether it profile originated from saliva, vaginal fluid, or blood – information which may be able to provide context to a criminal investigation.

For example, consider the case of an alleged sexual assault where DNA consistent with the victim is recovered from the neck of a bottle. The victim alleges that the suspect vaginally penetrated her with the bottle while the suspect claims that no sexual contact occurred – offering instead that the alleged victim had simulated oral sex on the bottle. Both stories "explain" the presence of the victim's DNA profile on the bottle. The ability to unambiguously detect saliva vs. vaginal fluids or a mixture of saliva and vaginal fluids could potentially help investigators to refute one of these opposing claims. A number of other scenarios could easily be imagined where the ability to characterize body fluids and to clearly detect and identify even microscopic traces of urine, seminal fluid and saliva or to differentiate between menstrual blood and peripheral blood, would have important probative value. Such capabilities would have clear benefits for criminal investigations and would enable forensic analysts to make more definitive statements about the potential tissue source of a DNA profile.

### 1.2 Serological Testing Assays

Serological testing relies upon four general categories of testing. They include chemical reaction assays, enzyme activity assays, immunoassays, as well as microscopy based assays. Each individual test can be further classified as being presumptive or

confirmatory for its designated target fluid. Presumptive tests will reveal the possible presence of a particular body fluid whereas a confirmatory test will unambiguously identify an unknown stain as having a specific biological origin. The following section will describe the four general testing categories.

### 1.2.1 Chemical Reaction Assays:

Chemical reaction tests are based on the formation of crystals or a color change in assay reagents as they interact with substrates in a target fluid. One example of a colorimetric assay, the Kastle-Meyer test, produces a pink color when exposed to hemoglobin present in blood[8, 9]. In this reaction **(Figure 1),** phenolphathalin is prepared in an alkaline zinc solution and is added to the stain along with hydrogen peroxide. Hemoglobin in the blood has peroxidase-like activity producing water as well as oxygen-free radicals. The oxygen radicals, in turn, catalyze the oxidation of the phenolphathalin to phenolphathalein, producing a bright pink color.

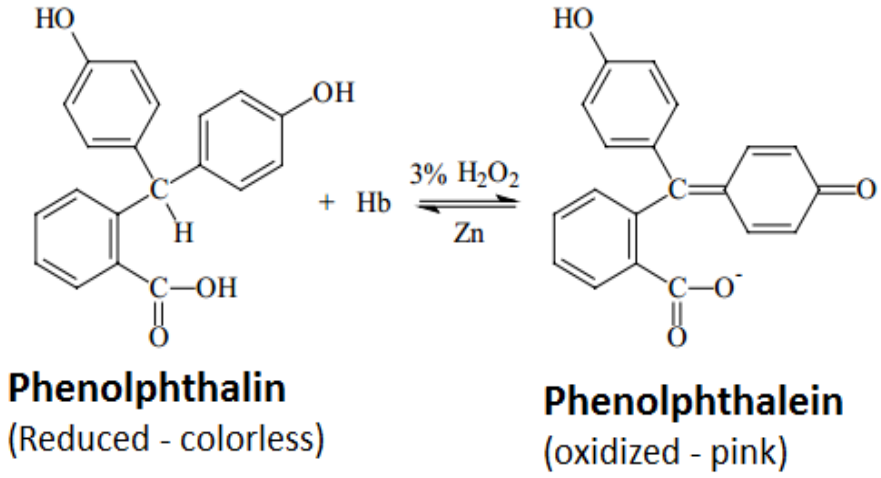



Figure 1 - Kastle-Meyer reaction, phenolphthalein is oxidized in the presence of hydrogen peroxide and hemoglobin giving a chemical indication of blood. Image adapted from Winchester[10].

An example of a crystal-producing test is the Teichmann test for blood[11]. Crystals form as a result of the chemical reaction between the iron component of hemoglobin and the Teichmann reagent, a solution of potassium bromide, potassium chloride, potassium iodide and glacial acetic acid.

While chemical reaction based assays are generally sensitive, false positives are common. Peroxides in plants as well as chemical oxidants are common causes of false positive results[12]. In addition, assays requiring the formation of detectible crystals lack the sensitivity and reliability necessary to be of utility in cases where only trace quantities of evidentiary material are available.

### 1.2.2 Enzyme Activity Assays

Enzymes are proteins which catalyze chemical reactions between substrates to generate a specific product[13]. For enzyme molecules which are present in high concentration in a given body fluid, the specific catalytic activity of the enzyme can be used to screen for the possible presence of an unknown stain. One example, prostatic acid phosphatase (also known as seminal acid phosphatase or SAP), hydrolyzes the phosphate group from its substrate molecule. This enzyme is present in high amounts in the prostate gland and detection of its activity is one of the most common methods to screen evidence for the presence of seminal fluid [14]. In this test, the phosphate group on α-naphthyl phosphate (the assay substrate) is cleaved forming a reactive species. To carry out this test, sodium α-naphthyl phosphate and the dye Brentamine Fast Blue are applied to a suspected seminal fluid stain. Enzymatically active SAP cleaves the phosphate group on

α-naphthyl phosphate which reacts with Brentamine Fast Blue to generate a dark purple color (**Figure 2**).

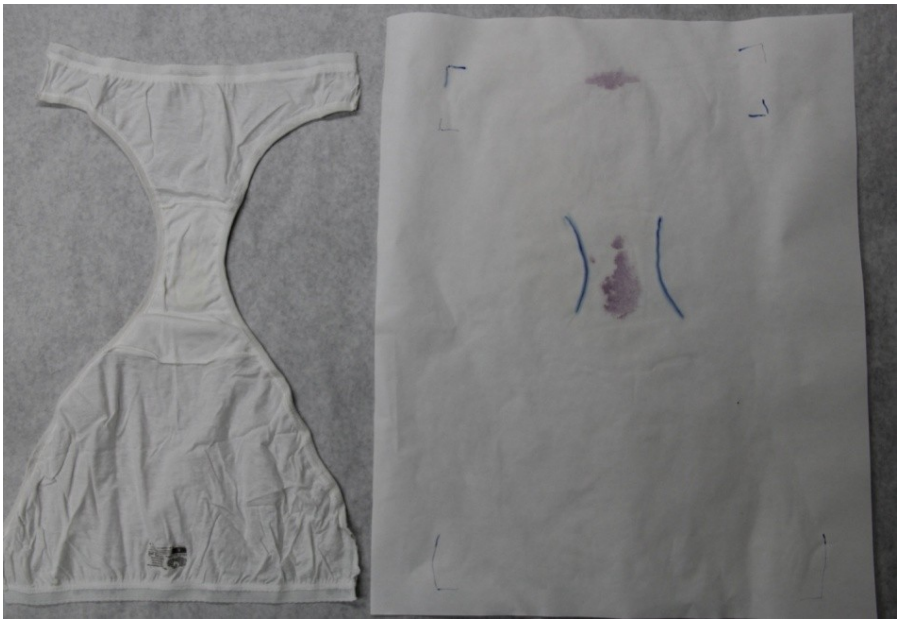


**Figure 2 – SAP overlay to localize potential seminal fluid stains. Moistened filter paper is pressed on top of the evidentiary material to draw out active enzyme by capillary action. Following the addition of α-naphthyl phosphate and Brentamine Fast Blue, a dark purple color indicates the presence of seminal fluid.**

While these tests are typically sensitive, the results are only presumptive for a target body fluid. This is due primarily to the fact that although a given enzyme may be an abundant component of the body fluid to which the assay is targeted, the same enzyme may also be present in one or more non-target body fluids or tissues, albeit at lower concentrations. The level of SAP in blood serum, for example, has been used as a marker for prostate cancer in males[15]. Additionally, proteins are subject to a loss of activity due to factors such as microbial degradation, exposure to chemical agents and inhibitors and extremes of temperature or pH[16-18].

### 1.2.3 Immunoassays

Immunoassays are analytical methods which employ antibody-antigen binding events to detect specific target molecules. When properly and thoroughly characterized, it is usually possible to isolate antibodies that are highly specific and sensitive for nearly any target molecule of interest. In relation to forensic serology, antibodies are selected to target molecules which are specific to a particular body fluid. Early immunoassays commonly employed in case work labs were based on the visualization of a precipitation reaction between antigens and their corresponding antibody. Using the Ouchterlony double diffusion method, a series of wells are punched into an agarose gel such that there is a central "antibody well" surrounded by additional "antigen wells". A solution containing antibodies against a specific protein is placed in the central well. Controls and extracts of questioned stains are added to the surrounding antigen wells. Diffusion of the sample antigens and the antibodies toward each other will occur and a precipitation line, indicating a positive result for the test, will appear if a sufficient quantity an antibody-antigen complex forms (**Figure 3**).

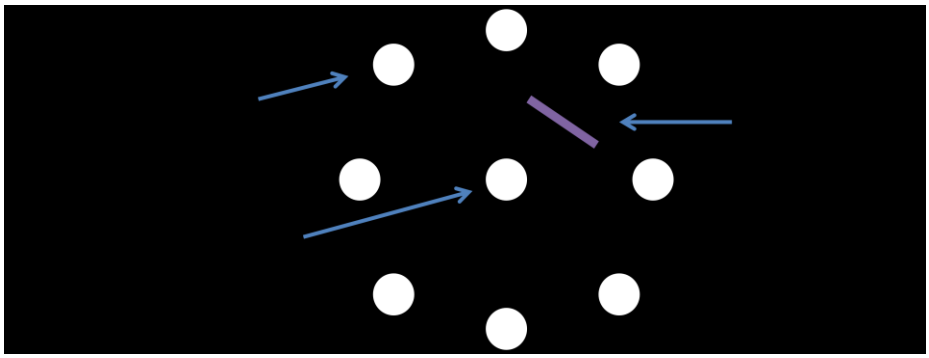


Figure 3 – Antigen-antibody precipitate formation indicating a positive test result.

More sophisticated methods, such as cross-over electrophoresis, use an electric field as an alternative to diffusion to drive antigens across an agar gel. As with the Ouchterlony method, a precipitate will form if a sufficient quantity of a target antigen is present.

Immunochromatographic assays have gained popularity in recent years for their simplicity, specificity, cost, and speed[19, 20]. Commercial immunochromatographic assays are sold as single-use lateral flow strips that target specific proteins present in a given body fluid. The strips (**Figure 4**) contain a sample loading well, lateral flow membrane, and three antibodies. To run the test, a liquid extract from a questioned sample is introduced onto the conjugate pad where it mixes with colloidal gold labeled antibodies. Samples which contain the target protein form an antigen-antibody-colloidal gold complex that migrates by bulk fluid flow across the lateral flow membrane. Immobilized antibodies, specific for the target protein, capture the antigen- antibody-colloidal gold complex as it migrates down the test strip. The aggregation of colloidal gold conjugated antibodies creates a visible band at the Test Line. Free antibodies not captured at the Test Line continue toward the control line where they are captured by immobilized antibodies targeted directly to the antibody portion of the antibody-colloidal gold conjugate. This produces a second visible band at the Control Line indicating that the test performed as designed.

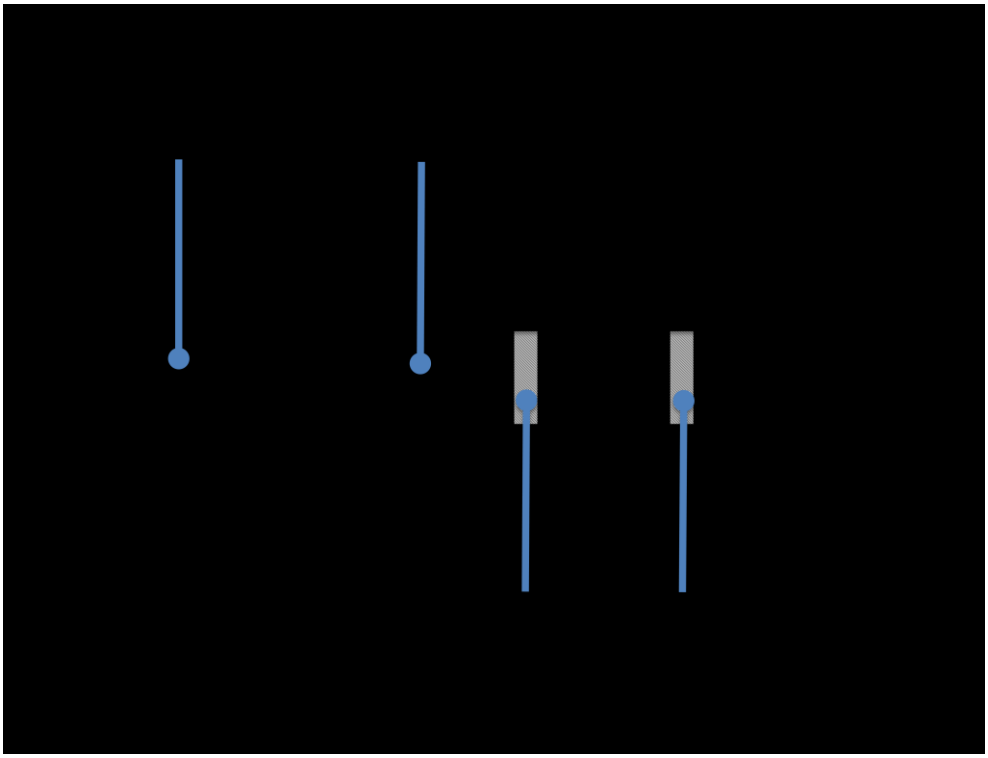


**Figure 4 - Lateral Flow Immunochromatographic Strip. A positive result is indicated by visible bands at both the Test Line and the Control Line.**

While sensitive and easy to use, antibodies have limitations which limit their probative value. The primary issue is that protein antigen-antibody binding is rarely 100% specific; thus, false positive results are possible. This arises from several sources: first, proteins are known to have variability in structure due to genetic polymorphisms as well as differential post-translational modification[21]. Because immunoassays are based on binding a specific epitope of a target protein, any modification to this structure can result in a false negative result. Second, along with variability of the individual protein, it is possible for non-target molecules with a similar or identical epitope to be recognized by the antibody generating false positive results. Finally, it is possible for the target antigen to be at such a high concentration that the immobilized antibodies at the test line

are more likely to bind free antigen than the antigen-complexed mobile-phase antibodies. Referred to as the "hook effect", this can produce false negatives or falsely indicate very low amounts of the target body fluid when, in fact, the exact opposite is true[21].

### 1.2.4 Microscopy Based Assays

Microscopy is simply the use of a series of lenses to view objects which are too small to see with the naked eye. Applied to forensic serology, microscopy has primarily been used as a means of identifying sperm cells on sexual assault evidence. Sperm cells can be both detected and identified as being of human origin using visible light or fluorescence microscopy by a skilled analyst knowledgeable in the morphological characteristics of human sperm. Because spermatozoa are much smaller than the vaginal or other epithelial cells with which they are often comingled in a forensic context, they can be difficult to localize in the microscopic field. With visible light microscopy, detection is facilitated by the use of histological stains such as the combination of hematoxylin and eosin or nuclear-fast red and picroindigocarmine[22, 23]. The latter more commonly known as "Christmas tree'' stain is particularly useful because it differentially stains spermatozoa and epithelial cells. Using this stain, the sperm head will be red with a pink tip, the mid section will be blue and the tail will be yellow-green. Epithelial cells by contrast will appear greenish to blue.

The more recent introduction of fluorescence microscopy has greatly streamlined the process of identifying spermatozoa[24, 25]. The strategy takes advantage of the specificity of proprietary fluorochrome-conjugated antibodies targeted to a protein that is unique to the spermatozoa head. The approach, which requires a fluorescence microscope

fitted with appropriate excitation and emission filters, makes it possible to selectively and readily visualize human sperm heads even against a background crowded with epithelial cells, microbes and non-cellular debris such as fecal matter. The spermatozoa heads appear as bright fluorescent ovoid bodies against a dark background.

The major limitations with microscopy are the need for relatively intact spermatozoa and the laborious nature of visually scanning microscopic fields for the presence of relatively few spermatozoa where non-sperm cells and other debris often predominate. Spermatozoa may be difficult to separate from cotton and other clothing fibers since as they degrade they typically lose their characteristics tails and ovoid shape such that they become more difficult to distinguish in the presence of similarly-sized microbes such as yeast. While fluorescence microscopy circumvents many of these problems, it requires a greater investment in costly microscopes and, because it is antibody based, may fail to yield optimal results as spermatozoa degrade and antigenic target molecules become unrecognizable by the antibodies.

### 1.3 Current Testing Methodologies

Seminal fluid, saliva, and blood are some of the most commonly encountered forms of serological evidence in criminal investigations. While not as common, vaginal fluid, anal secretions, fecal matter, urine, sweat, tears and nasal secretions could also constitute potentially probative evidence depending on the context of the case. Unfortunately, the range of body fluids for which serological assays would be valuable exceeds the number

for which assays actually exist. Currently, there are no reliable tests for vaginal fluids, anal secretions, sweat, tears or nasal secretions. However, there are numerous assay options available for several body fluids with forensic relevance which offer a wide range of strengths and weaknesses.

### 1.3.1 Semen and Seminal Fluid

Semen is a mixture of spermatozoa (sperm cells) and seminal plasma. Seminal plasma is a mixture of secretions from six different glands/tissues of the male genital system. Semen is ejected as a heterogeneous mixture of these secretions during ejaculation. The initial ejaculate (5% of the total volume) consists of secretions produced by the Cowper (bulbourethral) and Littre glands. The second portion of the ejaculate (15% to 30% of total volume) consists of secretions from the prostate gland. The secretions of the ampulla and epididymis then provide a small contribution while the seminal vesicles account for the balance (approximately 60% of total volume)[26-28]. Spermatozoa (sperm cells) make up a small portion of whole semen (1% to 5% of the total ejaculate volume)[29]. Pre-ejaculation fluid is secreted almost exclusively by the Cowper's gland and can contain traces of SAP and p30 but not semenogelin (Sg). The presence of sperm in pre-ejaculation fluid appears to be rare and may be due to carryover from a prior ejaculatory event. Because of the obvious association with sexual activity, seminal fluid and spermatozoa represent some of the most potentially probative evidence that can be collected as part of a sexual assault investigation.

Chemical reaction-based tests are no longer commonly used for the detection of seminal fluid due to their lack of specificity and reproducibility. Historically, however,

the chemical reaction-based tests for seminal fluid were the Barberio and Florence crystal tests.

Enzyme activity based tests such as the Brentamine test for seminal fluid target SAP activity, which is an abundant, but not unique, component of seminal fluid. SAP typically retains activity in a vaginal environment for approximately 48 hours post-coitus, but in an oral or rectal context, this period is typically shorter. A positive SAP result is indicted by a color change in the presence of a coupling agent and an indicator dye (*e.g.*, α-naphthyl phosphate and Brentamine Fast Blue). The test may be performed as a spot test, commercially prepared test strips or as an overlay technique to localize possible seminal fluid stains. Although SAP is a sensitive test, positive results cannot confirm the presence of seminal fluid. This is due to the fact that although SAP activity levels are elevated in seminal fluid, phosphatase activity in other fluids or materials can produce positive results. These include a number of non-seminal fluid associated substances such as vaginal fluid, fecal matter, plant material, spermicides and some feminine hygiene products.

Historically, immunoelectrophoresis, Ouchterlony double diffusion and Enzyme-Linked Immunosorbent Assays (ELISA) were employed to detect antibody binding. In recent years, these laborious and slow techniques have been supplanted by lateral flow immunochromatographic strip tests that are faster and commercially available at a relatively low cost. For seminal fluid detection, p30 and Sg are the primary targets of immunochromatographic test strips used in sexual assault investigations. The p30 protein is secreted by the prostate and is present in seminal fluid in a concentration range

between 70 to 5500 mg/mL with the majority of samples being between 820-1290 mg/mL. Semenogelin (Sg) is produced by the seminal vesicles and serves as the substrate for PSA and is the predominant protein in seminal fluid. In living individuals, the ability to detect PSA and Sg tends to be lost within 24 hours in a vaginal context and even faster in an oral or rectal context. Although antibodies targeted to PSA and Sg are generally sensitive and highly specific for human proteins, positive results must still be interpreted with caution and cannot confirm the presence of seminal fluid. There are two reasons: first, while PSA and Sg are abundant proteins in seminal fluid, they can also be found in a variety other bodily fluids and tissues, although not typically at concentrations that overlap significantly with the concentration range seen in seminal fluid. In the case of p30, however, positive results may be obtained with post-ejaculate urine; urine from adult males and vaginal fluids (**Figure 5**).

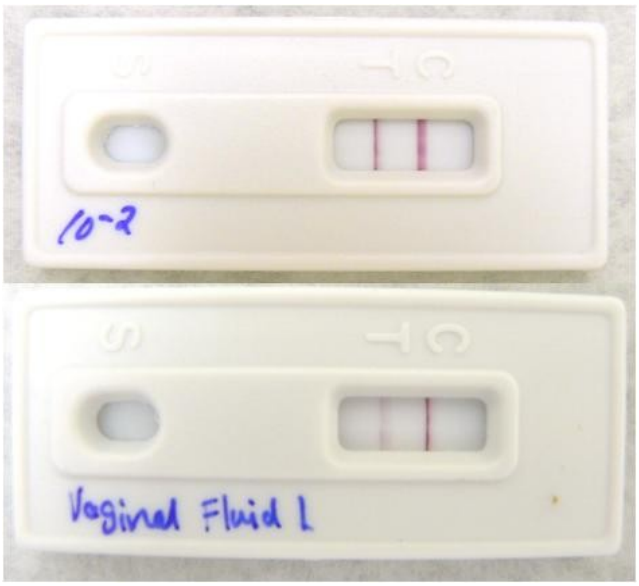



**Figure 5 - Abacus Diagnostics ABAcard® p30 immunochromatographic test strips. TOP: Visible bands at the Test Line (T) and the Control Line (C) indicates the detection of prostate specific antigen in a semen sample. BOTTOM: Visible bands at the Test Line (T) and the Control Line (C) indicate the detection of prostate specific antigen in a pre-coital vaginal fluid sample (*i.e.*, in the absence of semen). Image courtesy of Jillian Fesolovich, NMS Labs, Willow Grove, PA.**

Similarly, positive results with Sg may be obtained with post-ejaculate urine. The second reason is that, by their very nature, antibodies are subject to cross-reactivity with non-target antigens that possess structurally similar epitopes but which may be completely unrelated to the intended target.

Microscopy-based tests provide the most reliable means of indicating the presence of semen as they are based upon the direct identification of spermatozoa. On average, male ejaculate contains 3.5 mL of seminal fluid containing 10-50 million spermatozoa per mL. Human spermatozoa have a distinct morphology allowing species differentiation and can be identified for a longer postcoital interval (approximately 72 hours) when compared to chemical or protein components of semen. There are a variety of histological stains used for light microscopy of spermatozoa. Using methylene blue and acid fuchsin

(Baecchi's stain) the heads of the sperm stain red, while the midpieces and tails are blue. The use of hematoxylin and eosin (H&E stain) yields heads characterized by an intense blue-colored anterior spike and base-plate and a paler blue nucleus with a pink tail. Most commonly used is the aforementioned combination of nuclear-fast red and picroindigocarmine (Christmas tree stain) with which the heads are stained red and tails are yellow-green. The major drawback to light microscopy is the specificity of the dyes themselves. Regardless of the stain used, epithelial cells, microbes, bacteria and cotton fibers will also take up stain making the process of identifying the relatively small spermatozoa laborious and difficult. The use of fluorescence microscopy circumvents this drawback by allowing the selective visualization of spermatozoa through binding of commercially-available, sperm-specific antibodies conjugated to a fluorochrome. By combining the use of antibodies with microscopy, this approach also circumvents the possibility that antibody cross-reactivity might lead to an undetectable false positive result.

When using microscopy in a forensic context, a failure to observe spermatozoa must be interpreted with caution. Some males have abnormally low sperm counts (oligospermia) or may not produce spermatozoa at all (aspermia). Spermatozoa will be absent in the seminal fluid of vasectomized males approximately 2-4 months following surgery. The presence of spermatozoa in seminal fluid can also vary with the age of the male, and can be impacted by a broad range of factors including disease, genetics, radiation exposure, environmental toxins, undescended testis, varicocele, trauma, diet, tobacco use and/or illicit drug consumption.

### 1.3.2 Vaginal Fluid

Similarly to semen and its connection with sexual activity, the ability to detect vaginal fluid on evidentiary material has enormous potential to be probative as part of a sexual assault investigation. However, in contrast to current testing methods for semen, efforts to develop reliable tests for evidence of vaginal fluids have been largely unsuccessful.

Over the years, there have been numerous attempts to reliably identify vaginal epithelial cells in evidentiary samples. While the use of histochemical stains to detect sperm cells[23] is routine, staining to differentiate epithelial cells types (*e.g.,* skin, buccal and vaginal cells) has not been as successful. In the 1960s, the iodine-based Lugol's test held promise for identifying vaginal cells[30, 31]. This was based on studies suggesting that vaginal cells contained more glycogen than other epithelial cells. Unfortunately, Lugol-positive cells are also present in the male urethra[32], in male urine deposits[33], on >50% of penile swabs from males who had abstained from sex for several days and in the oral mucosa[34]. More recently, a methanol fixation protocol and Dane's staining technique (stains targeted to pre-keratin, keratin and mucin) was able to differentiate pure samples of vaginal, buccal and skin cells[35]. This approach, however, was unable to distinguish between a pure buccal cell sample and a mixed preparation of vaginal and skin cells. Given that forensic samples often contain such cell mixtures, this essentially negates the potential forensic utility of this approach as even a moderately reliable means of identifying vaginal epithelial cells.

### 1.3.3 Saliva

Saliva is a fluid produced by the major (90%) and minor (10%) salivary glands. The fluid is made up of 99% water but also contains a complex mixture of low-abundance proteins. Saliva provides lubrication, aids in digestion, maintains pH, assists with tooth enamel re-mineralization and plays a role in innate immune defense. The protein component of saliva is rich in the enzyme α-amylase which breaks down complex carbohydrates into smaller sugar molecules. As such, α-amylase activity or the ability to detect it through antibody binding serves as the basis of most serological saliva assays.

Enzyme activity tests for α-amylase commonly use the starch-iodine radial diffusion test. Starch in the gel is incubated with suspected saliva samples. Iodine is then added which stains starch a dark bluish-black. The passive radial diffusion of α-amylase into the gel and the subsequent breakdown of starch molecules produces an unstained "halo effect" around the test well that is characteristic of a positive result. There is a semi logarithmic correlation between the size of the halo and the amount of α-amylase loaded into the well[36]. Another approach is based on the use of degradable starch which contains a chemically-bound blue dye. Degradation of the starch due to α-amylase activity releases the blue dye to produce a color change that is visible to naked eye. The test may be performed as a spot test; using commercially prepared test strips or; as an overlay technique which may help to localize possible saliva stains on an article of clothing or other evidentiary material.

As with other fluids, immunochromatographic tests have improved the presumptive detection of saliva via α-amylase specific antibodies. Similar to seminal

fluid assays, lateral flow immunochromatographic strip tests are employed and a positive test result is indicated by the appearance of visible lines at the “test” and “control” positions on the assay membrane.

Regardless of whether the detection of α-amylase is based on enzyme activity or antibody binding, a positive result can only be considered presumptive. The α-amylase protein has been well conserved in evolution, and as a result, enzyme activity assays will yield positive results regardless of whether the amylase is from bacteria, fungi, or domestic pets. Similarly, it is not possible on the basis of enzyme activity to distinguish between variants of α-amylase (*i.e.*, salivary and pancreatic α-amylase). α-Amylase is also found in many body fluids other than saliva at levels that yield positive results with both enzyme activity and antibody-binding tests. These include breast milk, vaginal fluid, fecal matter, urine, blood and semen. Finally, cross-reactivity between antibodies and non-target antigens that possess structurally similar epitopes is also a possibility.

### 1.3.4 Blood

One of the most commonly encountered body fluids in criminal investigations is blood. Blood includes a cellular fraction consisting of erythrocytes, leukocytes and platelets (45%) suspended in a liquid fraction consisting of blood plasma (55%). Plasma contains 92% water along with a complex mixture of proteins, glucose, mineral ions, hormones as well as lipids. The metalloprotein hemoglobin is responsible for oxygen transport and is the major protein contained within erythrocytes. Because of its high abundance, the majority of blood detection strategies target this specific protein.

Among the many chemical reaction-based presumptive tests for blood are two crystal tests, the Takayama and aforementioned Teichmann tests, which rely on the heme group ferrous iron-catalyzed formation of crystals of heme derivatives. A positive result with the Takayama test is based on the formation of salmon-pink feathery crystals of pyridine hemochromogen resulting from the reaction of the ferriprotoporphyrin with pyridine. The Teichmann test positive result is indicated by the detection of brownish-yellow rhombic crystals of chlorohemin formed through the combination of a halogen with ferriprotoporphyrin. While crystal tests have not been found to yield positive results with body fluids other than blood, they are only presumptive for human blood for two reasons. The first is that a positive result will be obtained with both human and non-human blood. The second reason is that crystal formation only reflects a chemical reaction with the ferriprotoporphyrin group of hemoglobin. Thus false positives are technically possible with non-blood substances since iron porphyrin-containing hemoproteins such as cytochrome P450 and a number of other cytochromes occur frequently in biological systems[37]. As a practical consideration, however, these proteins are typically found in natural systems at concentrations well below the detection sensitivity of the crystal tests.

A number of other chemical reaction-based assays widely employed are based on reactions involving the iron-containing heme group of hemoglobin. This protoporphyrin IX ring structure with its associated iron ion displays peroxidase-like activity that can be harnessed to produce chemoluminescence or colorimetric changes in any of several indicator dyes. The Luminol test is most sensitive of these assays. Luminol and its contemporary derivatives are chemoluminescent substrates, the oxidation of which is

catalyzed by heme in the presence of hydrogen peroxide to form an unstable intermediate. Decay of this intermediate, liberates light[38]. Similarly, the aforementioned Kastle-Meyer (KM) and the Leucomalachite Green (LMG) tests rely on the peroxidase-like activity of heme to catalyze the reduction of hydrogen peroxide. This reaction liberates oxygen which then oxidizes either phenolphthalein to produce a pink color in the case of the KM test or the triphenylmethane dye leucomalachite green to produce a blue/green color which indicates a positive LMG result. Some commercially available test strips employ filter paper that has been pre-treated with diisopropylbenzene dihydroperoxide and the indicator dye tetramethylbenzidine (*i.e.,* Hemastix®). The peroxidase-like activity of the heme group catalyzes a change from yellow to green as a presumptive positive for the detection of blood. Positive results with these peroxidase reaction-based tests should be viewed as indicating peroxidase-like activity and by extension the presumed presence of blood. False positive results have been obtained with non-blood hemoprotein containing substances such as potato and horseradish extract as well as chemical oxidizing agents such as bleach, rust, ferric sulphate and copper-salts (*e.g.*, cupric sulphate which is an aquatic algaecide commonly used in aquaculture)[39, 40].

As with other body fluids, immunochromatographic tests have greatly improved the presumptive detection of blood. Hemoglobin has historically been targeted as the protein of choice for antibody-based tests with glycophorin-A being implemented more recently by the RSID™ Blood kit[41]. Glycophorin-A is the primary membrane protein located in red blood cells and is involved with cell-cell binding interactions[42]. The specificity of the assay is such that it is even capable of distinguishing human from primate

glycophorin-A. As with the commercially available antibody-based tests for seminal fluid and α-amylase, lateral flow immunochromatographic strip tests are employed; a positive test result is indicated by the appearance of visible lines at the "test" and "control" positions on the assay membrane. Although antibodies targeted specifically to human hemoglobin or glycophorin-A are generally sensitive and specific for human and often primate proteins, positive results should still be interpreted with caution. Antibody cross-reactivity to non-human blood is possible. For example, one antibody binding-based assay which specifically targets human hemoglobin has also been found to yield a positive result with ferret blood[43]. In addition, the potential for binding to unrelated proteins with structurally similar epitopes is difficult or impossible to predict *a priori* and cannot be eliminated as a possibility.

### 1.3.5 Menstrual Blood

Menstrual blood is present between 3 to 7 days during an average 28-day female menstrual cycle. Menstrual blood has not typically been considered a critical fluid with regards to forensic serology, particularly in comparison to the identification of semen and/or vaginal fluid. The reality, however, is that the ability to identify menstrual blood components also has the potential to provide critical evidence as part of a sexual assault investigation.

Initial work on the identification of this fluid attempted to utilize the polymorphic isoenzyme Lactate Dehydrogenase (LDH)[44]. As with Hemoglobin, LDH quaternary structure consists of a tetramer of two subunits, M and H. Unlike Hemoglobin, LDH can exist as five separate isoforms; LDH-1 with 4H subunits, LDH-2 or 3H1M, LDH-3 or

2H2M, LDH-4 or 1H3M, and LDH-5 or 5M. Serological differentiation relies upon the specific localization of each isozyme. Peripheral blood contains LDH-1, LDH-2, and LDH-3 where menstrual blood also contains LDH-4 and LDH-5[3]. Similar to other protein-based serological markers, LDH can be also be found in non-target tissues, thereby limiting its value as a specific test for menstrual blood[45].

### 1.3.6 Urine

Urine contains a variety of inorganic ions including chloride, phosphate, and sulfate which are not specific to urine but which are found at concentrations that are significantly higher in urine than other body fluids[37]. In addition, urine contains a variety of peptides and amines which can serve as presumptive indicators for the presence of urine including urea, uric acid, and creatinine. A number of chemical reaction- and enzyme activity-based tests have been developed to identify possible urine stains. Two that are in more common use are tests for urea and creatinine. Both tests provide a presumptive indication of the presence of urine since both urea and creatinine are found at detectable concentrations in other bodily fluids including sweat and seminal fluid[46].

Tests for urea are based on the detection of exogenously added urease enzyme activity which breaks down urea to liberate ammonia and carbon dioxide. The production of ammonia is then detected by Nessler's reagent (mercuric iodide in potassium iodide) which turns yellow or *p*-dimethylaminocinnamaldehyde (DMAC) which produces a strong red color. One of the first commercially available test strips based on this reaction employ paper impregnated with urease and bromothymol blue as a pH indicator to measure the formation of ammonia hydroxide. This same chemical system can be used to

locate suspected urine stains by spraying a combination of urease and bromothymol blue on an item of evidence. The presence of urine is indicated by the appearance of a blue spot[47].

Tests for creatinine include the Jaffe and Salkowski tests. The former mixes a solution of picric acid in toluene or benzene with an extract of a suspected urine stain. A positive reaction is indicated by the formation of red creatinine picrate. The latter uses sodium nitroprusside so that a positive indication of creatinine is indicated by formation of a blue product upon heating. Commercially available test strips based on the detection of creatinine employ a proprietary indicator system which produces a purplish-brown color in the presence of creatinine.

Antibody-binding-based assays for urine currently provide the only specific means of identifying urine. A commercially available lateral flow immunochromatographic strip test is based on the use of antibodies that are targeted to the Tamm-Horsfall Protein (THP) (*i.e.*, uromodulin) which is the most abundant protein present in urine. It is secreted into urine at a rate of 80-200 mg/day by the thick ascending limb of the loop of Henle[48]. As with other immunochromatographic strip tests, a positive test result is indicated by the appearance of visible lines at the "test" and "control" positions on the assay membrane. A positive test result, however, should be considered presumptive for the identification of human urine. As currently formulated, the test does not display cross-reactivity with human saliva, semen, whole blood, vaginal fluid, or menstrual blood but it does show cross-reactivity with urine from non-human samples.

### 1.3.7 Fecal Matter

Fecal material is most often identified by testing for the presence of urobilin which is formed as the terminal degradation product of heme in the enterohepatic urobilinogen cycle. Urobilin is responsible for the characteristic color of urine and fecal matter. In the large intestine, urobilinogen (a precursor of urobilin) is present at high levels and can be readily detected using the Edelman test which oxidizes the urobilinogen to urobilin by treatment with alcoholic mercuric chloride. This is followed by the addition of alcoholic zinc chloride to yield an urobilin-zinc salt compound which fluoresces a bright apple green under UV light. A positive result is not human specific and is only presumptive for fecal matter since urobilin is also present in urine, albeit at lower concentrations.

Light microscopy can also be used for the identification of fecal matter. This approach requires a skilled analyst familiar with the appearance of various enteric microbes as well as undigested fibrous particles of food which may include both plant and animal products. Although many microbial species will be present in both human and non-human fecal matter, a careful analysis of partially decomposed food debris may provide an indication of whether the material is consistent with a human diet.

## 1.4 Emerging Techniques

While serological tests for blood, semen, saliva and urine have long existed[19, 49-51], many of these are laborious, some pose health and safety risks, others consume significant amounts of a valuable evidentiary material, others fail to provide adequate

sensitivity or specificity. Additionally, there are some fluids (*i.e.* vaginal and menstrual fluids) which lack presumptive or confirmatory tests entirely. A number of promising methodologies have been in development to resolve the limitations of the current techniques. Emerging proteomic, transcriptomic, epigenomic and other technologies are being explored as a way to rapidly identify human body fluids with greater specificity and sensitivity than that of assays currently in use.

### 1.4.1 mRNA/miRNA Biomarkers

Messenger ribonucleic acid (mRNA) acts an intermediary molecule for the transfer of genetic information stored in DNA to the cell's functional units, *i.e.*, proteins[52]. Significant research interest has been devoted to these molecules as a means of identifying bodily fluids on the basis of differential expression profiles[53].

Several research groups have been investigating the use of mRNA-based techniques for body fluid identification. In 2003, pioneering work identified markers for saliva by searching the Cancer Genome Anatomy Project for proteins expressed only in the salivary glands. Combining reverse transcription PCR, visualization through gel electrophoresis, and quantitation by RiboGreen®, several of these candidate markers, including histatin 3, statherin, and Proline-Rich Protein BstNI Subfamily (PRB1/2/3), were shown to be expressed in saliva but not blood or semen[54].

More recently, a comprehensive prototype method for identifying blood, saliva, seminal fluid, and menstrual fluid was successfully developed using reverse transcriptase PCR. This multiplex method paired mRNA genetic markers along with a housekeeping gene to normalize for differences in mRNA expression between different body fluids.

This allowed for more confident expression profiles to be developed for each bio fluid[55]. An in-depth study of RNA recovery under a variety of conditions also demonstrated that under optimal conditions mRNA remained stable and detectable in some casework-type samples even after 547 days[56].

The use of mRNA biomarkers for body fluid identification has the advantage of enabling forensic practitioners to employ familiar sample preparation methods that are fundamental to any work involving nucleic acids. Methods for the co-extraction of DNA and RNA from the same stain have been developed[57] so that both mRNA and STR profiles can be developed from the same extract thus facilitating both individualization and body fluid identification in a single step[58].

Micro RNAs (miRNA) are small, 21-25 base pair long, non-coding nucleotide RNA molecules which are involved in the regulation of gene expression[59]. Along with mRNA, these molecules have been explored as promising markers for the characterization of body fluids. An initial study on the forensic potential of these molecules, published in 2009, used a combination of database evaluation and empirical testing to identify miRNA candidates which showed differential expression among blood, semen, saliva, vaginal fluid, and menstrual fluid. The results of this work found that each body fluid (with the exception of semen) could be differentiated on the basis of a unique miRNA expression pattern. Further work to evaluate the reliability of miRNA biomarkers, examined twenty one different tissues (*e.g.*, cervix, esophagus *etc.*) and found no expression patterns matching those of the forensically relevant body fluids[60]. In a more recent study, however, in which a microarray strategy was used to identify

candidate miRNA markers for saliva, semen, vaginal and menstrual fluids and peripheral blood, miRNA biomarkers that differentiated peripheral blood and semen were identified. In this later study, the approach was not able to identify reliable markers for the detection of saliva, vaginal or menstrual fluids[61].

While this technology is in its infancy, the approach has the same key advantages as mRNA based-assays with respect to integration with the current forensic workflow for DNA. As an added advantage, the small size of miRNA molecules relative to mRNA makes them potentially more robust for the analysis of aged or degraded samples[62].

### 1.4.2 Epigenetic Markers

Epigenetics is the study of transmissible changes in DNA expression that are not related to modifications in the DNA sequence[63]. These changes are generally associated with DNA methylation-related gene silencing. DNA methylation typically targets cytosine residues and localizes at CpG islands in promoter regions upstream of protein coding genes[64]. What makes these sites significant is their capability for gene silencing[65]. Put simply, a gene with an unmethylated promoter region will freely transcribe genetic information whereas a methylated promoter region will tend to have its genetic transmission quenched[66].

The use of epigenetic biomarkers for body fluid identification, therefore, relies on detecting tissue-associated differences in DNA methylation patterns[63, 66]. As with RNA markers, this approach employs pattern analysis but has the advantage of making it possible to directly "query" the DNA in a sample to determine the tissue from which it originated. Initial studies of epigenetic markers demonstrated the potential utility of the

approach using semen, saliva and skin tissue[67, 68]. A commercially available assay uses methylation-specific PCR to target genetic loci which are differentially methylated between semen and non-semen samples. In stains containing semen, three semen specific methylated loci will preferentially amplify creating a strong signal which can be visualized on an electropherogram whereas non-semen samples will not amplify strongly at these loci thus generating a weaker signal[69, 70].

A more extensive study aimed at identifying methylation response ratios that could be used to differentiate between semen, blood, saliva and skin cells identified 38 different loci which showed differential methylation patterns[68]. Of those, 15 were selected for inclusion in a multiplex assay which successfully differentiated target stains that were assayed as either pure or mixed samples. Tests with aged casework-type samples demonstrated that semen and blood could be clearly distinguished from each other in 20-month old samples. As with other nucleic acid-based assays, this methodology can also be readily incorporated into the standard DNA workflow.

### 1.4.3 Raman Spectroscopy

Non-destructive strategies for body fluid identification are particularly attractive to forensic practitioners for whom conventional methods of serological testing can be unacceptably consumptive of what are often minimal samples. Raman spectroscopy is based on a spectroscopic technique which determines the identity of an unknown substance[71] by measuring the inelastic scattering of light produced from the chemical constituents within the sample. Every unique "substance", therefore should produce a unique signature of spectroscopic peaks upon which substances can be differentiated

from each other[72]. The technology is robust and sensitive such that little if any sample preparation is needed, and picogram quantities of a sample can be detected[72].

This technique has been applied to body fluids in order to identify unique "multidimensional spectroscopic signatures" for specific body fluids of interest. In an effort to accommodate sample heterogeneity, advanced statistical algorithms are also employed to search for the best match between an expected "spectroscopic signature" and that of a questioned sample. Encouraging results have been reported for the identification of single-source vaginal fluid, sweat, blood, and semen stains using this approach[43] [73]. Additional studies have suggested that Raman spectroscopy may be capable of differentiating menstrual fluid from peripheral blood[48] as well as being capable of differentiating between human, canine, and feline blood samples[64].

Because of its reliance on statistical pattern fitting, however, additional studies will be needed to ensure that the pattern-matching algorithms can accommodate more challenging mixed stains as well as those containing contaminants or that are degraded, *i.e.*, any forensic type sample that deviates substantially from the pristine reference "spectroscopic signature".

Because of its ability to analyze samples *in situ*, an added potential benefit of this technique would be the ability to screen evidentiary material at the crime scene, effectively eliminating the need in some cases to transport cumbersome items of evidence to the lab to screen for potentially probative material.

### 1.4.4 Proteomic Assays

Among the most promising of the emerging strategies for body fluid identification are those made possible by advances in proteomics. The proteome is the final product of genome expression and comprises all the proteins present in a cell at a particular time[74]. In contrast to traditional approaches in molecular biology where a single or small subset of proteins are analyzed, proteomic techniques make it possible to identify and monitor hundreds to thousands of proteins in a single experiment[75]. It is now possible, for example, to map entire proteomes using 2-D gel electrophoresis or multidimensional HPLC based techniques coupled to mass spectrometry to identify potentially useful biomarkers. Once identified, targeted-ion mass spectrometry can be used to facilitate the detection and quantitation of low abundance protein biomarkers of interest[76]. This has resulted in a wealth of new opportunities to develop protein-based assays for medical and forensic purposes.

The application of proteomics to the analysis of challenging samples in forensic serology has been the focus of research and development activities of multiple research groups[77]. By comparing the proteomes of different body fluids, researchers seek to discover protein biomarkers that are unique to a given body fluid and thus could serve as the basis for the development of confirmatory tests for the identification of such forensically relevant body fluids as saliva, seminal fluid, peripheral blood, vaginal and menstrual fluids, and urine.

One research group with the New York Office of the Chief Medical Examiner has reported the successful development of a multiplex assay for blood, saliva, and semen on

a MALDI-TOF/TOF tandem mass spectrometer[77, 78]. Using protein biomarkers identified through database mining as well as empirical research results, several front-end sample preparation protocols were evaluated. These include 2-dimensional fractionation by combining isoelectric-focusing and reverse-phase liquid chromatography (IEF-LC-MALDI); 1-dimensional fractionation by reverse-phase liquid chromatography (LC-MALDI) and; direct spotting of unfractionated samples followed by MALDI-TOF/TOF analysis. These approaches made it possible to successfully detect several body fluid protein biomarkers including cystatin-SA and α-amylase 1 as markers of saliva; semenogelin I/II and kallikrein 4 as markers of semen; and hemoglobin subunit α and β as markers of blood.

A major strength of the MALDI-TOF/TOF platform is the potential speed of analysis. Using the direct spotting method, a 96-well plate of samples can be processed in a matter of minutes. As with all mass spectrometry strategies, however, there is a relationship between the amount of pre-fractionation and the sensitivity of the assay. The trade-off is that fractionation steps are consumptive in terms of time and cost but achieve superior sensitivity. For example, the direct spotting approach was only able to detect semenogelin-1 and 2 in 0.1nl of semen, whereas the LC-MALDI approach detected several additional biomarkers in the seminal fluid panel. This relationship also carried over to the analysis of mixtures by MALDI-TOF/TOF. The more rigorous sample pre-fractionation achieved by the IEF-LC-MALDI strategy made it possible to detect salivary α-amylase in a 10:1 mixture of semen to saliva while the LC-MALDI and direct spotting approaches were not able to detect the minor saliva component. Overall, however, these

approaches made it possible to characterize single and mixed body fluid stains in the subnanoliter range and worked well with forensic type samples aged up to 20 months[79].

## 1.5 Proteins as Serological Candidates

One of the significant advantages of a protein biomarker approach is the tremendous diversity of potential targets that are made possible due to post-translational modification in different tissues. As a result, a single protein may be differentially modified by one's metabolism in two different body fluids, making it possible to develop highly specific assays in cases where epigenomic patterns or mRNA expression profiles might not differ. Another key advantage is the stability of many proteins under conditions that lead to the degradation of other molecules. Proteins are among the most long-lasting of all biological molecules having been routinely isolated from even ancient biological material[80, 81]. In a more forensically applicable study, a 99.5% decrease in mRNA levels was observed in post-mortem brain tissue while protein levels remained relatively constant[82]. Still, as is the case with all biological molecules, proteins do fragment and degrade over time. Assays based on the use of protein biomarkers, however, can be readily adapted to detect fragmented proteins. Thus, even partially degraded target biomarkers may still be reliably detected[83].

## 1.6 Proteomics

Proteomics is a large scale study of expressed proteins. Because of recent advancements in instrumentation and techniques, the field of proteomics has expanded such that research scientists have the ability to rapidly and reliably profile entire proteomes of nearly any tissue or fluid. The primary objective of most proteomic-based research projects has been to develop novel diagnostic tests for the clinical setting. These assays are based on the use of proteins as diagnostic or prognostic markers of disease states such that they serve as tools for the personalization of medical treatments[84]. While the identification of disease biomarkers is not the goal of a proteomics project focused on forensic serology, the same principals, methodologies, and instrumentation can be utilized in order to identify proteins that can act as biomarkers for the identification of a specific biological fluid. These methodologies follow a well established proteomics pipeline that begins with a discovery phase to experimentally derive a panel of candidate biomarkers followed by a validation phase to confirm their specificity across a larger sample population.

### 1.6.1 Proteomics Front End Technologies

One of the primary challenges of proteomics discovery research is the enormous number of different proteins that are present within most biological fluids. Human plasma, for example, is estimated to contain several thousand different proteins. Additionally, the difference in concentration between the most and least abundant proteins can vary across ten orders of magnitude[85]. As a consequence, high-abundance proteins (*e.g.*, serum albumin or immunoglobulins in blood) which are not highly specific for a single body fluid can easily mask the detection of less abundant proteins that are

more “diagnostically relevant” for forensic applications. To circumvent this, proteomes are fractionated to reduce complexity of analysis.

As stated above, the primary goal of proteomics discovery research is to identify protein biomarkers. In a medical diagnostics setting, biomarkers should be differentially expressed between two samples – typically one from a patient with a disease state of interest and the other from a healthy individual. Adapted to forensic serology, useful biomarkers should be uniquely expressed in a target body fluid. Regardless of whether a protein is being used for a clinical or forensic application both rely on the combination of qualitative and quantitative assessments of a proteome. In this vein, researchers have employed gel-based methods, metabolic and chemical labeling methods, label free methodologies, as well as targeted quantitative methods in order to identify promising biomarkers.

#### 1.6.1.1 Gel-Based Methods

Traditionally, proteome analysis has relied on two-dimensional polyacrylamide gel electrophoresis coupled with mass spectrometry. These methods have been successful in separating highly complex mixtures where protein spots in gels are compared, picked, and subsequently identified by mass spectrometry. However, these techniques have been shown to be highly variable from gel to gel making it difficult to distinguish between actual protein variability versus gel-to-gel experimental variation[86]. More recent improvements to this traditional approach have utilized two-dimensional difference gel electrophoresis (2D-DIGE). In this technique, two or more samples are differentially labeled with fluorescent tags, pooled, and run on a single gel. The proteins in each sample

are then differentially detectable by their fluorescent tags. Protein spots that are differentially expressed between samples are subsequently characterized through mass spectrometry[87]. While these techniques have improved the ability to detect differentially expressed proteins, there still are limitations that impede the accurate analysis of proteins with very high or low pI values, proteins with large transmembrane domains as well as limitations to proteins with extremely large or small molecular masses[88, 89].

#### 1.6.1.2 Labeling-Based Methods

Metabolic and chemical labeling strategies offer a different approach than in-gel methods. Stable Isotope Labeling with Amino acids in Cell culture (SILAC) is one option. In this method, cells are grown in cultures with differentially-labeled amino acids. One culture media will contain normal amino acids and the other will have heavy isotope labels, *e.g.* C-13 vs. C-12 or N-15 vs. N-14. The amino acids in culture will be incorporated into the protein. When run on the mass spectrometer, peptides will behave identically but will have a measurable mass shift associated with their unique isotopic makeup, allowing quantitation and detection of differential expression between two states[90]. This strategy has been successfully used in many cell culture based systems. For example. Nirmalan *et al.* successfully measured proteomic changes in the parasite *Plasmodium falciparum* with regard to its protein expression during development as well as responses to anti-malaria drugs[91]. However, a drawback in this approach is the requirement for the incorporation of labeled amino acids in cell culture. This leaves many human/clinical samples (*e.g.,* serum and tissues) incompatible with the assay.

Chemical labeling strategies exist as well. Two well known methods of chemical labeling are Isotope-Coded Affinity Tag or the ICAT system and Isotope Tagging for Relative and Absolute Quantitation or the ITRAQ system. Both methods use a chemical label that is incorporated into the protein. In the ICAT system a heavy or light label plus an affinity tag is incorporated into cysteine SH groups. The biotin affinity tag is used to isolate the ICAT peptides through affinity chromatography so that quantitative differences can be measured[92]. ITRAQ uses a different system with the tag attaching to the N-terminal of digested peptides. The tagging molecule consists of three groups, a reporter group, an isobaric tag, and a reactive group that will link to the peptides. Differentially labeled peptides behave identically but have different reporter ions that liberated during collision-induced dissociation. Quantitative information is achieved by examining reporter ion abundances [93]. Numerous challenges, however, are present in these workflows. ICAT, for instance, relies on cysteines to be present. Without the presence of the appropriate amino acid, the affinity isolation process fails resulting in missed tryptic peptides, which may be diagnostically useful for quantification. Another issue relevant to both methods is the addition of labels downstream in the methodology. This can be compromised by potential errors in the efficiency of any process before the addition of tags such as fractionation, depletion, or digestion[94]. In addition, these methods are typically expensive, rely on proprietary software, and aside from ITRAQ, are limited to two or three comparisons at a time[95].

#### 1.6.1.3 Label-Free Methods

Two major label-free methodologies exist. Both methods follow the same process for sample preparation, digestion, chromatographic fractionation, and analysis on a tandem mass spectrometer. Data are quantified either through spectral peak intensities or spectral counting after identification by database search tools[95]. Using spectral peak intensities, extracted ion chromatograms for each peptide are integrated and peak areas are used for quantitative comparisons. The spectral counting method compares the number of spectra identified for each peptide in an analytical run. Label-free methods have the advantage of eliminating laborious and potentially expensive labeling steps; the capability of running any sample type, without limitation on how many samples can be compared. Conversely, these techniques rely heavily on reproducible chromatography and numerous sample replicates to ensure accurate and precise quantitative measurements[96].

#### 1.6.1.4 Targeted Proteomics

The final methodology, targeted proteomics, utilizes a triple quadrupole (QQQ) mass spectrometer operating in Multiple Reaction Monitoring (MRM) mode[97]. This method employs tandem quadrupole mass analyzers in order to isolate specific precursor and product ions. Quantitation is achieved through incorporation of synthetic peptides labeled with stable isotopes. Both the synthetic peptides and the native digested peptide will behave identically in the instrumentation. Calculating the ratio between standard and native peptide allows quantitation of the peptides[98] . This method has become popular

due to the high multiplexing capability and high sensitivity that can be achieved when targeting known compounds[99].

### 1.6.1.5 High-performance Liquid Chromatography

Dependable and sensitive proteomics experiments are directly linked to separation technologies used to simplify samples before analysis[100] . One additional technique that has gained popularity is multidimensional high-performance liquid chromatography (2D-HPLC). This method has left gel-based methods behind in favor of precise HPLC-based separations. One specific instrument, the ProteoSep™ 2D-HPLC system (Eprogen, Downers Grove, IL; formerly the Beckman Coulter ProteomeLab™ PF2D), employs chromatofocusing and reverse-phase chromatography to fractionate the proteome.

**Figure 6 - Eprogen ProteoSep™ multidimensional high-performance liquid chromatography system. System uses two dimensions of separation (chromatofocusing and reverse-phase chromatography) to separate entire proteomes from a single injection.**

Using the ProteoSep™ system as seen in **Figure 6**, a chromatofocusing column first separates proteins based off their isoelectric point. More specifically, proteins, which are carrying a negative net charge, are loaded into a column at pH of 10 followed a gradient to pH 4. As the pH decreases, proteins will begin to elute off the column into a series of first dimension fractions on a 96-well fraction collector. Specifically, proteins

elute at their isoelectric point (*i.e.*, the pH at which the protein net charge is zero). Reverse-phase chromatography is then performed on each of the pH fractions from the first dimension. In this step, a stationary phase with octadecyl carbon chains, or C18, and mobile-phase gradient (organic solvent) elute proteins based on their hydrophobicity. Proteins are initially bound to the hydrophobic stationary phase until a sufficient concentration of organic solvent disrupts the interaction between the protein and the stationary C18 particles[101]. As proteins elute from the reverse-phase column they are collected in a series of fractions across eight 96-well plates. During the reverse-phase runs, a 214 nm UV detector monitors the eluent for the presence of proteins as they are released from the column. Reverse-phase runs are then digitally combined to generate a conceptual proteome map (**Figure 7**) in which all the proteins successfully fractionated by the system are represented by pH (*x*-axis) / hydrophobicity (*y*-axis) / quantity (*z*-axis) coordinates. The multidimensional maps can then be compared using suitable software programs to identify differential protein expression across multiple samples. Based on the user's scientific question, individual reverse-phase fractions can be recovered and tandem mass spectrometry performed to identify the specific proteins present in the fraction of interest.

Figure 7 - Example of the ProteoSep™ pH/hydrophobicity/quantity map (i.e., proteome map) obtained using a body fluid sample. Proteins, which are represented as colored bands across the image, are separated based on pH on the x-axis followed by hydrophobicity on the y-axis. Relative protein quantity is indicated by the color (red=high; blue=low).

### 1.6.2 Liquid Chromatography Tandem Mass Spectrometry for Proteomics

The most powerful tools in proteomics research are the mass spectrometry systems which allow scientists to identify, characterize, and quantitate proteins in complex biological matrices. These systems incorporate high-performance liquid chromatography to fractionate samples, an ion source to generate charged gas-phase molecules, and a mass spectrometer to analyze the resulting ions. Once the data have been acquired, software tools are employed to identify individual proteins by comparing the acquired data to a protein database.

#### 1.6.2.1 Reverse-Phase High-Performance Liquid Chromatography

Chromatography is a general term for a variety of techniques which are used to separate mixtures. All chromatographic techniques utilize a mobile phase as well as a stationary phase. A mixture is dissolved in the mobile phase and is then passed through a stationary phase. Every compound in a mixture will have a higher or lower affinity for the mobile and stationary phases, thus allowing complex mixtures to be separated.

The most popular chromatographic technique for protein analysis is reverse-phase high-performance liquid chromatography (HPLC). This technique uses a pumping system to deliver a gradient over a stationary phase. The mobile phase consists of a blended aqueous and organic buffer where the stationary phase consists of porous silica particles coated with carbon chains.

During separation, a protein mixture is loaded onto the stationary phase under high aqueous (*i.e.*, polar) conditions causing hydrophobic moieties on the protein/peptides to bind the non-polar solid-phase particles. A linear gradient with an increasing percentage of organic solvent is applied to the stationary phase causing the mobile phase to become increasingly non-polar – conditions under which hydrophilic proteins adsorb to the stationary phase poorly whereas hydrophobic proteins adhere strongly. As a result, hydrophilic proteins elute quickly where a higher percentage of organic solvent will be required to elute hydrophobic proteins from the column (**Figure 8**). Once eluted from the stationary phase, individual proteins/peptides are measured via a detector.

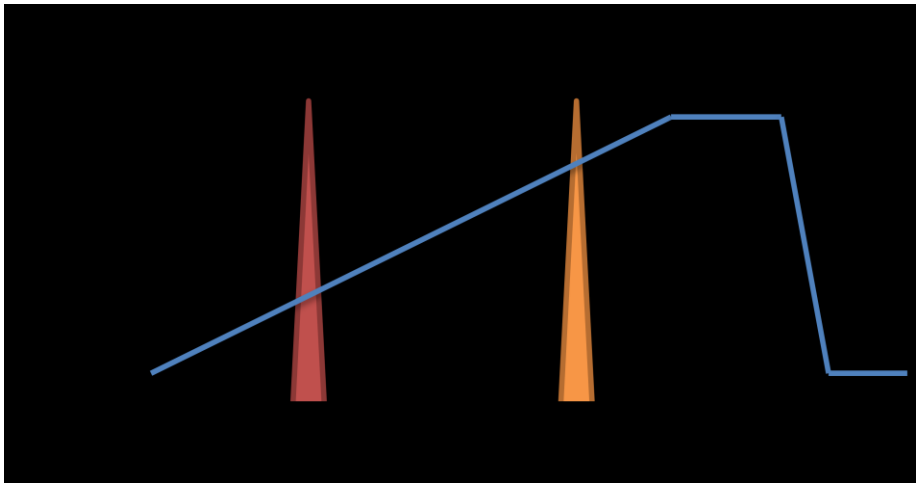


Figure 8 – Acetonitrile (buffer B) gradient used to elute peptides (orange and red peaks) using reverse-phase HPLC.

### 1.6.2.2 Electrospray ionization

Many modes of ionization exist for generating charged particles for analysis via mass spectrometry. Currently, peptide and protein analysis is restricted to 'soft' ionization techniques via Electrospray Ionization (ESI) or Matrix-Associated Laser Desorption Ionization (MALDI). The restriction to ESI and MALDI is due to the need to avoid fragmentation in the source from "harsh" ionization methods such as chemical or electron ionization.

When coupling an HPLC system to a mass spectrometer, an electrospray ionization source is commonly used as it allows molecules to become ionized directly from the liquid phase (**Figure 9**).

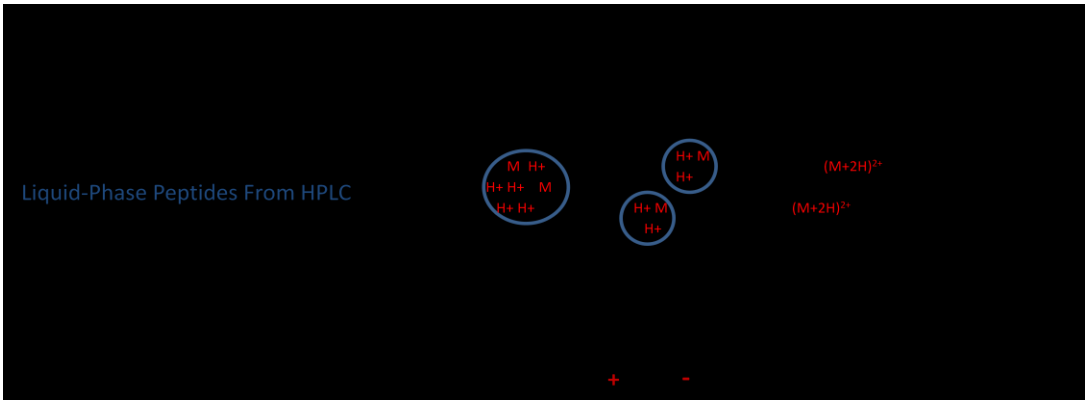



**Figure 9 - Electrospray ionization source. Peptides eluting into the source are charged, dried, and travel into the mass spectrometer inlet.**

In this process, the volatile organic solvent and acidic eluent (*i.e.*, acetonitrile with an acid modifier such as formic acid) from the HPLC is forced through a capillary under a high voltage potential (*i.e.*, ≥2000 Volts) between the spray needle and entrance to the mass spectrometer. This process generates positively charged droplets containing solvent, acid and analyte which will travel together towards the negatively charged MS inlet. Heated nitrogen drying gas (approximately 300°C) aids in evaporation of the solvent. As the solvent evaporates, the repulsive forces increase due to the excess charge contained in the droplets[102]. The repulsive force eventually becomes strong enough that a Coulomb fission occurs splitting the initial droplet into a number of smaller droplets. Repeated evaporation and fission events lead to protonated gas-phase molecules which enter the mass spectrometer[103].

Proteome analysis by LC-MS/MS represents one of the most difficult analytical challenges when trying to identify biomarkers in complex biological matrices. Sample quantities are typically limited and individual proteins in a matrix are generally low in

abundance. To circumvent these problems, nanoscale liquid chromatography is typically employed to increase the sensitivity of the LC-MS platform.

As the method implies, nanoscale liquid chromatography is simply a scaled-down version of the standard electrospray technique where column inner diameters and flow rates are significantly reduced. As a comparison, a standard HPLC column with a 4.6 mm inner diameter and a 150 mm length would have a column volume of approximately 2.5 mL. By contrast, a nanoscale column with a .075 mm inner diameter and a 150 mm length would have a column volume of only 700 nL. This ultimately leads to a much more concentrated sample on the column. In addition to the concentration advantage in nanoflow systems, low flow (400 nL/min *vs.* 400 μL/min) nanoscale electrospray generates smaller charged droplets compared to standard HPLC

**1.6.2.3 Mass Spectrometry**

Modern mass spectrometers consist one or more mass analyzers as well as a detector to record ion abundance[104]. More specifically, mass spectrometers measure compounds using mass-to-charge (m/z) ratios - the molecular mass divided by compound charge. For example, a peptide sequence Serine-Glycine-Alanine-Valine-Methionine-Lysine (**Figure 10**) under acidic LC conditions will protonate and enter the mass spectrometer with a +2 charge, one charge at each basic functional group. The m/z ratio for this compound would be computed by determining the mass of the compound 591 Da, adding two protons, and dividing by a charge state of 2 to give the following formula (591+2) / 2 = 296.5 m/z.

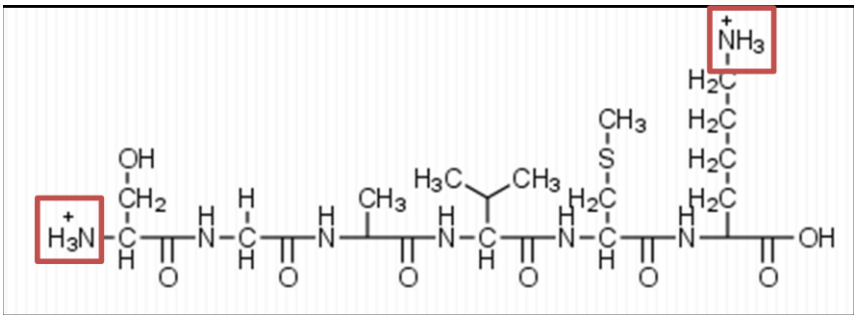


Figure 10 – Peptide with the sequence Serine-Glycine-Alanine-Valine-Methionine-Lysine. The m/z ratio for this doubly charged compound would be 296.5 m/z.

Various mass spectrometers are capable of performing proteomics experiments. Instruments are differentiated from each other by the mass analyzer(s) integrated into the system. Some of the most common mass analyzers are the quadrupole, ion trap, and time-of-flight tube[104, 105]. Modern proteomics instruments are capable of multiple steps of mass measurements. This process, also called tandem mass spectrometry, can generate a mass spectrum of both the intact ionized peptide (MS) as well as its fragmentation spectra (MS/MS). Some instruments such as an ion trap can perform MS and MS/MS using the same mass analyzer whereas other instruments require multiple analyzers working in tandem to achieve the same result.

**Ion Trap Mass Spectrometers** (**Figure 11**) utilize alternating electric fields to trap and eject ions. Using this system, peptides are trapped and subsequently ejected from an oscillating electric field produced by static direct and alternating radio frequency (RF) currents[106]. More specifically, a multi-stage tandem-in-time process is carried out where ions are first stored or "trapped" in orbitals specific to their m/z ratios. The RF amplitude is then increased which ejects peptides from the trap to a detector. Acquisition software examines the MS spectra and selects specific precursor ions for fragmentation. Ions not

selected for fragmentation are ejected from a new batch of ions in the trap. An increased RF signal is then applied at a specific resonance frequency to induce fragmentation of the isolated compound. Product ions are then ejected in the same manner as the initial MS scan, with the spectra representing product ions generated during fragmentation[102].

One notable feature of ion traps is the ability to perform additional MS scans on product ions ($MS^n$). $MS^n$ allows for a compound to be fragmented and the resulting fragments further isolated in the ion trap. This feature allows ion traps to provide greater information on the chemical structure of unknown compounds [107].

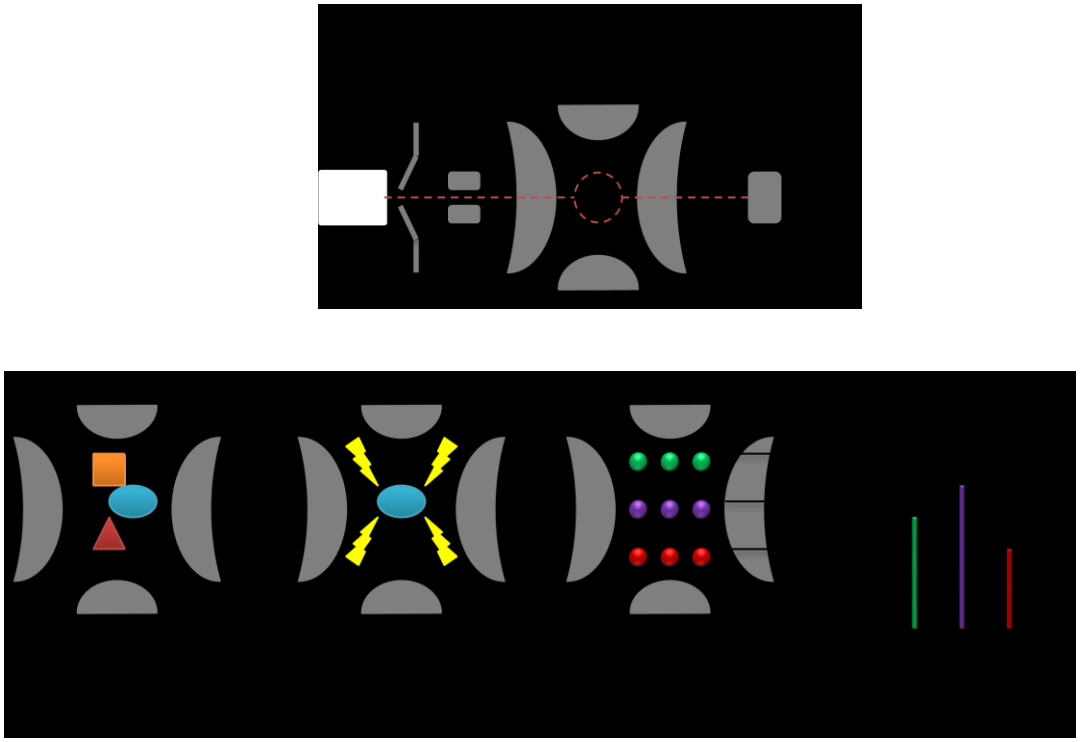
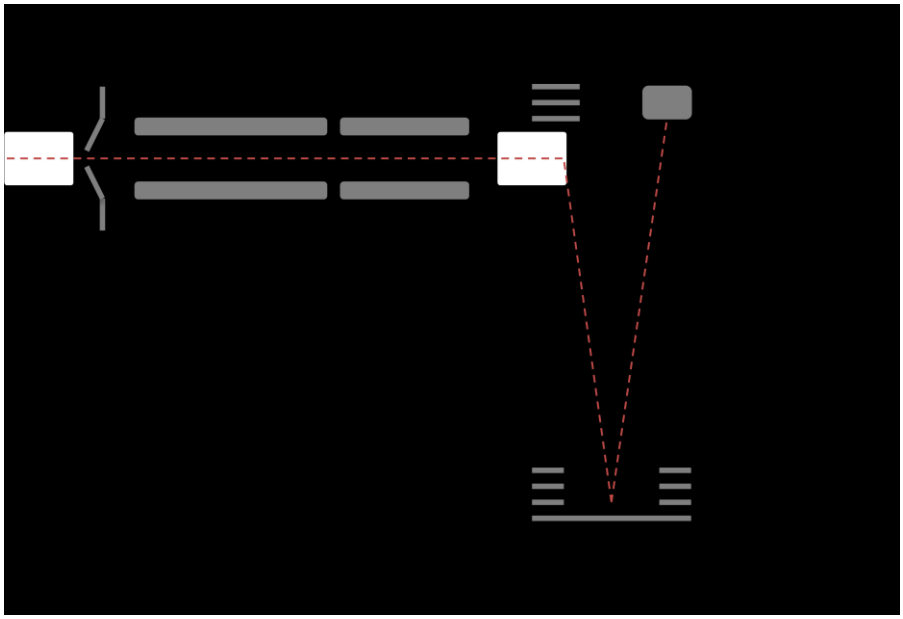


**Figure 11 Top – Schematic of an ion trap mass spectrometer. Red line indicates path of ions which are trapped in an electric field in the ion trap prior to ejection and detection. Bottom – MS/MS spectra generation through ion isolation, fragmentation, and ejection from the ion trap. In this example, the target ion (blue) is isolated in the ion trap, undergoes fragmentation, subsequent fragments (green, purple, red) are ejected to generate a spectrum.**

**Quadrupole Time-of-Flight** (QTOF) is a hybrid mass spectrometer which combines quadrupole and time of flight mass analyzers (**Figure 12**). Like the ion trap, these instruments are capable of acquiring tandem MS data but with the added functionality from the quadrupole mass filter and accurate mass TOF analyzer. In standard data-dependent acquisition mode, the instrument generates a survey scan of all eluting peptides (MS scan). Acquisition software then selects specific precursor ions from the MS scan and the quadrupole filters for only that specific m/z. The selected precursor ion is then fragmented in the collision cell and the product ions are accelerated and resolved in the TOF for detection. The precursor isolation, fragmentation, separation, and detection process produces the MS/MS fragmentation spectra for each peptide.

The advantages to the Q-TOF are twofold. First, the quadruple mass analyzer allows the system to isolate ions with specific m/z ratios allowing targeted acquisition. Additionally, the system can achieve high mass accuracy (typically <5 ppm) coupled with high resolving power. Together, these factors give the QTOF the advantage of being able to isolate and more accurately measure and resolve peptides when compared to a triple quadrupole or ion traps. This leads to more confident and qualitative protein identifications[108, 109].



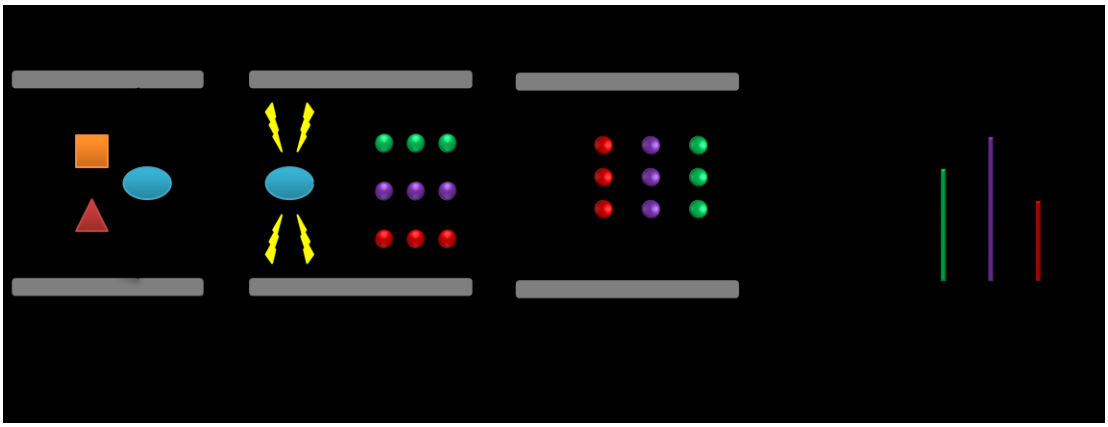


Figure 12 - Bottom – Schematic of a Quadrupole Time-of-Flight mass spectrometer. Red line indicates ion path which enters the instrument in the inlet, is isolated in the quadrupole mass filter (Q1), fragmented (q2), and resolved on the TOF before detection. Bottom – MS/MS spectra generation through ion isolation in quadrupole one (Q1), fragmentation in the collision cell (q2), followed by separation by the TOF mass analyzer. In this example, the target ion (blue) is isolated in the quadrupole, undergoes fragmentation in the collision cell, and the subsequent fragments (green, purple, red) are resolved in the TOF analyzer before detection to generate a spectra.

**Triple quadrupole** mass spectrometers (QQQ) have been described as the gold standard of modern quantitative analysis[110]. This class of instrument (**Figure 13**) is made up of two quadrupole mass filters as well as a collision cell between them. The gold standard classification is due to the sensitivity, selectivity, and ease of operation of these

instruments due to the ability to perform Selective Ion Monitoring (SRM) and Multiple Reaction Monitoring (MRM) experiments. In MRM mode, two separate selection steps are performed. First, quadrupole one (Q1) isolates a select precursor ion, precursor ions are selected and fragmented in the collision cell q2, and then specific product ions are selected in quadrupole three (Q3) and subsequently detected. While this instrument has much lower mass accuracy/resolution compared to the QTOF, the QQQ mass spectrometers in MRM mode achieve a greatly enhanced signal-to-noise ratio which ensures that target ions are detected with improved sensitivity and confidence even with low amounts of starting material[111].

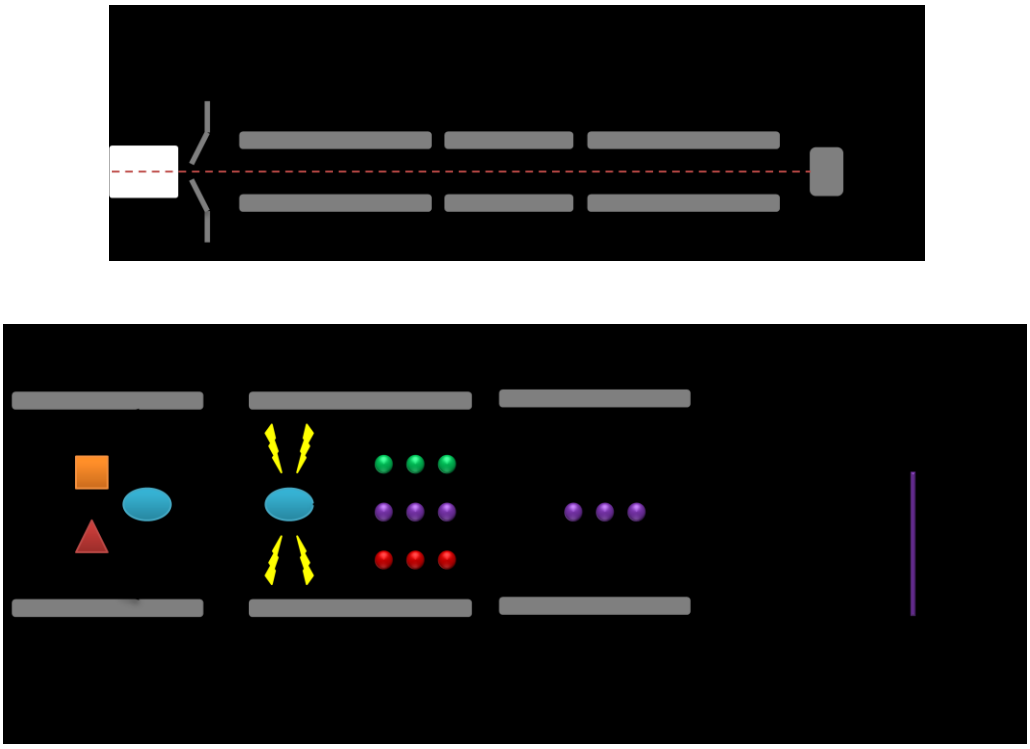


Figure 13 - Schematic of a triple quadrupole mass spectrometer. Red indicates the ion path starting from the inlet, isolation in quadrupole 1 (Q1), fragmentation in the collision cell (q2), and fragment isolation in (Q3) prior to detection. Bottom – MS/MS spectra generation through MRM mode with peptide isolation in quadrupole 1 (Q1), fragmentation in the collision cell (q2), product ion isolation in quadrupole 2 (Q3). In this example the instrument the target ion (blue) is isolated in the first quadrupole, undergoes fragmentation in the collision cell, then a specific fragment is isolated in quadrupole 3 (purple) before detection and spectra generation.

## 1.7 Research Objectives

In the current research study, proteome mapping and sensitive protein detection technologies have been applied to six forensically significant body fluids (*i.e.*, semen, saliva, urine, peripheral blood, vaginal and menstrual fluids) with the overarching goal of developing a multiplex body fluid assay on a mass spectrometry platform. To do this, a standard proteomic workflow was employed encompassing a biomarker discovery and verification phase followed by the development of a prototype assay for implementation in the environment of a standard caseworking laboratory.

The discovery phase of the research was designed to generate a panel of candidate protein biomarkers which were unique to the six target body fluids. Front-end fractionation was performed with the ProteoSep™ system in order to generate 2D-HPLC proteome maps. These were then analyzed using a novel, in house software package designed to flag potentially unique biomarkers. Unique, but unknown, proteins represented by specific pH/hydrophobicity/quantity coordinates were then identified by tandem mass spectrometry. This generated a database of candidate high-specificity protein biomarkers.

The biomarker verification phase was designed to evaluate the reliability and specificity of the candidate biomarkers. A multiplex, targeted, Q-TOF assay was designed to analyze samples of each body fluid collected from fifty unrelated individuals for each fluid. The results were evaluated for consistency of biomarker expression as well as specificity for the intended target body fluid. A small series of casework-type samples were also tested to evaluate the performance of the assay with casework-type samples.

The prototype validation phase examined the viability of a mass spectrometry-based serological assay for use in a higher-throughput caseworking laboratory. The most promising targets from the verification phase were then used to develop a rapid multiple reaction monitoring assay on a triple quadrupole mass spectrometer. Additionally, a series of casework-type samples were prepared to evaluate possible DNA/Protein co-extraction procedures.

## 1.8 Hypothesis

The general hypothesis being tested by this dissertation research is that a mass spectrometry based proteomics assay will provide an accurate and robust method for the simultaneous identification of semen, saliva, urine, peripheral blood, vaginal and menstrual fluids for use with evidentiary samples in a criminalistics lab. Furthermore, it is predicted that the assay will improve upon current methodologies by providing true confirmatory results for all the target fluids in a single multiplex assay. The specific core hypotheses of this dissertation research are therefore:

**(1)** Proteomic fractionation in combination with tandem mass spectrometry is capable of generating a database of body-fluid specific proteins for the identification of biological stains.

**(2)** A mass spectrometry-based assay can successfully designed to simultaneously screen samples from a larger sample population for six forensically relevant body fluids.

**(3)** Mass spectrometry is sufficiently robust for processing casework samples in a production lab environment.

**(4)** Mass spectrometry is compatible with existing DNA workflows as well as with the demands of a testing laboratory.

## 1.9 Dissertation Structure

Each chapter of the current dissertation contains a brief introduction, description of the experimental methods employed, an overview of the results obtained, as well as a discussion of the research findings. Chapter 2 focuses on the discovery of the panels of

candidate high-specificity biomarkers for each body fluid which were identified using three protein identification approaches. The focus of chapter 3 is on the development and testing of a multiplex assay on a Q-TOF which was designed to screen for six biological fluids simultaneously. Chapter 4 assesses the feasibility implementing a mass spectrometry-based body fluid assay in a casework lab environment. Chapter 4 will also include the results of experiments which examine the compatibility of the sample preparation protocols for this approach with existing DNA methodologies.

# Chapter 2:

# Biomarker Discovery

## 2 Introduction

The objective of this phase of the dissertation research was to develop a database of candidate biomarkers specific to individual biological fluids which includes saliva, semen, urine, peripheral blood, vaginal menstrual fluids. In order to ensure that at least two to three high-specificity biomarkers were successfully validated for each body fluid, a minimum of five to ten candidate biomarkers were to be identified for each fluid of interest. This was based on the expectation that some of the identified as promising candidate biomarkers would eventually be found to be expressed at unacceptably high levels in non-target tissues or would not be reliably detectible in the general population.

Traditionally, proteomics-based discovery research begins with whole proteome fractionation by 2D gel electrophoresis and manual excision/purification of proteins of potential interest for subsequent identification by mass spectrometry[112]. While these methods have a long history of successful use, they are not without limitations. For example, it is difficult to resolve proteins that are lipophilic, very large (>150 kDa), very small (<5 kDa) or less abundant. Also, poor reproducibility between gels often necessitates that numerous gels be run to obtain data sets that can be reliably compared with each other to identify proteins of interest[113].

More recently, multidimensional liquid chromatography platforms have been introduced to overcome some of the more significant limitations of traditional gel-based approaches[114]. The use of liquid-phase separation avoids many of the solubility problems associated with gels. Second, as a result of improved solubility liquid-phase separation ensures that a more complete proteomic profile of each body fluid can be generated because a greater range of proteins are fractionated and recovered. Third, less

abundant proteins in complex mixtures are more readily screened as potential markers of specific body fluids because there was a higher efficiency of recovery (>95%) of eluted proteins in liquid fractions and more total protein (up to 5 mg) can be injected initially onto the HPLC column without the band distortion that occurs with 2DGE.

While the limitations of 2DGE have been well documented, it is also likely that yet to be characterized limitations exist in the liquid chromatography based fractionation technique as well. To account for this possibility and to ensure a comprehensive view of the proteomes of each of the six target fluids, multiple separation approaches were employed to characterize each proteome. These included 1D fractionation, 2D fractionation, as well as a "whole" proteome analyses. Thus, a series of experiments were performed in order to populate a database of candidate body fluid-specific protein biomarkers. ProteoSep™ 2D-HPLC fractionation and tandem mass spectrometry served as the core technologies to fulfill these goals. First, a combination of chromatofocusing and reverse-phase fractionation (*i.e.*, 2D liquid-phase separation) was used to generate global proteome maps followed by cluster analysis and mass spectrometry analysis. Next, pH fractionation alone (*i.e.*, 1D liquid-phase separation) was performed followed by mass spectrometry analysis. Finally, a whole proteome approach using unfractionated samples was analyzed using accurate mass tandem mass spectrometry (**Figure 14**).

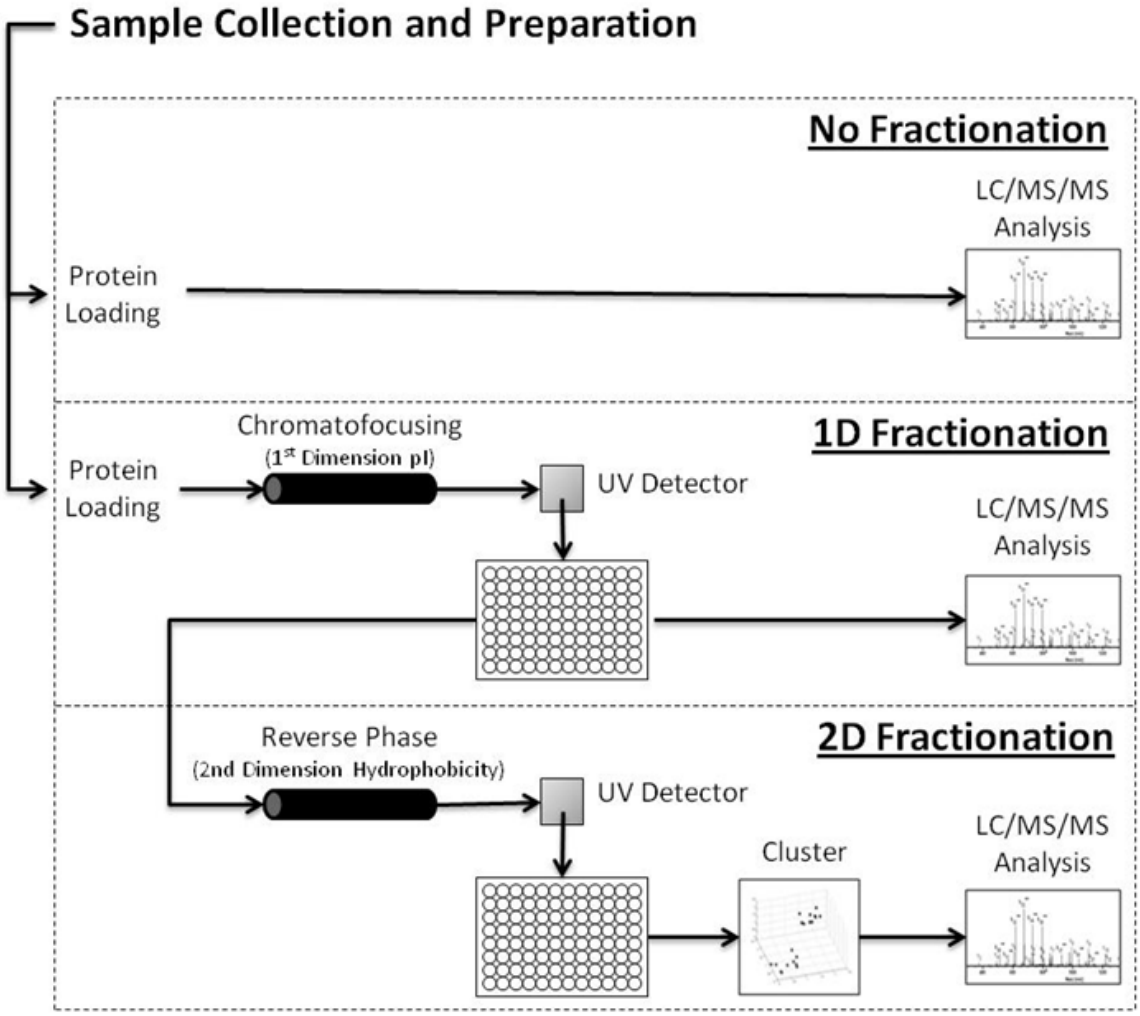



**Figure 14 - Three analytical strategies were employed for biomarker identification. These included non fractionated analysis of whole body fluids, analysis of 1D chromatofocusing fractions as well as cluster analysis of 2D chromatofocusing/reverse phase fractions**.

## 2.1 Materials and Methods

All research conducted under this project was reviewed, approved by the University of Denver Institutional Review Board (IRB) for research involving human subjects and conducted in full compliance with U.S. federal policy for the protection of human subjects. In total, 10 subjects (five males and five females) were recruited from

the undergraduate population at the University of Denver. Female donors were compensated $80 and males $20 for their participation in the research study. All research volunteers agreed and signed a letter of informed consent acknowledging they have received, read, and understood the protocols involved in the collection of samples. Additionally, all donated samples were given a random alphanumeric label to ensure confidentiality. Finally, males were instructed to refrain from ejaculation for 24 hours prior to sample collection and all donors were instructed to refrain from sexual intercourse for 5 days prior to the collection of urine, and vaginal fluids and menstrual blood for females.

### 2.1.1 Sample Collection and Preparation

An overview of the workflow performed from sample collection through biomarker identification is provided in **Figure 15**. Saliva, seminal fluid, urine, vaginal fluid, and menstrual blood were self collected where peripheral blood samples were collected at the University of Denver health clinic. After collection, saliva, seminal fluid, urine, vaginal fluid samples were desalted, concentrated, quaantified for total protein and stored at -80°C prior to 2D, 1D, or unfractionated analysis. All blood based samples underwent additional preparation steps via high abundant protein immunodepletion. Menstrual blood was further processed via hemoglobin removal.

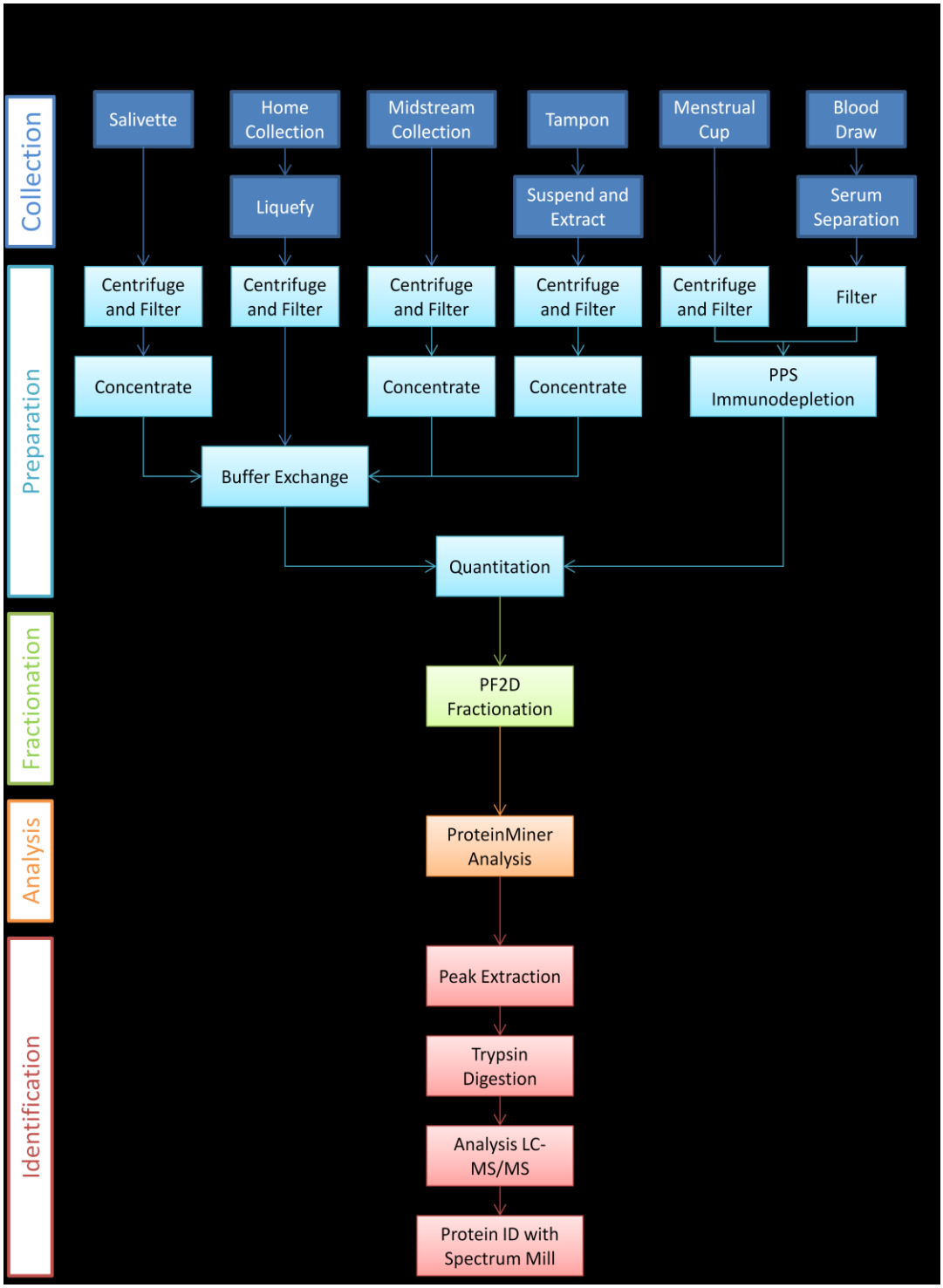



**Figure 15 - Overview of Biomarker discovery phase from sample collection to protein identification. Briefly, samples were collected, cellular debris was removed, followed by concentration and buffer exchange. Blood samples were processed through an immunodepletion column to remove high abundant proteins. Samples were then quantified for total protein, fractionated on the Pf2D system prior to mass spec analysis.**

#### 2.1.1.1 Saliva

Saliva samples were self-collected in the laboratory setting using Salivette™ saliva collection sponges (Sarstedt, Newton NC). Each donor was directed to thoroughly rinse their mouth with sterile water to remove residual food particles. After five minutes to allow secretion of saliva, the donor was instructed to place a Salivette™ saliva collection sponge into their mouth and to gently chew and roll the sponge around in their mouth for approximately three and a half minutes. The sponge was then placed into a sterile plastic conical tube. This process was then repeated five additional times to ensure that sufficient quantities of saliva were collected for proteome mapping.

Salivette™ sponges were centrifuged at 1500 x g for 2 minutes at 4°C to recover saliva. Individual Salivette™ sponge fractions were then pooled for each donor for a total volume between 5 and 10 mL. Pooled samples were centrifuged again at 10,000 x g for 20 minutes at 4°C to pellet cells and insoluble material. The supernatant was filtered through a 0.45 μm filter to remove remaining debris prior to concentration. Each sample was then concentrated at 4000 x g for 30 minutes at 4°C using an Amicon Ultra-15 (EMD Millipore, Billerica, MA) with a 5K nominal molecular weight cut off (NMWCO) centrifugal filter unit to concentrate the sample to a final volume of 2.5 mL.

#### 2.1.1.2 Seminal Fluid

Semen samples were self-collected by male donors in the privacy of their home. Donors were directed to refrain from sexual activity for a minimum of 24 hours and then to collect a 3-6 mL sample of seminal fluid by masturbation. The subject was instructed to directly deposit the fluid into a sterile plastic collection cup provided by the laboratory and then to refrigerate the sample until it could be transported to the lab within one hour.

The freshly collected semen was then incubated at room temperature for 30 minutes to allow it to liquefy. Liquefied samples were then transferred to a 15 mL conical vial and diluted with a 1/3 volume of phosphate buffered saline. The sample was then centrifuged at 10,000 x g for 20 minutes at 4°C to pellet spermatozoa. The protein-rich supernatant was then passed through 0.45 μm filter to ensure cellular removal.

#### 2.1.1.3 Urine

Urine samples were self-collected by donors in the privacy of their home. Donors were instructed to deposit a morning urine sample (>50mL) into a sterile collection cup provided by the laboratory. Protein concentration varied substantially between individuals thus >20 mL was typically used to ensure that a sufficient quantity of protein was recovered for accurate proteome mapping.

Freshly collected urine was centrifuged at 10,000 x g for 20 minutes at 4°C and passed through a .45 μm filter to ensure cellular removal. Two 15 mL volumes of urine were then concentrated at 4000 x g for 30 minutes at 4°C using an Amicon Ultra-15 with a 5K NMWCO centrifugal filter units to a final total sample volume of 2.5 mL.

#### 2.1.1.4 Vaginal Fluid

Vaginal fluid samples were self-collected by female donors in the privacy of their home. Donors were instructed to insert a commercially available 100% cotton tampon and were encouraged to use lubricant to minimize the risk of tissue abrasion and/or microbial infections. The tampon was left in place for the period of approximately 10 minutes, gently removed and placed in a 15mL conical tube. Donors were directed to refrigerate the sample until it could be transported to the lab within one hour.

Tampons were saturated with PBS and allowed to sit at room temperature for 30 minutes with occasional vortexing to elute proteins. The tampon was then placed in a 50 cc syringe to force out the fluids and eluted proteins. The liquid was transferred to 50 mL conical vials and centrifuge at 10,000 x g for 20 minutes at 4°C. The resulting supernatant was passed through a 0.45 μm filter to ensure cellular removal prior to concentration. The sample was then concentrated using an Amicon Ultra-15 with a 5K NMWCO centrifugal filter units at 4000 x g for 30 minutes at 4°C to a final total sample volume of 2.5 mL.

#### 2.1.1.5 Peripheral Blood

Donors were escorted to the Student Health Center where a 15mL sample of whole blood was obtained by a certified nurse using venipuncture. The blood was drawn into a sterile vacuum tube containing an anticoagulant (EDTA) and gel barrier to facilitate separation of the blood serum from cellular material.

EDTA gel barrier tubes were spun at 1500 x g for 10 minutes to separate the blood serum. The isolated serum was then passed through a 0.45 um filter unit to ensure cellular removal from the sample.

#### 2.1.1.6 Menstrual Fluid

Menstrual fluid was self-collected by female donors the privacy of their home. The collection protocol employed an FDA-approved over-the-counter latex-free, DivaCup™ (Diva International, Kitchener, Ontario) hypoallergenic cup for the collection of menstrual flow. The cup was then gently removed and the contents were poured into a sterile 50mL conical tube and refrigerated until delivered to lab within one hour.

Fresh menstrual fluid was transferred to a 15 mL conical vial and diluted with 1/3 total volume of phosphate buffered saline. The sample was homogenized by vortexing and added to an EDTA gel barrier blood collection tube which was then allowed to incubate for 15 minutes at room temperature. Tubes were then spun at 1500 x g for 10 minutes at 4°C to separate the fluid/serum component. Menstrual fluid/serum was then passed through a 0.45 μm filter unit to ensure cellular removal from the sample.

#### 2.1.1.7 Hemoglobin Removal from Menstrual Blood

HemogloBind™ (Biotech Support Group, Monmouth Junction, NJ) was used to remove hemoglobin from lysed erythrocytes commonly encountered in menstrual fluid samples. Effective hemoglobin removal was achieved by the addition of 500 μL of HemogloBind™ to 250 μL of menstrual fluid/serum. The solution was vortexed and then gently mixed for 15 minutes. The sample was then centrifuged at 9000 x g for 2 minutes at room temperature. The supernatant containing hemoglobin-free menstrual serum was recovered.

#### 2.1.1.8 Protein Partitioning System (PPS): Immunodepletion

Both menstrual fluid and peripheral blood were depleted of twelve non-specific high-abundance serum proteins which had the potential to mask the detection of lower-abundance proteins with greater body fluid specificity. These included albumin, IgG, IgA, IgM, fibrinogen, transferrin, HDL, haptoglobin, α1-antitrypsin, α1-acid glycoprotein and α2-macroglobin. For protein partitioning, 250 μL of sample was diluted with 375 μL of dilution buffer (10 mM Tris-HCL, 150 mM NaCl, pH 7.4). The partitioning column was first equilibrated for 30 minutes at a flow rate of 2 mL/min with dilution buffer. Proteins were loaded onto the column in a dilution buffer at a flow rate of 2 mL/min.

Unbound proteins in the flow through were collected in 50 mL conical tubes using a FC/I Module (Beckman-Coulter, Brea, CA). Bound high abundance proteins were then eluted with stripping buffer (100 mM glycine, pH 2.5) at a flow rate of 2 mL/min. Lastly, the column was equilibrated with dilution buffer at a flow rate of 2 mL/min This process was repeated with 7 total runs for each sample. Flow through from each individual sample was pooled and concentrated with an Amicon Ultra-15 with a 5K NMWCO centrifugal filter units at 4000 x g for 30 minutes at 4°C to concentrate samples to 1 mL total volume. ProteoSep™ (proprietary) start buffer was added for a total volume of 4 mL and concentrated again to exchange buffer contents. ProteoSep™ start buffer was added again to a final volume of 2 mL.

#### 2.1.1.9 Buffer exchange

Non-blood containing samples which had not been treated by the PPS immunodepletion column required a buffer exchange prior to injection into the ProteoSep™ system to equilibrate samples to the chromatofocusing starting pH. This was achieved through use PD-10 Desalting Columns (GE Healthcare, Buckinghamshire, UK) which were equilibrated with 25 mL of ProteoSep™ start buffer. A 2.5 mL protein sample was added to the PD-10 column and the flow through was discarded. The column-bound proteins were then eluted by the addition of 3.5 mL of ProteoSep™ start buffer to the column in a clean conical tube.

#### 2.1.1.10 Protein Quantitation by Micro BCA™ kit

The total protein content of the cell-free extract of each body fluid sample was quantitated by visible light spectrophotometry using the Micro BCA™ kit (Pierce Biotechnology, Rockford, IL) based on comparison to a standard curve generated for

Bovine Serum Albumin (BSA). A total of eight serial dilutions of BSA were prepared with concentrations ranging from 0.5 µg/mL to 200 µg/mL. The working stock reagent for the assay was prepared by combining 25 parts of proprietary "BCA Reagent A", 25 parts of proprietary "BCA Reagent B" and 1 part of proprietary "BCA Reagent C". In a 1.5 mL microcentrifuge tube 150 µl of working stock reagent was added to 150 µL of sample or standard. Tubes were vortexed, spun briefly, and allowed to incubate in a Thermomixer® (Eppendorf, Hauppauge, NY) at 37ºC for 2 hours. Samples were then allowed to cool to room temperature for 10 minutes and read on a visible spectrophotometer at 562 nm. The BSA standard curve was used to generate a linear equation from which the protein concentrations of the unknown samples were determined.

### 2.1.2 ProteoSep™ 2D-HPLC Fractionation

Fractionation using the ProteoSep™ 2D-HPLC system involved a two-step fractionation process – High-Performance Chromatofocusing (HPCF) followed by High-Performance Reverse-Phase (HPRP) chromatography. An HPCF column was used to fractionate protein mixtures based on the isoelectric points of the liquid-phase proteins. An in-line pH meter, linked to the instrument's automated fraction collector, controlled the output of the eluent to a 96-well plate from pH 8.5 to pH 4.0 in 0.1 pH increments. The first dimension HPCF column was equilibrated with 30 column volumes of ProteoSep™ start buffer at pH 8.5 at a flow rate of 0.2 mL/min for 130 minutes. Following equilibration, 5 mg of protein was injected into the HPCF module followed by 20 minutes of ProteoSep™ start buffer at a flow rate of 0.2 mL/min. At 20 minutes, eluent buffer (pH 4) was run at a flow rate of 0.2 mL/min for 115 minutes to create a pH

gradient after which 1 M NaCl was run as a high ionic strength salt wash. During the gradient phase, eluted fractions were collected at 0.1 pH intervals and stored in a chilled 96 deep-well autosampler plate. These first-dimension fractions were used both for direct analysis by mass spectrometry and for further second-dimension HPRP fractionation.

For liquid phase second-dimension fractionation, each fraction collected from the first dimension module was automatically and sequentially injected into an HPRP column where proteins were separated on the basis of their hydrophobicity. This second-dimension HPRP column was initially flushed with 5 column volumes of 0.08% trifluroacetic acid (TFA) in acetonitrile followed by 10 column volumes of 0.1% TFA in $H_2O$ running at 0.75 mL/min. Sample proteins were bound with 2 min of 0.1% TFA in water at a flow rate of 0.75 mL/min. At 2 minutes, a 0-100% gradient of 0.08% TFA in acetonitrile was run over a 30-minute interval creating a 3.33% change in solvent concentration/minute. At 0.5 minute intervals, fractions were collected with an FC-204 fraction collector (Gilson, Middleton, WI) into a series of 96 deep-well plates. The collected fractions which contained intact proteins (approximately 400 fractions) were stored in a locked -70°C freezer until required for further characterization.

### 2.1.3 Protein Identification by Mass Spectrometry

First-dimension protein fractions, second-dimension protein fractions and unfractionated total protein extracts from each body fluid of interest were analyzed by mass spectrometry to generate a comprehensive series of proteomes for each body fluid. In preparation for mass spectrometry, first-dimension fractions were concentrated and

desalted and second-dimension fractions and 30 µg of unfractionated total protein were both lyophilized.

#### 2.1.3.1 pH Fraction Desalting

Direct analysis of ProteoSep™ pH fractions (*i.e.*, first dimension fractions) by mass spectrometry required that samples be desalted using Spin-X UF 500 concentrators (Corning, Lowell, MA) with a 5000 Da Molecular Weight Cut Off (MWCO). For each sample, 500 µL of sample was added to the filter unit and spun at 10,000 x g until the filtrate had passed through the column. Three 500 µL volumes of 50 mM ammonium bicarbonate were then passed through the column to remove residual salts. The retained supernatant was then recovered and brought up to a total of 500 µL and quantified for total protein using the Thermo Scientific Micro BCA assay described previously.

#### 2.1.3.2 Trypsin digestion

First- and second-dimension protein fractions were reconstituted in 40 µL of 100 mM Tris-HCl pH 8.5 while unfractionated proteins were reconstituted in 40ul of 8M urea in 100 mM Tris-HCl pH 8.5. Each of the reconstituted protein extracts was mixed with 1.2 µL of 100 mM Tris(2-carboxyethyl)phosphine (TCEP) as a reducing agent. The resuspended proteins were then shaken for 20 minutes at room temperature. After incubation, 0.88 µL of 500 mM Iodoacetamide (IAA) was added and the samples were shaken in the dark for an additional 15 minutes to alkylate the proteins. The alkylated proteins were digested in one of three ways based on the sample being analyzed. These were:

- resuspended proteins from first dimension fractions were mixed with 1.25 µg of trypsin and incubated for 16 hours at 37°C.
- resuspended proteins from second dimension fractions were mixed with 1.25 µg of trypsin and incubated for 8 hours at 37°C after which the samples were sonicated and digested with a second equal volume of trypsin for an additional 8-10 hours at 37°C.
- resuspended proteins from unfractionated body fluid extracts were mixed with 120 µl of 100 mM Tris-HCl to dilute the urea in the sample to 2M after which the sample was mixed with 1 µg of trypsin and incubated for 14-16 hours at 37°C.

The digested proteins from the ProteoSep™ first- and second-dimension fractions were then vacuum dried and re-suspended in 3% acetonitrile and 0.1% formic acid for subsequent analysis by mass spectrometry. Due to the urea present in the unfractionated protein samples, an additional post-digest cleanup and desalting step was necessary.

#### 2.1.3.3 Post Trypsin Peptide Cleanup and Desalting

Pierce C-18 spin columns (Pierce Biotechnology, Rockford, IL) were used to cleanup and concentrate digested peptides from the unfractionated body fluid samples. Each spin column was placed in a 1.5 mL microcentrifuge tube and activated by adding 200 µL of 50% acetonitrile followed by centrifugation at 1500 x g for 1 minute at room temperature. This was repeated one additional time. The column resin was equilibrated with 200 µL of 5% acetonitrile and 0.5% triflouroacetic acid and spun at 1500 x g for 1 minute at room temperature. This step was repeated one additional time. Digested

samples were then loaded onto column and spun at 1500 x g for 1 minute at room temperature. The flow through was collected and passed through the column again at 1500 x g for 1 minute at room temperature. The column was washed with 200 µL of 5% acetronitrile and 0.5% triflouroacetic acid and spun at 1500 x g for 1 minute at temperature. This process was repeated three additional times after which the digested peptides were eluted by adding 20 µL of 70% acetonitrile followed by centrifugation at 1500 x g for 1 minute at room temperature. This was repeated one additional time. The recovered peptides were then lyophilized in a vacuum centrifuge and resuspended in 3% acetonitrile and .1% formic acid for analysis by LC-MS/MS.

#### 2.1.3.4 MS/MS Acquisition

For the analysis of first- and second-dimension ProteoSep™ fractions by mass spectrometry an Agilent 6300 series ion trap mass spectrometer (Agilent Technologies, Santa Clara, CA) was used while for the analysis of unfractionated proteins an Agilent 6510 series Q-TOF mass spectrometer was used. These were all coupled to a 1200 series HPLC-Chip/MS system (Protein ID chip specifications: 150mm 300 Å C18 chip for first-dimension fractions and unfractionated protein samples; Protein ID "short chip" specifications 43mm 300 Å C18 chip for second dimension fractions).

Columns were equilibrated in 0.1% Formic acid in water. For first-dimension fractions and unfractionated proteins, 1 µL of protein digest (equivalent to 1 µg of digested protein) per injection was used. For second-dimension fractions 0.1 µl to 1.5µl of protein digest (variable protein content) per injection was used. A 0-45% of 0.1% formic acid in acetonitrile gradient was performed over 40 minutes followed by a 6

minute column re-equilibration (first-dimension fractions), 11-minutes followed by a 4 minute column re-equilibration (second-dimension fractions) or 44 minutes followed by a 6 minute column re-equilibration (unfractionated proteins).

#### 2.1.3.5 Data Analysis

Data analysis was performed using Spectrum Mill software suite by Agilent Technologies. Data were searched against the Swiss-Prot database. Search parameters included setting the species selection to "Human"; designating a missed cleavage allowance of 3; fixed modifications included carbamidomethylation and variable search modifications included methionine oxidation and deamidation. The search criteria employed with data from the Agilent 6300 series ion trap allowed for a 2.5 Da precursor and a 0.7 Da product ion mass tolerance with a minimum matched peak intensity of 60%. On the 6510 series Q-TOF, the search criteria allowed a 20 ppm precursor and 50 ppm product ion mass tolerances with a minimum matched peak intensity of 60%.

## 2.2 Results and Discussion

In total, over 1000 proteins were identified in the course of the biomarker discovery experiments. This included candidate proteins identified by: (1) Analysis of peaks from the ProteoSep™ which were identified as unique using an in-house software package; (2) Analysis of protein rich pH fractions and; (3) Analysis of unfractionated body fluid samples. Table 2.1 summarizes all candidate biomarkers identified under these three approaches. The details of each experiment are described separately.

**Table 1 - Summary of identified protein biomarkers across all three separation approaches.**

| Fluid | Protein | Accession Number | ProteoSep™ 2D Fraction | ProteoSep™ 1D pH Fraction | Unfrac-tionated |
|---|---|---|---|---|---|
| **Semen** | Epididymal secretory protein E1 | P61916 | | √ | √ |
| | Glycodelin | P09466 | | | √ |
| | Semenogelin-1 | P04279 | √ | √ | √ |
| | Semenogelin-2 | Q02383 | √ | √ | √ |
| | Prolactin-inducible protein | P12273 | √ | √ | |
| | Prostate-specific antigen | P07288 | √ | √ | √ |
| | Prostatic Acid Phosphatase | P15309 | √ | √ | √ |
| **Saliva** | Cystatin-D | P28325 | | | √ |
| | Cystatin-SA | P09228 | | √ | √ |
| | Histatin-1 | P15515 | √ | | √ |
| | Salivary acidic proline-rich phosphoprotein 1/2 | P02810 | √ | | |
| | Statherin | P02808 | √ | | √ |
| | Submaxillary gland androgen-regulated protein 3B | P02814 | | | √ |
| **Vaginal Secretions** | Cornulin | Q9UBG3 | √ | √ | √ |
| | IgGFc-binding protein | Q9Y6R7 | | | √ |
| | Involucrin | P07476 | √ | √ | √ |
| | Ly6/PLAUR domain-containing protein 3 | O95274 | | √ | √ |
| | Matrix metalloproteinase-9 | P14780 | | | √ |
| | Mucin-5B | Q9HC84 | | | √ |
| | Neutrophil gelatinase-associated lipocalin | P80188 | | √ | √ |
| | Periplakin | O60437 | √ | √ | |
| | Suprabasin | Q6UWP8 | | | √ |
| | Vimentin | P08670 | √ | | |
| **Urine** | Osteopontin | P10451 | √ | √ | √ |
| | Uromodulin | P07911 | | √ | √ |
| **Menstrual Fluid** | Calpastatin | P20810 | | √ | |
| **Peripheral Blood** | α-1-antitrypsin | P01009 | | √ | √ |
| | Complement C3 | P01024 | √ | √ | √ |
| | Hemoglobin subunit beta | P68871 | | √ | √ |
| | Hemopexin | P02790 | √ | √ | √ |

### 2.2.1 Biomarker Identification: ProteoSep™ 2D Fractionation

Five individuals were recruited to donate samples of each body fluid being analyzed. This redundancy was intended to help to discriminate between proteins that are consistently expressed in a specific body fluid versus those that might reflect inter-individual variability in protein expression - and thus not be suitable as biomarkers. Five individuals reflected an additional effort to obtain the maximum amount of proteomic information within a reasonable budget and timeframe. Additionally, while mapping proteomes from a larger number of study participants would provide a stronger assessment of protein expression among different humans, the primary objective of this stage of the search was to identify candidate biomarkers of body fluids that could be subsequently validated for use across a larger population.

Samples were prepared as summarized in **Figure 15.** Saliva, seminal fluid, urine, and vaginal fluid were prepared in a similar manner. In all cases, samples were centrifuged at high speed to remove cellular material followed by concentration and buffer exchange with ProteoSep™ loading buffer.

Menstrual and peripheral blood needed to be treated separately primarily due to the complexity of the blood proteome – and the presence of abundant non-specific proteins. Peripheral blood was collected in an EDTA gel barrier tube in order to generate blood serum devoid of cells and clotting factors. Menstrual blood, however, was typically contaminated with hematosed erythrocytes. As hemoglobin comprises 32-36% of all the proteins found in red blood cells, the serum from menstrual blood samples containedlarge

quantities of hemoglobin which acted to mask the detection of less abundant menstrual fluid-specific proteins. For this reason, hemoglobin was removed from collected serum samples prior to proteome fractionation through the use of HemogloBind™.

Because the dynamic range of protein concentration in serum spans more than ten orders of magnitude, the presence of these high-abundance proteins (*e.g.* albumin, IgG, transferrin) makes it difficult to identify proteins that are abundant but which are more likely to be specific markers of each body fluid[115]. To circumvent this potential problem, commercially available IgY-12 Proteome Partitioning columns were employed. These antibody-based columns made it possible to remove twelve highly abundant proteins (albumin, IgG, transferrin, fibrinogen, IgA, α-2-Macroglobin, IgM, α-1-acid antitrypsin, haptoglobin, α-1-acid glycoprotein, apolipoprotein A-I and A-II) from human blood serum. This yielded an enriched pool of the less abundant but more body fluid-specific blood proteins in the flow-through fraction.

Each body fluid sample was quantified using the Micro BCA kit as described above and a standardized 3 mg of total protein was loaded onto the ProteoSep™ system. In total, thirty proteome maps were generated. In each of these maps, data from the chromatofocusing column (X axis) was combined with data from the reverse-phase column (Y axis) to yield a .dat file containing X, Y, and Z (height) values which can be processed and visualized with the ProteoVue™ software suite. Examples of the 2D pH/hydrophobicity/quantity maps for peripheral blood, urine, semen and saliva are shown in **Figure 16**. The intensity and color of the bands represent the abundance of the protein

detected. Red, orange and yellow bands represent more abundant proteins while green and blue represent less abundant proteins.

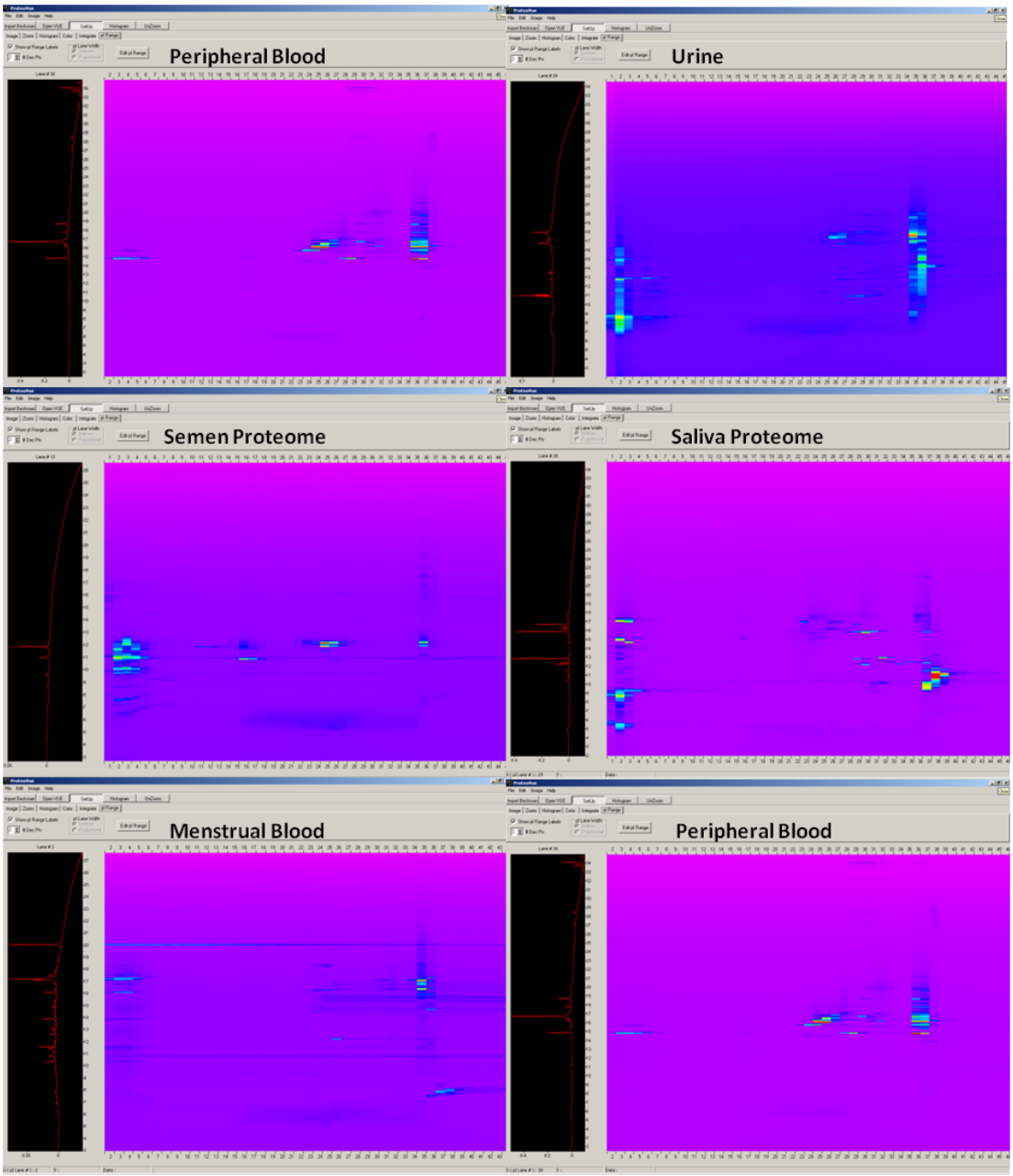



**Figure 16 - Examples of the 2D pH (*x*-axis)/hydrophobicity (*y*-axis) maps (*i.e.*, proteome maps) obtained from five different body fluid samples (peripheral blood, urine, semen, saliva, and menstrual fluid/blood). Differences in the position of bands between the proteome maps for different body fluids are indicative of potential protein markers of interest.**

### 2.2.1.1 Proteome Map Comparison

Rigorous pairwise comparisons among the 2D-fractionated pH / hydrophobicity / protein expression "proteome maps" for each body fluid were made as part of the effort to identify potential high-specificity biomarkers. Initially, the commercial DeltaVue™ software suite which was integrated into the ProteoSep™ System was used for comparing proteome maps. The software requires a .dat file input to generate color-coded "differential display" maps in which differences between the proteomic profiles of any two samples (**Figure 17**) are highlighted.

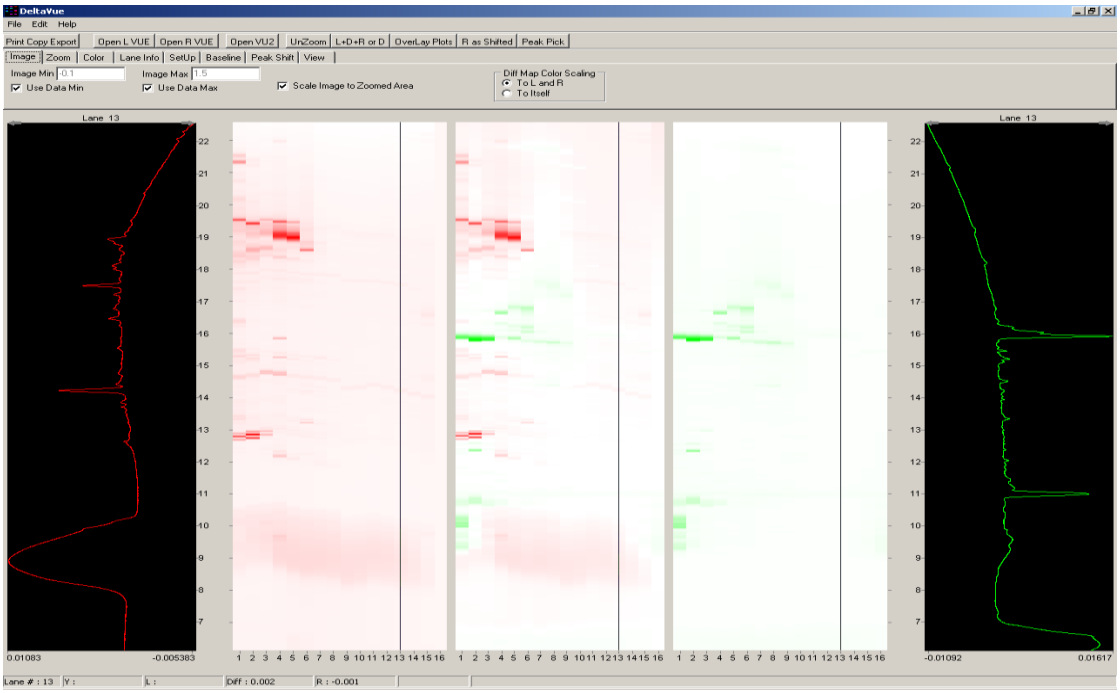



**Figure 17 - DeltaVue™ difference profile (center panel) comparing the urine proteomic map (center left panel in red) with the saliva proteome map (center right panel in green). Loss of resolution of lower abundance proteins and the inability to normalize protein position for subtle retention time differences and pH variances between proteome maps imported into DeltaVue™ make it difficult to reliably identify true protein differences between body fluids.**

Following several comparisons, however, the limitations of the software became apparent. The software was designed for traditional biomedical-based proteomics projects where diseased samples are compared to healthy samples. As an example of the more typical workflow in a disease diagnostics context, tissue samples from multiple individuals with a given biological state of interest (*e.g.,* healthy *vs.* diseased hepatic tissues) would be run on the ProteoSep™. In that case, two nearly identical proteomes would be compared to identify quantitative differences in protein expression. By contrast, using the ProteoSep™ for comparisons of six highly dissimilar biological matrixes introduced an unanticipated level of complexity – specifically the need to accurately align and compare very dissimilar proteomic maps. Doing such comparisons using the standard DeltaVue™ software would require hundreds of one-to-one comparisons between different proteomes.

Given the inherent limitations of the instrument's DeltaVue™ software, it was necessary to develop an in-house software suite to facilitate the rigorous and simultaneous comparison of multiple proteome maps at once. To this end, a series of data optimization procedures were developed and combined with a data mining algorithm to identify body fluid-specific peak clusters (*i.e.*, *x*-/y-/z-axis coordinates representing the HPCF pH fraction, HPRP retention time and expression level, respectively of individual proteins) within the ProteoSep™ proteome maps. Those ProteoSep™ fractions containing proteins corresponding to the peaks flagged by the software as unique to the proteome of a specific body fluid were then recovered and the protein(s) represented by the peak(s) of interest were identified using tandem mass spectrometry.

### 2.2.1.2 Dataset Optimization

As stated above, each proteome map is stored as a .dat file which is produced by the ProteoSep™ System. Each proteomic map file consisted of >450,000 data points in a 45 x 10,501 data point matrix. A large portion of these data points, however, represent "uninformative" background noise created when no proteins are eluting from the system or represent what is essentially "uninformative" space between peaks during the reverse-phase separation (**Figure 18**).

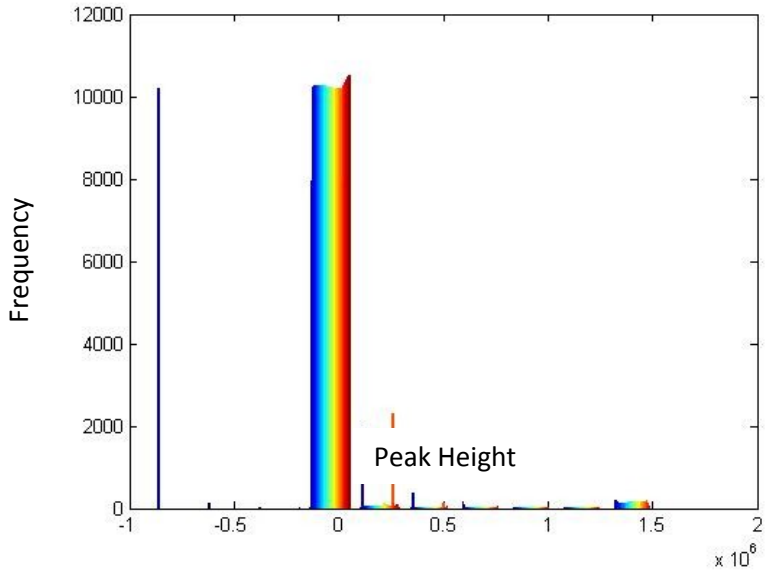



**Figure 18 – Histogram of peak height distribution generated by the ProteoSep™ system. The majority of the stored data exists as background/non-protein regions. This can be seen as the majority of data center around 0, which represents background signal in the instrument.**

Due to the enormous quantity of data points generated for each proteome map, coupled with the fact that the majority of the data exists as background signal, a protein peak identification algorithm was implemented to transform each proteome map into a simplified dataset containing only the coordinates of peak locations (**Figure 19**). The algorithm works on a simple principle, that the point at which a slope changes from

positive to negative (or the point where a derivative changes signs) represents the coordinates on the proteome map where a protein peak is likely to exist. Using this technique virtually all of the underlying noise and uninformative data can be eliminated while increasing the resolution of the signal itself.

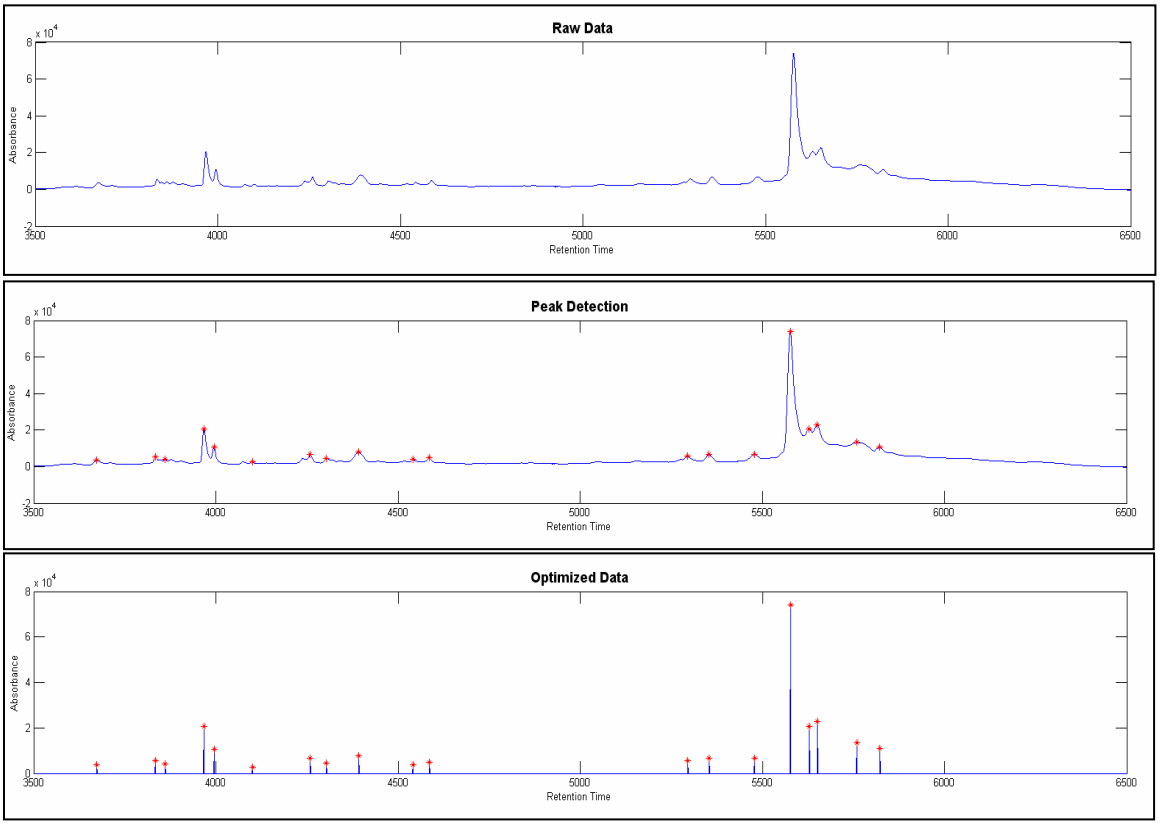



**Figure 19 - Application of a protein peak identification algorithm to second-dimension raw hydrophobicity data (top) generated from a single first-dimension ProteoSep™ fraction. Protein peaks are called based on the point at which the slope of the trace changes from positive to negative or the point where a derivative changes signs (middle). Once identified, called protein peaks are used to create a normalized representative dataset from which > 99.9% of the underlying noise and "uninformative" data has been eliminated (bottom).**

### 2.2.1.3 Data Mining of Optimized Proteome Maps

It was also necessary to create a reliable means of comparing different proteomes while taking into consideration the subtle differences between sample runs (*x*/*y* axis) and interindividual (*z* axis) differences in protein expression levels. For this task, a data

mining algorithm was implemented to combine individual proteome maps for the same body fluid into a single consensus proteome map. Data mining for this purpose was defined as grouping like objects together. This "clustered/consensus map" was then used to facilitate comparisons among different body fluids. The specific algorithm that was implemented is known as the k-means clustering - an algorithm that organizes a dataset into k subsets. The algorithm involves a four step procedure. First, a location is assigned for each of the subset centers k (centroids); second, each data point is assigned to its nearest center; third, the optimal position of each center is calculated based off a distance measure to each point assigned to it and; fourth, steps 2 and 3 are repeated until the centers are "stable" with each center representing the set of individual protein points from multiple proteome maps from the same body fluid. The overall product of this procedure was a set of points split into k partitions, meaning that all of the peaks representing a given protein across multiple samples of the same body fluid from different individuals were grouped together in three dimensional space and were represented by a single central "consensus point". The clustered "Consensus" proteome maps for each of the body fluids of interest were then overlaid with each other and cluster centers were compared between each fluid. Any cluster center which appeared in a single fluid and did not overlap with a cluster center in any other fluid was flagged as the map position of a potential protein biomarker for subsequent identification through tandem mass spectrometry.

### 2.2.1.4 Recovery and Analysis of ProteoSep™ 2D Fractions

The translation of identified proteome map coordinates by Protein Miner™ into actual candidate protein biomarkers was a relatively straight forward process. The first step of this workflow began with the retrieval of ProteoSep™ second dimension fractions containing potentially unique proteins. A fractional analysis software package from the University of Michigan was used to rapidly locate a specific protein fraction that had been eluted from the ProteoSep™ second dimension HPRP column. By placing the cursor over a desired peak from the second dimension chromatogram, the software identifies the specific deep-well plate location of the protein fraction containing the peak of interest (**Figure 20**).

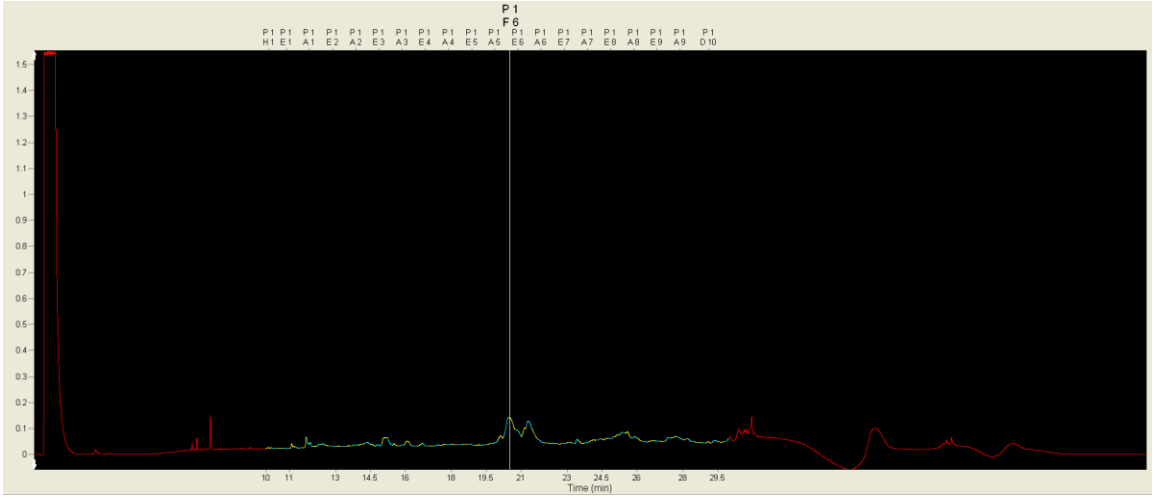
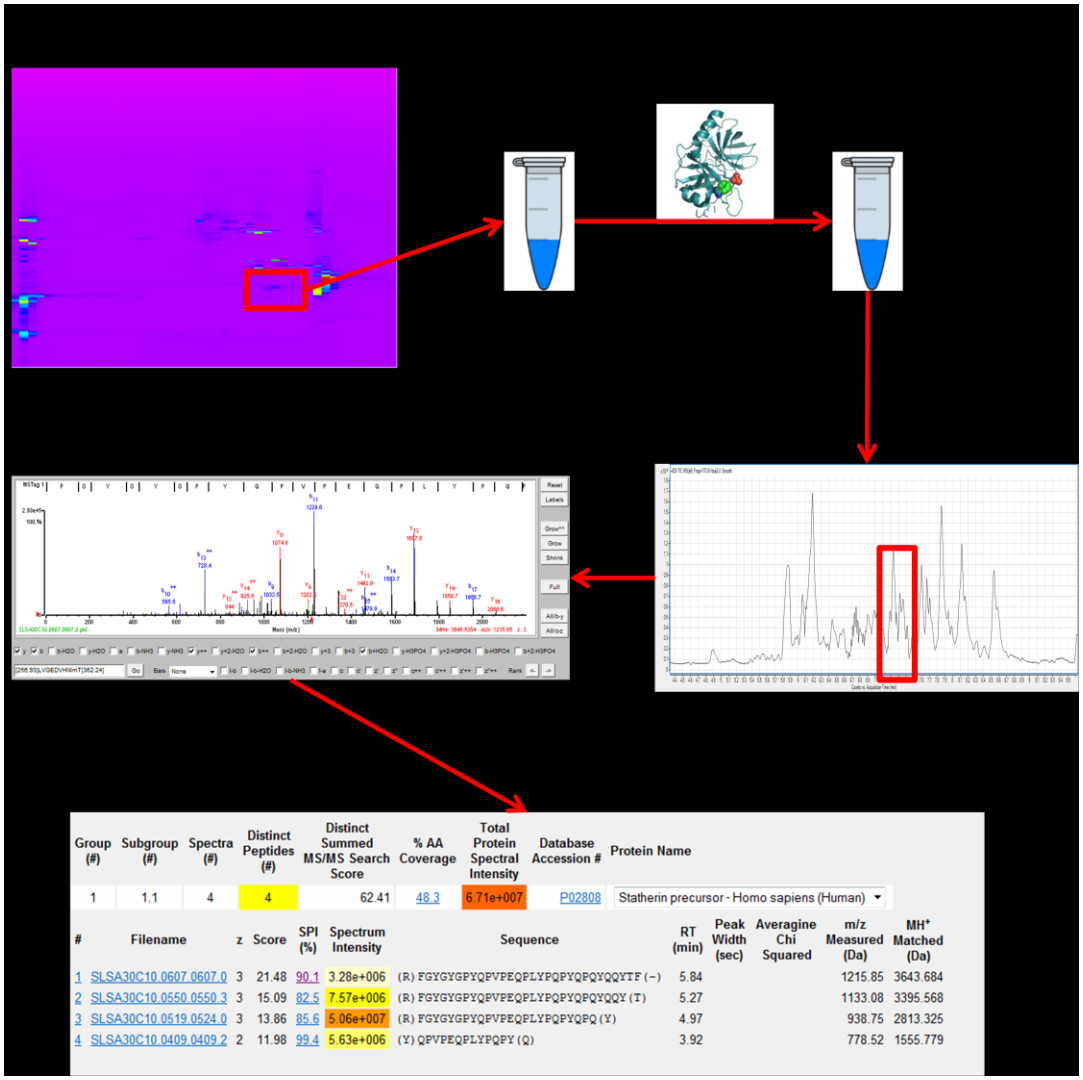



**Figure 20 – Fraction analysis software acquired from the University of Michigan. Cursor placement reveals the exact 96 deep-well plate and well number where the proteins in a peak of interest are expected to be found.**

The proteins contained in these fractions were prepared as stated in section 2.1.3. Briefly, the protein(s) contained in a fraction of interest were denatured and digested to form peptide fragments using the enzyme trypsin. Mass spectrometry was performed on an Agilent 6300 series ion trap mass spectrometer coupled with a nanoflow-LC 1200

series HPLC-Chip/MS system. Because of the molecular simplicity of each ProteoSep™ second-dimension fraction (*i.e.*, only 1-3 proteins/fraction are expected to be eluted), it was possible to use a short 43mm analytical column with only an 11-minute gradient was to obtain accurate protein identifications.

Database identification of specific proteins was performed using Agilent's Spectrum Mill search engine as described previously. **Figure 21** shows an overview of the workflow and search results for a cluster of peaks representing potentially unique protein biomarkers in saliva sample #SLSA. Using the peak selection software the peak of interest was localized to HPCF pH fraction 30/HPRP fraction C10 which correlated to well F6 of 96 deep-well plate P1. After recovery of the liquid-phase fraction from this well, digestion, MS/MS analysis and database search of the resulting data, the protein statherin was identified as a potential biomarker of human saliva. This result was encouraging as statherin had also been independently identified as a possible saliva marker by gene expression database searches and by forensic researchers working on the development of mRNA biomarkers for saliva[54, 116]. Similarly, the identification of semenogelin-1 and -2 as markers of seminal fluid[117] and periplakin as marker of vaginal secretions were consistent with what has been reported by biomedical researchers[118]. The accuracy with which the proteomic approach was able to identify these markers was seen as boding well for the likely specificity of numerous other candidate biomarkers that were identified - but for which information on tissue specificity in the literature was lacking.



| Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Distinct Summed MS/MS Search Score | % AA Coverage | Total Protein Spectral Intensity | Database Accession # | Protein Name |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | 4 | 4 | 62.41 | 48.3 | 6.71e+007 | P02808 | Statherin precursor - Homo sapiens (Human) |

| # | Filename | z | Score | SPI (%) | Spectrum Intensity | Sequence | RT (min) | Peak Width (sec) | Averagine Chi Squared | m/z Measured (Da) | MH+ Matched (Da) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SLSA30C10.0607.0607.0 | 3 | 21.48 | 90.1 | 3.28e+006 | (R) FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF (-) | 5.84 | | | 1215.85 | 3643.684 |
| 2 | SLSA30C10.0550.0550.3 | 3 | 15.09 | 82.5 | 7.57e+006 | (R) FGYGYGPYQPVPEQPLYPQPYQPQYQQY (T) | 5.27 | | | 1133.08 | 3395.568 |
| 3 | SLSA30C10.0519.0524.0 | 3 | 13.86 | 85.6 | 5.06e+007 | (R) FGYGYGPYQPVPEQPLYPQPYQPQ (Y) | 4.97 | | | 938.75 | 2813.325 |
| 4 | SLSA30C10.0409.0409.2 | 2 | 11.98 | 99.4 | 5.63e+006 | (Y) QPVPEQPLYPQPY (Q) | 3.92 | | | 778.52 | 1555.779 |

**Figure 21 – Protein identification workflow. Briefly, proteins flagged as unique are recovered, digested and analyzed on an ion trap mass spectrometer. Acquired data is then processed using Agilent's Spectrum Mill search engine to determine the identity of the unknown proteins in the fraction.**

This strategy of cluster analysis, unique peak selection, ProteoSep™ 2D fraction recovery, and biomarker identification through tandem mass spectrometry was repeated for each body fluid of interest in order to populate a database of candidate biomarkers. Where available, the potential utility *(i.e.*, tissue specificity) of the candidate biomarkers was bolstered based on searches of independent proteome database projects and published reports in the professional proteomic literature. These included: semenogelin-

I/II, prolactin inducible protein, prostate specific antigen, and prostatic acid phosphatase as markers of seminal fluid; histatin-1, salivary acidic proline rich phosphoprotein 1/2, and statherin as markers of saliva; cornulin, involucrin, periplakin, and vimentin as markers of vaginal fluid; osteopontin as a candidate urine biomarker; and α-1-antitrypsin, complement C3, hemoglobin subunit β, and hemopexin as markers of peripheral blood. Unfortunately, no high-specificity menstrual fluid markers were identified using this approach. Analysis of each candidate protein based on information in Swiss-Prot /uniprot/ncbi as well as published biomedical literature made it possible to arrive at a reasonably specific listing of candidate biomarkers. Biomarker candidates will be described further in relation to their biological function and tissue specificity in a later section.

While this particular approach provided promising results, the original goal of generating five to ten candidate biomarkers for each fluid was not met. Additionally, no viable menstrual fluid proteins were identified using the ProteoSep™ mapping strategy. The fundamental assumption employed in this approach is that clusters flagged by the software would contain unique proteins – that is to say, a peak cluster found only in saliva should contain a saliva-specific protein. The reality, however, was that not all unique peak clusters were found to contain viable biomarkers. **Figure 22** illustrates a search result from the analysis of a vaginal fluid fraction that was flagged as containing a unique protein. The search results of the acquired MS/MS data identified several serpin isoforms (protease inhibitors) as well as S100 proteins (zinc-binding multidisciplinary

proteins)[119, 120], all of which are known to be ubiquitously expressed in a range of human tissues.

| Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Distinct Summed MS/MS Search Score | % AA Coverage | Total Protein Spectral Intensity | Database Accession # | Protein Name |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | 3 | 3 | 39.46 | 31.1 | 5.14e+007 | P05109 | Protein S100-A8 OS=Homo sapiens GN=S100A8 PE=1 SV=1 |
| 2 | 2.1 | 3 | 3 | 34.30 | 9.4 | 1.57e+007 | P29508 | Serpin B3 OS=Homo sapiens GN=SERPINB3 PE=1 SV=2 |
| 2 | 2.1 | 2 | 2 | 24.14 | 5.8 | 1.10e+007 | P48594 | Serpin B4 OS=Homo sapiens GN=SERPINB4 PE=1 SV=2 |
| 2 | 2.1 | 1 | 1 | 13.19 | 3 | 6.23e+006 | Q9UIV8 | Serpin B13 OS=Homo sapiens GN=SERPINB13 PE=2 SV=2 |
| 3 | 3.1 | 2 | 2 | 27.11 | 24.5 | 2.64e+007 | P06702 | Protein S100-A9 OS=Homo sapiens GN=S100A9 PE=1 SV=1 |
| Totals: | 11 | 11 | | | | | | |

**Figure 22 – Spectrum Mill results from the analysis of a ProteoSep™ 2D fraction flagged as containing a peak cluster unique to vaginal fluid. The identified proteins, however, are known to be ubiquitously expressed in non-target tissues.**

These results established that unique clusters of peaks may not necessarily contain body fluid-specific proteins. Conversely, not all clusters with similar (or even identical) map coordinates between two fluids may contain the same proteins. To circumvent this limitation of the cluster analysis approach, two alternative proteome strategies were employed to identify additional biomarker candidates.

### 2.2.2 Biomarker Identification: ProteoSep™ 1D Fractionation

As stated above, corresponding map coordinates and peak clusters from two separate fluids may not necessarily indicate that the peaks represent the same protein. As a result proteins with potential utility as high-specificity biomarkers may be missed using a 2D cluster analysis based approach. To test this, protein-rich ProteoSep™ first-dimension pH fractions for each body fluid were desalted, quantified, and run on a longer gradient in an effort to identify additional biomarkers which may have been missed by the initial cluster analysis. **Figure 23** shows a portion of a vaginal fluid proteome map corresponding to first-dimension pH fractions 20 through 27 which were selected for MS/MS analysis using the methods described.

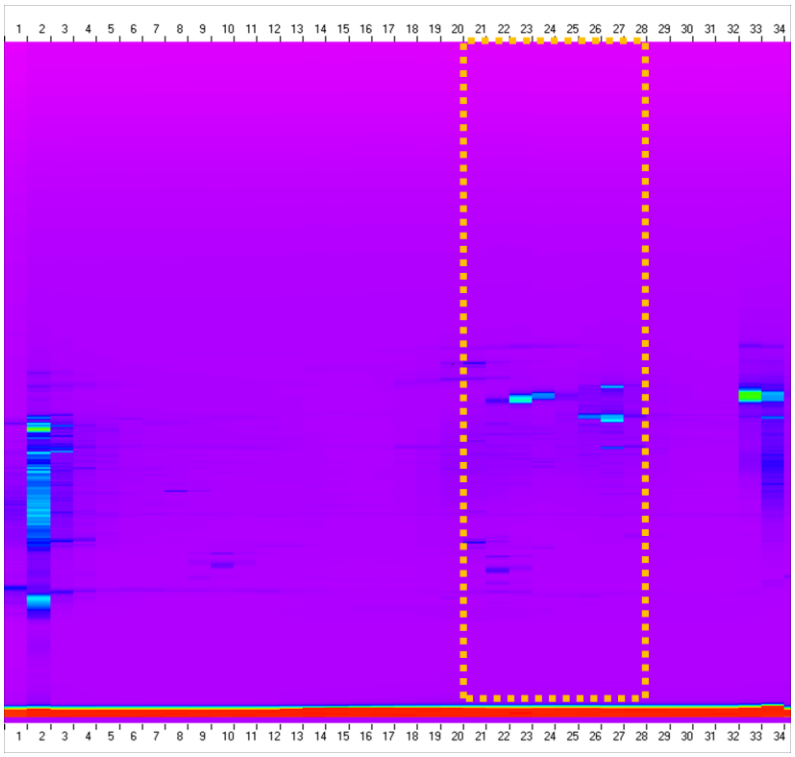



**Figure 23 – Proteome map of a vaginal fluid sample showing protein-rich ProteoSep™ first-dimension pH fractions 20-27 which were recovered from the HPCF eluent deep-well plate for MS/MS analysis.**

Analysis of the ProteoSep™ first-dimension pH fractions resulted in the addition of several candidate biomarkers for each body fluid and successfully identified the first candidate biomarker for menstrual fluid (calpastatin). Analysis of seminal fluid detected all previously identified biomarkers and identified one additional candidate protein, epididymal secretory protein E1. Cystatin-SA was identified as an additional candidate biomarker for saliva. Vaginal fluid analysis was also successful in identifying several new candidate biomarkers including Ly6/PLAUR domain-containing protein 3 and neutrophil gelatinase-associated lipocalin. Finally, uromodulin was added to the candidate list for urine.

These results confirmed the hypothesis that a single pass approach to biomarker discovery based on a cluster analysis of ProteoSep™ 2D proteome mapping results may

not provide a sufficient resolution of the proteome to ensure the identification of all possible body fluid specific proteins. This alternative approach achieved greater protein resolution and as a result added six new candidate biomarkers while confirming the identification of many other biomarkers (*e.g.* semenogelin-I/II). The confirmation provided by this approach provided support to the view that those candidate protein biomarkers that were identified by cluster analysis are reliable and reproducibly detectable.

### 2.2.3 Biomarker Identification: Unfractionated Body Fluids

Given the success achieved by direct analysis of the ProteoSep™ first-dimension pH fractions, an additional series of experiments were performed in which unfractionated body fluids were analyzed by tandem mass spectrometry. It is well known that not all proteins are compatible with all separation approaches. Thus, this second alternative strategy represented an effort to obtain an even more comprehensive dataset of the proteins expressed in each body fluid of interest. Additionally, data from unfractionated body fluid extracts would offer a "real world" assessment of the biomarker detection characteristics as no laborious fractionation or enrichment steps are involved – an important issue that may impact the ultimate feasibility of implementing mass spectrometry-based assays in a forensic caseworking environment.

To reduce the potential impact of interindividual variability in protein expression, three samples of each body fluid were prepared as described under sections 2.1.1 "Sample Collection and Preparation" and 2.1.3 "Protein Identification by Mass Spectrometry". However, neither menstrual fluid nor peripheral blood samples were

passed through an immunodepletion column nor was any effort made to remove hemoglobin for these experiments.

The results obtained by MS/MS analysis of these samples further supplemented the candidate database for saliva as well as vaginal fluid while reconfirming the apparent tissue specificity of several candidate markers that had been previously identified by cluster analysis and first-dimension pH fraction strategies. The analysis of unfractionated body fluid protein extracts resulted in the addition of cystatin-D, histatin-1, as well as submaxillary gland androgen-regulated protein 3B as candidate biomarker list for saliva. Immunoglobulin G Fc-binding protein, matrix metalloproteinase-9, mucin 5B, as well as suprabasin were identified as high-specificity biomarker candidates of vaginal fluid.

### 2.2.4 Candidate Biomarker Review

A comprehensive list of candidate biomarkers (Table 1) was compiled after the completion of the biomarker discovery phase of this research. This list was propagated through a combination of empirical comparative proteomic data as well as an inspection of public proteome database annotations a review of the proteomic and biomedical literature. A description of the candidate biomarkers that comprised the initial assay panels for each body fluid of interest are provided in detail in the following sections. A more thorough investigation and validation of the tissue specificity (and thus ultimate utility) of these candidate markers was also performed and is described in Chapter 3.

#### 2.2.4.1 Seminal Fluid

Seminal fluid showed highly consistent results between all three experimental methods; only epididymal secretory protein E1 and prolactin-inducible protein were not consistently expressed across all experiments. Epididymal secretory protein E1 was first

identified as a major secretary protein in the epididymis[121]. The protein is involved in the transportation of cholesterol from lysosomes and is proposed to have a role in sperm formation[122, 123]. This protein was not identified in the ProteoSep™ 2D fractions but was found in the comparative proteomic analyses of ProteoSep™ 1D pH fractions as well as unfractionated samples.

Semenogelin-I/II are the most abundant proteins in seminal plasma and are responsible for the gel-like matrix of human semen. Both isoforms act as substrates for prostate specific antigen (p30), where upon lysis, sperm are able to move freely through the seminal matrix[124, 125]. Both protein variants were identified by all three biomarker discovery methods.

Prolactin-inducible protein is a small 17 kDa glycoprotein located in saliva, sweat glands, but with very high expression in seminal fluid[126]. Depending on the expression levels in saliva, this protein may not be suitable for differentiating these two fluids. The protein has numerous biological functions including protection of spermatozoa from damage by endogenous IgG antibodies[127]. Prolactin-inducible protein was detected in the comparative proteomic analyses of ProteoSep™ 2D fractions as well as the ProteoSep™ 1D pH fractions but was not detected in analyses of unfractionated samples.

Prostate-Specific Antigen (also known as PSA or p30) is a serine protease produced by epithelial cells located in the prostate[128]. The primary function of prostate-specific antigen is to cleave semenogelin-I/II thus creating a soluble, liquid medium, for spermatozoa movement[129]. Prostate-specific antigen has been well studied as an indicator for prostate cancer when serum levels reach approximately 4-10 ng/mL[130, 131]. While this protein is not absolutely seminal fluid-specific, the detection limits of most

assays make it difficult to detect it in whole blood[132]. As a result, this protein may have utility, in combination with other biomarkers, for the detection of seminal fluid. As with semenogelin, PSA/p30 was reproducibly detected in by three biomarker discovery strategies.

Prostatic Acid Phosphatase (also known as Seminal Acid Phosphatase or SAP) is a glycoprotein secreted by the epithelial cells of the prostate gland which is capable of hydrolyzing phosphate groups from substrate molecules [133]. SAP is another seminal fluid protein which has seen utility as a clinical marker for prostate cancer[134, 135]. While largely replaced by PSA/p30 for screening purposes, the combination of low expression in non-target tissues and assay detection limits makes this protein useful as a potential marker of seminal fluid.

All of the proposed protein biomarkers of seminal fluid have been well characterized in proteomic experiments and have been reported in numerous independent studies of the seminal fluid proteome thereby demonstrating the potential reliability of these candidate markers[136, 137].

**2.2.4.2 Saliva**

Salivary biomarkers were more variable in their detection across the three experiments. Cystatin D and cystatin SA originate from the submandibular and sublingual glands and act as protease inhibitors in saliva[138]. Both of these proteins are salivary specific with no expression in the seminal vesicles, prostate, ovaries or other tissues[139]. Cystatin D was only detected in analyses of unfractionated protein samples whereas cystatin SA was detectable in analyses of both unfractionated protein samples and the ProteoSep™ 1D pH fractions.

Histatin-1 originates from the submandibular and sublingual glands and is associated with the regulation of salivary calcium[140, 141]. Histatin-1 was identified in the ProteoSep™ 2D fractions as well as the unfractionated protein samples. Salivary acidic proline-rich phosphoprotein is secreted into saliva where it also acts to regulates calcium phosphate salts providing a protective / reparative environment for dental enamel[142, 143]. This protein was identified only in comparisons of the ProteoSep™ 2D fractions. Statherin, also secreted into saliva, has a similar function to that of salivary acidic proline-rich phosphoprotein 1/2. Statherin inhibits potentially harmful calcium phosphate precipitation in saliva[144, 145]. Statherin was detected in both the comparative proteomic analyses of ProteoSep™ 2D fractions and unfractionated protein samples.

All the proposed biomarkers of saliva have also been well characterized in numerous proteomics studies thereby increasing the likelihood that these will prove to be reliable biomarkers for saliva[146, 147].

**2.2.4.3 Vaginal Secretions**

Vaginal fluid was among the most difficult body fluids to analyze as the proteins associated with the fluid appear to be less consistent across multiple experiments and are less rigorously documented in the literature in terms of tissue localization and biological function. Involucrin is a structural protein contained within squamous epithelial cells[148] that was detected only in vaginal fluid samples by all three biomarker discovery strategies. Similar to involucrin, cornulin is also expressed in squamous where it plays a role in epithelial cell differentiation. It may also play a role in mucosal-epithelial immune response. The protein has been characterized in the cervix and in esophageal tissues[149, 150]. This being said, it was not detected in any saliva samples in the current

study. This protein is one of only two candidate biomarkers of vaginal fluids that was identified by all three comparative proteomic approaches.

IgGFc-binding protein is involved with the mucosal layers of epithelial cells, it has been found in the colon and placenta, and is thought to aid in immunity[151]. IgGFc-binding protein was only detected in analyses of unfractionated protein samples.

Ly6/PLAUR domain-containing protein 3 is involved in the regulation between extracellular structural support scaffolding and epithelial cell layers[152]. Ly6/PLAUR domain-containing protein 3 was identified in the comparative proteomic analyses of both ProteoSep™ 1D pH fractions and unfractionated protein samples.

Matrix metalloproteinases are a family of proteins which break down and remodel components of the extracellular matrix[153]. Matrix metalloproteinase-9 was found only in unfractionated protein samples of vaginal fluids. Although this protein has been reported in serum and blood plasma, its expression levels appear to be below the detection limits of most assays making it a potentially viable candidate biomarker for use in conjunction with other candidate proteins[154, 155].

Mucin-5B is a member of a class of large highly glycosylated proteins which form elastic matrices on mucosal surfaces[156]. Mucin 5B is reportedly expressed in the mucosal glands of the uterine endocervix as well as the airway epithelium[157-159]. In the current study mucin-5B was found to be a very abundant protein in vaginal fluid but may also detected in saliva samples. Nevertheless, it may have utility as a vaginal marker if the protein is undetectable in the saliva samples. Mucin-5B was only reliably detected in the comparisons of unfractionated protein samples.

Neutrophil gelatinase-associated lipocalin belongs to the lipocalin family of transport proteins which have been associated with innate immunity though iron sequestration[160]. As such, this protein can be found in tissues prone to exposure to bacterial and other microorganisms including the respiratory tract, salivary glands, uterus, and prostate[160, 161]. Neutrophil gelatinase-associated lipocalin was detected in analyses of ProteoSep™ 1D pH fractions and unfractionated protein samples.

Periplakin is a protein component of the cornified envelope in human skin[162] and was identified in comparative analyses of both the ProteoSep™ 1D pH fractions and unfractionated protein samples. Suprabasin is expressed in keratinocytes and plays a role in epidermal differentiation. It has been reported to be expressed in the uterus as well as the esophagus[163]. Suprabasin was only identified in analyses of unfractionated protein samples. Vimentin which acts as the major intermediate filament in mesenchymal stem cells[164], was identified only through cluster analysis of the ProteoSep™ 2D fractions.

Vaginal fluid was the most difficult body fluid to analyze with respect to specificity and thus the potential utility of candidate biomarkers. The presence of these markers as being consistently identified in a series of vaginal proteomics projects was reviewed in the professional literature. Table 2 presents a compiled list of eight vaginal fluid proteomics research reports and lists the study in which the candidate biomarker was identified.

**Table 2 – Meta-Analysis of Biomarkers from Eight Vaginal Fluid Proteomics Studies[a].**

| Protein | Accession Number | Study 1[165] | Study 2[166] | Study 3[167] | Study 4[168] | Study 5[169] | Study 6[170] | Study 7[171] | Study 8[172] |
|---|---|---|---|---|---|---|---|---|---|
| Cornulin | Q9UBG3 | | | | | √ | | | √ |
| IgGFc-binding protein | Q9Y6R7 | | | | | √ | | | |
| Involucrin | P07476 | √ | | √ | √ | √ | | | √ |
| Ly6/PLAUR domain-containing protein 3 | O95274 | | | | | √ | | | √ |
| Matrix metalloproteinase-9 | P14780 | √ | | | √ | √ | | | |
| Mucin-5B | Q9HC84 | √ | | | √ | √ | | | √ |
| Neutrophil gelatinase-associated lipocalin | P80188 | √ | | √ | √ | √ | √ | √ | √ |
| Periplakin | O60437 | √ | | | √ | √ | | | √ |
| Suprabasin | Q6UWP8 | | | | | √ | | | √ |
| Vimentin | P08670 | √ | | | √ | √ | | | √ |

[a]Table adapted from Tang[170].

Each of the candidate protein biomarkers of vaginal fluids selected for forensic validation was also identified in at least two independent projects with several being identified in three to six of the eight published studies. Because the candidate biomarkers form the current study are consistent with what has been independently reported by other biomedical researchers, it is believed that these proteins have significant potential utility as reliable biomarkers. However, it is also appreciated that due to the limited literature available on some of these biomarkers, more rigorous forensic validation studies may reveal that some of the proposed candidates are not actually unique to vaginal fluids.

#### 2.2.4.4 Urine

Two prospective urine biomarkers were identified in the course of the current research. The protein osteopontin provides protection from kidney stone formation through the inhibition of calcium crystal formation[173]. Osteopontin can also be found in plasma with connections varying from autoimmune disorders to tumor progression [174, 175]. Osteopontin was identified by all three comparative proteomics strategies.

Uromodulin, also known as Tamm-Horsfall urinary glycoprotein, is the most abundant protein found in human urine[176]. It has been proposed that uromodulin is involved in preservation of water and electrolyte levels as well as being linked to infection prevention[176, 177]. Uromodulin is also involved as an extracellular protein in bone matrix formation[178].Uromodulin was detectable in analyses of both ProteoSep™ 1D pH fractions and unfractionated protein samples. Both of these proteins are reproducibly found in urine and thus would appear to make good targets for reliable urine identification[179, 180].

#### 2.2.4.5 Menstrual Blood

Only a single menstrual blood-specific biomarker was identified, that being calpastatin, which was initially flagged through comparative analysis of the ProteoSep™ 1D pH fractions. Calpastatin, is another cysteine protease inhibitor. More specifically, this protein is thought to be involved in the regulation of a subset of cellular apoptotic events[181] and to play a role in tissue preservation and muscle degradation through proteolysis inhibition[182]. To date, no menstrual blood proteomics projects have been published. The failure to readily detect this marker in unfractionated menstrual fluid samples does raise some concern. It is possible that the protein may not be universally

expressed which would explain the absence in the other samples analyzed. It is also a possibility that removing the immunodepletion step as part of sample processing has placed the protein below the detection limits of even modern mass spectrometers.

However, there were multiple vaginal-specific candidate biomarkers detected in menstrual blood as might be anticipated. Table 3 summarizes the prevalence of these proteins in menstrual blood. Because vaginal fluid will be a component of any vaginal sourced sample, these data suggest that it may be possible to generically differentiate a sample as being of potential vaginal origin. However, there would be no way to confidently identify a sample as menstrual fluid versus a mixture of peripheral blood and vaginal fluid.

**Table 3 – Vaginal Fluid Proteins Identified in Menstrual Blood.**

| Fluid | Protein | Accession Number | ProteoSep™ 2D Fraction | ProteoSep™ 1D pH Fraction | Unfrac-tionated |
|---|---|---|---|---|---|
| **Vaginal Biomarkers in Menstrual Blood** | Cornulin | Q9UBG3 | | | √ |
| | IgGFc-binding protein | Q9Y6R7 | | | |
| | Involucrin | P07476 | | | |
| | Ly6/PLAUR domain-containing protein 3 | O95274 | | | |
| | Matrix metalloproteinase-9 | P14780 | | | |
| | Mucin-5B | Q9HC84 | | | |
| | Neutrophil gelatinase-associated lipocalin | P80188 | √ | √ | |
| | Periplakin | O60437 | | | |
| | Suprabasin | Q6UWP8 | | | |
| | Vimentin | P08670 | | | √ |

### 2.2.4.6 Peripheral Blood

Four peripheral blood biomarkers were selected during the course of the biomarker discovery phase of the current study. α-1-antitrypsin is a non-specific serine protease inhibitor found in human plasma. This protein's primary role is as an inhibitor of

neutrophil elastase thus protecting tissues from proteolytic damage[183, 184]. α-1-antitrypsin was identified in the comparative proteomic analyses of both the ProteoSep™ 1D pH fractions and unfractionated protein samples.

Complement C3 is a serum protein implicated in the complement cascade of the immune system. More specifically, complement C3 is involved with innate immunity and aids antibodies as well as phagocytes in response to pathogen infection[185, 186]. Complement C3 was identified by all three comparative proteomics methods.

Hemoglobin subunit beta - The metalloprotein hemoglobin is responsible for oxygen transport and is the major protein contained within erythrocytes. Hemoglobin exists as a tetramer containing two beta chains and two alpha chains[187]. Hemoglobin subunit beta was not flagged as a potential biomarker in the cluster analysis of the ProteoSep™ 2D fractionation experiments – attesting to the efficient removal of hemoglobin during sample processing. This protein was, however, detected as a potential biomarker in analyses of the protein rich ProteoSep™ 1D pH fractions and the unfractionated protein samples from which hemoglobin was not removed.

Hemopexin is produced in the liver and found in plasma. This protein is responsible for trapping free heme in plasma as well as iron recycling in the liver[188, 189]. As with complement C3, this protein was identified by all three biomarker discovery strategies.

These selected blood proteins have all been well characterized and all appear to represent excellent candidates for blood biomarkers[190].

## 2.3 Conclusion

Proteomics-based methodologies can routinely identify hundreds to thousands of proteins in a biological system[191]. The overall objective of these types of studies are to identify candidate protein biomarkers for use as diagnostic tools – be it for a clinical diagnosis or, in the case of the current study, for use as a diagnostic marker for a specific body fluid in forensic casework.

In the current study, over one thousand proteins were identified utilizing three approaches: 1) cluster analysis of proteomic maps developed from ProteoSep™ 2D fractionation and peak analysis by mass spectrometry; 2) comparative proteomic analysis of protein-rich ProteoSep™ 1D pH fractions and analysis following mass spectrometry and; 3) comparative proteomic analysis of unfractionated body fluids following mass spectrometry. After careful review of the data, a database containing thirty body fluid-specific protein biomarkers was compiled. These include candidate biomarkers for the identification of semen, saliva, urine, peripheral blood, vaginal/menstrual fluids.

While these results are highly encouraging, it is important to stress that these protein biomarkers were identified by mapping the protein profiles of just five individuals per bodily fluid and thus these proteins can only be considered candidate biomarkers. For example, the possibility cannot be ignored that some candidate biomarkers might be secreted into non-target fluids, particularly with the less well-documented proteins associated with vaginal fluid. In fact, the published literature and proteomic data from the current study, itself, clearly suggest that few of the candidate biomarkers will be absolutely specific for a single body fluid. Because of this, a larger-scale study must be

completed in order to more rigorously confirm the degree to which these candidate protein biomarkers can be interpreted as sufficiently specific to a given body fluid as to warrant their use in a forensic context.

# Chapter 3:

# Biomarker Specificity Verification

## 3 Introduction

The traditional proteomics research and development pipeline is characterized by a methodical discovery phase to identify putative biomarkers followed by a high-throughput verification phase to assess the diagnostic potential of the biomarker candidates. Historically, verification in the clinical setting has been performed using immunoassays in order to generate accurate and reproducible quantitative measurements of specific proteins of interest[192, 193]. While the use of immunoassays is well established approach, a disconnect exists in the translation from biomarker candidates to clinical utility. This exists primarily due to the lack of commercially-available antibodies/reagents as well as the prohibitive cost of *de novo* immunoassay development. This problem is compounded by the fact that typical proteomics discovery projects may identify as few as ten or as many as fifty candidate proteins with potential clinical utility[76, 194]. Recent advances in targeted mass spectrometry offer an alternative approach to traditional antibody-based testing. Using ion-scanning techniques on quadrupole mass spectrometers, it has become possible to identify, as well as quantify, hundreds of protein biomarkers in a single analysis without the added cost of immunoassay design and characterization[195, 196].

In developing a verification platform for a body fluid-specific serological assay, it must be emphasized that, with any antibody-based assay, the results are "presumptive by definition". This is because the potential for antibody cross-reactivity with non-target molecules is difficult to predict *a priori* and can never be completely eliminated[197]. Thus, for a biomarker verification assay, a targeted mass spectrometry-based platform offers a lower cost, a shorter development time, as well as the inherent accuracy of

detection by mass spectrometry rather than indirect identification based on an intermolecular binding event.

For the verification outlined in this phase of the dissertation research, a quadrupole time-of-flight system will be employed to carry out the analyses. This instrument utilizes target ion "inclusion lists" that are generated based on the candidate biomarkers being verified. Each of these target ions are diagnostic tryptic peptides of our candidate biomarkers which the first quadrupole isolates. The isolated ions are then fragmented into smaller peptides in the following collision cell. These fragments then enter a TOF mass analyzer which yields highly accurate product ion spectra which is uniqe to each peptide [198, 199]. The bottom line is that the Q-TOF assay allows unprocessed biological stains to be directly injected and even low-abundance biomarkers can be reliably detected in a background of hundreds to thousands of higher-abundance proteins. The core objective for this phase of the current research was to verify the body fluid specificity and the consistency of expression in the general population of the previously identified candidate protein biomarkers. This was performed utilizing a Quadrupole Time-of-Flight instrument.

## 3.1 Materials and Methods

All research conducted under this project was reviewed, approved by the University of Denver Institutional Review Board (IRB) for research involving human subjects and conducted in full compliance with U.S. Federal Policy for the Protection of Human Subjects. In total, 100 subjects (50 males and 50 females) were recruited from the undergraduate population at the University of Denver. Female donors were

compensated $80 and males $20 for their participation in the research study. All research volunteers agreed and signed a letter of informed consent acknowledging they have received, read, and understood the protocols involved in the collection of samples. Additionally, all donated samples were given a random alphanumeric label to ensure confidentiality. Finally, males were instructed to refrain from ejaculation for 24 hours prior to sample collection and all donors were instructed to refrain from sexual intercourse for 5 days prior to the collection of urine, and vaginal fluids and menstrual blood for females.

### 3.1.1 Sample Collection and Preparation

An overview of the workflow performed from sample collection through biomarker identification is provided in **Figure 24.** Saliva, seminal fluid, urine, vaginal fluid, peripheral, and menstrual blood were self collected whereas peripheral blood samples were collected by laboratory staff. After collection saliva, seminal fluid, urine, vaginal fluid, peripheral, and menstrual blood samples were centrifuged to remove cellular material and quantified for total protein. Urine was processed further via a methanol precipitation cleanup step. Samples were divided into 10-15 equal volume aliquots and stored at -80°C before analysis.

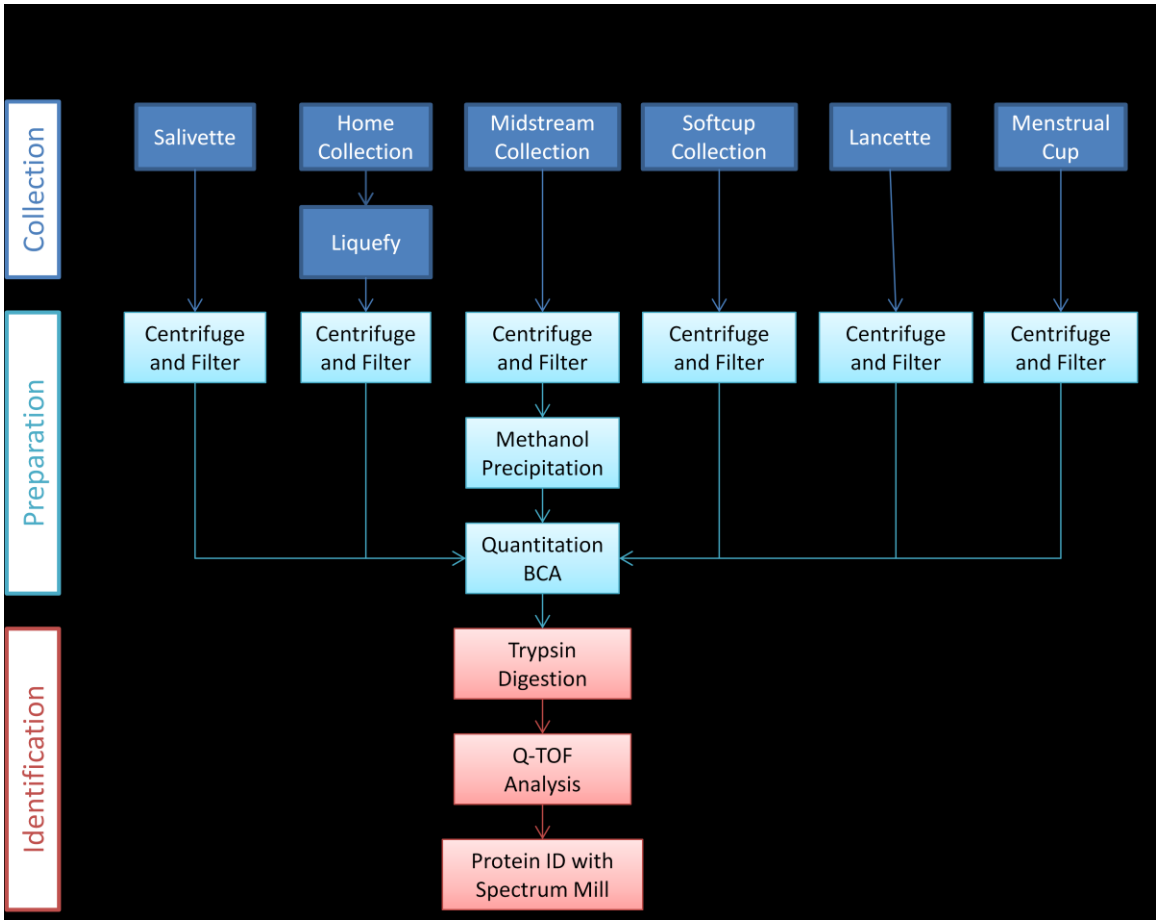



**Figure 24 – Biomarker verification workflow from sample collection to biomarker identification. Briefly, samples were collected and cellular debris removed via centrifugation and filtration. Samples were quantified for total protein and analyzed by mass spectrometry.**

### 3.1.1.1 Saliva

Saliva was self-collected as described in section 2.1.1.1 with slight modification. Samples were stored as aliquots after the 0.45 µm filtration step and no concentration by Amicon Ultra-15 centrifugal filtration was performed.

### 3.1.1.2 Seminal Fluid

Seminal fluid was collected as described in section 2.1.1.2

### 3.1.1.3 Urine

Urine was collected as described in section 2.1.1.2 with the following modifications. After the urine had been collected, it was centrifuged at 10,000 x g for 10 minutes at 4°C and then passed through a 0.45 μm filter to ensure cellular removal. The protein component of the urine was then precipitated from a 300 μL aliquot of the urine by the addition of 3 volumes (900 μL) of cold methanol and incubation for 10 min at -20ºC. The precipitated proteins were pelleted by centrifugation at 12.000 x g for 10 min at 4ºC after which the supernatant was discarded. One volume (300 μL) of -20ºC acetone was added to the pellet and the tube was allowed to stand for 10 minutes at min at 4ºC. The sample was then centrifuged at 12.000 x g for 10 min at 4ºC to repellet the protein precipitate. The supernatant was discarded and 100 μL of 25 mM ammonium bicarbonate was added and the tube was incubated at 37 ºC with gentle shaking for 10 minutes to resolublize the protein pellet.

### 3.1.1.4 Vaginal Fluid

Vaginal fluid collection employed an FDA-approved, over-the-counter, single-use latex-free, hypoallergenic cup SoftCup™ (Evofem, San Diego, CA). Donors were instructed to insert the SoftCup™ for one hour and then deposit the cup into a 50 mL sterile collection container. Donors were directed to refrigerate the sample until it could be transported to the lab within one hour. Upon receipt, 2-3 mL of phosphate buffered saline was added to the SoftCup™ to rehydrate and recover vaginal fluids from the sides of the device. The recovered liquid was transferred to two or more 1.65 mL

microcentrifuge tubes which were centrifuged at 10,000 x g for 10 minutes at 4°C. The supernatant was then passed through a 0.45 μm filter to ensure cellular removal.

#### 3.1.1.5 Peripheral Blood

Small volumes of peripheral blood were collected using the Unistik® 3 single use safety lancets (Owen Mumford, Marietta, GA) and the StatSampler EDTA collection vial (Statspin Technologies, Norwood, MA). Donors were instructed to wash their hands with disinfectant soap for two minutes under warm water. The donor's finger was then cleaned with an alcohol swab and punctured using the lancet on the side of the volunteer's ring finger to minimize discomfort. Blood was expressed and collected from the volunteer using the capillary tube from the StatSampler® collection kits. Once the capillary tube was full (200 μL) the volunteer was given a Band-Aid® to cover the puncture site. The StatSampler® capillary tube was then expelled into a gel barrier microcentrifuge tube and spun for 2 minutes at 1500 x g to recover the serum. Blood sera from StatSampler® collection kits were spun at 10,000 x g for 20 minutes at 4°C and passed through a 0.45 μm filter to ensure cellular removal.

#### 3.1.1.6 Menstrual Fluid

Menstrual fluid/blood was self-collected by study participants in the privacy of their homes. The collection protocol employed an FDA-approved, over-the-counter, latex-free, hypoallergenic DivaCup™ for the collection of menstrual flow. The donor was directed to insert the cup into the vagina during menses for a period of approximately one hour. The cup was then gently removed and the contents were poured into a sterile 50mL

conical tube. The donor was instructed to refrigerate the collected sample until delivered to the lab within one hour.

Menstrual blood was delivered in a sterile 50mL collection container. The blood was then transferred to a pink-cap 5 mL EDTA blood collection tube. The blood was rinsed with 2-3 mL of PBS and returned to a pink-cap EDTA blood tube. If necessary, due to the volume of material received, samples were divided between additional pink-cap EDTA blood tubes. Samples were allowed to stand for 5 minutes at room temperature and then centrifuged at 1500 x g for 10 minutes at room temperature to separate the menstrual serum from cells and debris. The menstrual serum was transferred to a fresh tube and centrifuged at 10,000 x g for 10 minutes at 4°C after which the supernatant was passed through a 0.45 µm filter to ensure cellular removal.

#### 3.1.1.7 Protein Quantitation with Pierce Micro BCA kit

Protein Quantitation was performed as described in section 2.1.1.10

### 3.1.2 Body Fluid Identification

#### 3.1.2.1 Trypsin Digestion and Peptide Cleanup

Trypsin digest and C-18 cleanup was performed as described in sections 2.1.3.2 and 2.1.3.3.

#### 3.1.2.2 Q-TOF MS/MS Acquisition

Mass spectrometry was performed on an Agilent Technologies HPLC-chip/MS system coupled to an Agilent 6510 Quadrupole Mass Spectrometer. The HPLC chip column used was a 150 mm 300 Å C18 analytical with a 160 nL enrichment column. Columns were equilibrated in 0.1% formic acid in water. A total of 500 ng of protein was

injected onto the column with the following run conditions: Buffer A (0.1% formic acid in water) and Buffer B (90% Acetonitile, 10% water, 0.1% formic acid). An initial 44-minute run employed a gradient of 3-36% Buffer B over 38 minutes. This was followed by 80% Buffer B from 40-44 min and then a 5-minute post-run equilibration with 3% Buffer A.

#### 3.1.2.3 Q-TOF Data Analysis

Data analysis was performed as described in 2.1.3.5

### 3.1.3 Preparation of Casework Samples

#### 3.1.3.1 Stains on Swabs and Substrates

Casework-type single body fluid samples were created by pipetting 50 µL of a desired body fluid, prepared as described in section 3.2.1, onto either standard cotton swabs or variable substrates (*e.g.* cotton, denim, leather ). Swabs and substrates were allowed to air dry in a biosafety hood for 40-60 minutes. Once dried, the cotton was removed from the swab or a cutting was removed from a substrate with a scalpel and placed into a 1.65 mL microcentrifuge tube. Proteins were extracted by adding 300 µL phosphate buffered saline to the swab and shaken at 37°C for 10 minutes. After incubation, the swab was placed into a spin basket and centrifuged for 30 seconds at 13,000 x g to recover all fluid. Samples were then quantified using Micro BCA protocol described in section 2.1.1.10 and processed as described in section 3.1.2.

### 3.1.3.2 Stains on Swabs with Contaminants

Biological fluids on casework type samples encountered in the lab may often be mixed with contaminants from the environment and/or chemicals used in the processing of the crime scene. In order to prepare samples to test the impact of these additives, swabs were prepared as described in section 3.1.3.1 with minor adjustment. Prior to addition of the body fluid of interest, swabs were dipped in various environmental contaminants (*e.g.*, bleach, Bluestar®, and spermicidal lubricant). The body fluid of interest was then applied to the swab and allowed to dry.

### 3.1.3.3 Mixed Stains on Swabs

Casework-type samples encountered in the lab may often contain mixtures of more than one biological fluid. In order to generate samples to evaluate the ability to detect target biomarkers representing more than one type of stain, swabs were prepared similarly to section 3.1.3.1 with minor adjustment. For volume-based mixtures, two equal volumes of a body fluid of interest (25 µL each) were added to a swab which was then dried.

## 3.2 Results and Discussion

In order to validate the consistency of detection and target body fluid specificity of the candidate biomarkers identified during the discovery phase of the current research, fifty single-source samples of saliva, vaginal fluid, seminal fluid, urine, peripheral blood, and menstrual fluid/blood were collected and run on a Q-TOF mass spectrometer operating in targeted MS/MS mode. Additionally, a variety of casework-type samples

were prepared for standard and sensitivity analyses which included dried biological stains on cotton swabs, a variety of commonly encountered substrates, stains mixed with typical chemical contaminants encountered in a forensic context, mixtures of different body fluids.

### 3.2.1 Sample Collection and Analysis

Fifty single-source samples were collected from donors for analysis by targeted mass spectrometry. These were collected and quantified as previously described. In order to appropriately represent the conditions under which samples will be analyzed in the context of a caseworking laboratory, no prefractionation or other exceptional pretreatments were employed at this stage. For example, menstrual blood was not treated with HemogloBind™ to remove hemoglobin. Additionally, neither menstrual fluids nor peripheral blood were passed through an immunodepletion column to remove abundant and thus potentially interfering proteins. In order to assess the utility of the candidate biomarkers for use with more challenging type of evidentiary material, a variety of casework-type samples prepared as described above were also analyzed during this phase of the research.

### 3.2.2 Assay Method Development

While targeted mass spectrometry based assays are highly reproducible and robust significant method development is required to design the analytical method. Due to the complexity of an analytical method which can process six biological matrixes simultaneously, the assay development activities were divided into a series of smaller objectives. These included: protein biomarker selection, target peptide selection, single-fluid multiplex design followed by the six-fluid multiplex assay design (**Figure 25**).

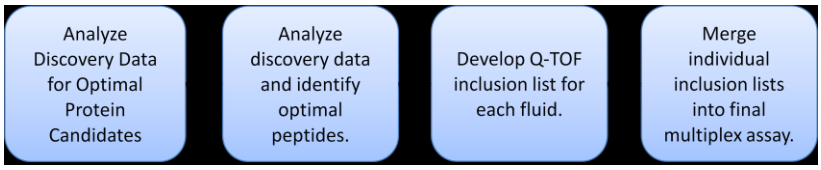



Figure 25 – Method development workflow from peptide selection through multiplex assay design.

By carefully reviewing the data generated during the initial biomarker discovery phase of the project, three to five of the most promising candidate protein biomarkers were selected for the inclusion in the multiplex assay for each body fluid of interest. Once the proteins were selected, optimal tryptic peptides for the detection of each protein were identified by mass spectrometry. Additionally, the analytical run parameters were optimized and a single-fluid multiplex assay was developed for evaluation. Finally, successfully-developed assays for each individual body fluid were merged and again evaluated for performance prior to analysis of the fifty individual fluids.

#### 3.2.2.1 Protein Selection

The discovery phase of the current research identified numerous candidate biomarkers with potential utility for body fluid differentiation. From these candidates, the most discriminating and reliable protein targets must be identified for integration into the targeted Q-TOF assay. The optimal proteins are those that were identified using multiple biomarker discovery approaches and whose tissue specificity was also supported by a review of the proteomic and biomedical literature.

In general, candidate seminal fluid biomarkers were identified consistently by all three comparative proteomic strategies. Only two proteins were not – epididymal secretory protein E1 was not identified as a potential biomarker by cluster analysis of the proteomic maps from the ProteoSep™ 2D fractions and prolactin inducible protein was

not identified in the comparative analysis of mass spectrometry data from unfractionated protein samples. All of the candidate biomarkers except for prolactin inducible protein were included in the development of the targeted assay for seminal fluid. The decision not to included prolactin inducible protein was that it has been reproducibly identified in several salivary proteomics studies and does not appear to be sufficiently specific to human seminal fluid[200, 201].

The selection of saliva biomarkers was slightly more difficult as no biomarker was detected by all three comparative proteomic approaches employed in the current study. In this case, more weight was given to proteins identified through analysis of unfractionated protein samples as it was felt that this would better ensure that the selected proteins were abundant enough to be reliably detected. Included as candidate biomarkers for the saliva assay, therefore, were statherin, histatin-1 and cystatin-SA which were all identified by two of the comparative proteomics approaches, where cystatin D and submaxillary gland androgen regulated protein were only identified by the comparative analysis of data from the unfractionated protein samples. Salivary acidic proline-rich phosphoprotein 1/2 was omitted due to the fact that it was only detectable as being potentially unique by the cluster analysis of the ProteoSep™ 2D fractions.

The selection of the most promising vaginal fluid proteins was also challenging as only one candidate biomarker, cornulin, was identified as unique by more than one comparative proteomics strategy. Additionally, according to a thorough literature review, the vaginal candidates appeared to have the least overall specificity. For example, cornulin, involucrin, Ly6/PLAUR domain-containing protein 3, periplakin and suprabasin were all reported to be associated with the epithelium. Thus, it can be assumed

that similar tissues may contain these proteins as well. Therefore, it is very possible that other non-keratinized tissues such as the mouth, esophagus, male reproductive tract, as well as the intestines may contain these proteins. As an example, cornulin has been linked to cancer involving esophageal tissues[202]. Table 4 provides a summary of the findings from four independent studies where candidate biomarkers of vaginal fluids were identified in salivary proteomic studies.

**Table 4 – Salivary Proteomic Studies where Candidate Biomarkers of Vaginal Fluid were Identified.**

| Body Fluid | Protein | Accession Number | Study 1[203] | Study 2[204] | Study 3[205] | Study 4[206] |
|---|---|---|---|---|---|---|
| Vaginal Secretions | Cornulin | Q9UBG3 | | | | |
| | IgGFc-binding protein | Q9Y6R7 | | | | |
| | Involucrin | P07476 | | | √ | |
| | Ly6/PLAUR domain-containing protein 3 | O95274 | | | | |
| | Matrix metalloproteinase-9 | P14780 | | | √ | |
| | Mucin-5B | Q9HC84 | √ | √ | √ | |
| | Neutrophil gelatinase-associated lipocalin | P80188 | | | √ | |
| | Periplakin | O60437 | | | √ | |
| | Suprabasin | Q6UWP8 | | | | |
| | Vimentin | P08670 | | | √ | |

Based on the results presented in Table 4, several of the proposed biomarkers of vaginal fluids have also been identified in saliva proteome projects. These results must be taken into consideration when designing the final targeted Q-TOF assay. However, it should be noted that none of these markers were identified in any saliva samples analyzed during this dissertation research. Based on these findings, cornulin, IgGFc-binding protein, Ly6/PLAUR domain-containing protein 3, as well as suprabasin were selected for inclusion in the targeted Q-TOF assay for vaginal fluid. Additionally, matrix metalloproteinase-9, mucin-5B, as well as neutrophil gelatinase-associated lipocalin were

added to the assay. While these markers have been identified in saliva, all four studies utilized extensive two dimensional fractionation strategies to detect proteins with very low expression levels. Thus, while several of these proteins may be expressed in both fluids, the quantity of protein that can be found may in one tissue versus another may still make it possible to discerng saliva from vaginal fluid. Neither vimentin nor involucrin nor periplaklin were detected in comparisons of proteomic data from unfractionated body fluid samples. Accordingly, they were removed from consideration for the targeted assay development.

Urine biomarkers, osteopontin as well as uromodulin, were both selected for integration into the targeted assay. Both markers appear highly specific to urine and both proteins were identified in nearly all the biomarker discovery strategies employed.

Calpastatin was the only protein identified as a potential biomarker of menstrual fluid. Unfortunately, this protein was only detected in the comparative analysis of proteins from the ProteoSep™ 1D pH fractions. Because of the potential value of identifying a reliable biomarker of menstrual fluid, an attempt was made to locate the protein manually in the dataset generated by the Q-TOF analyses of the unfractionated protein samples described in Chapter 2. Using the results from the ProteoSep™ 1D pH fraction where the protein was originally identified, two peptides GTVPDDAVEALADSLGK and LAAAISEVVSQTPASTTQAGAPP were used in an attempt to manually identify the protein in the raw data from the analysis of the unfractionated protein samples. By using the exact masses of these tryptic peptides, the extracted ion chromatograms (EIC) for the calpastatin peptides and a reference peptide from Hemopexin (NFPSPVDAAFR) were searched (**Figure 26).**

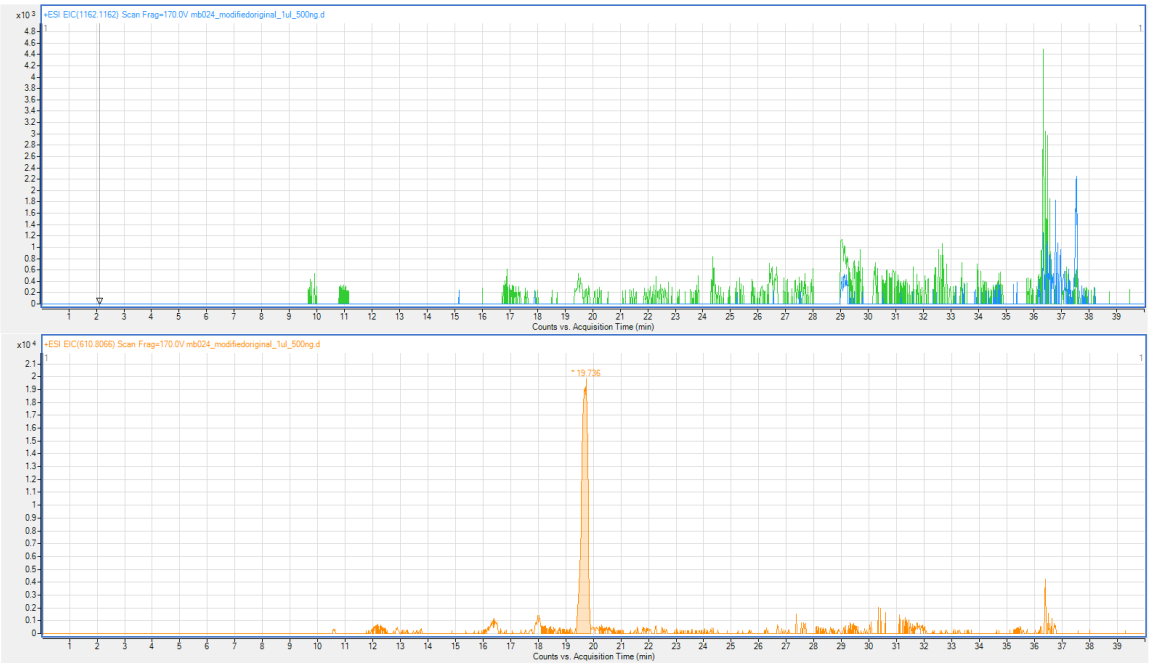



**Figure 26 –Extracted ion chromatogram for Calpastatin (top) and Hemopexin (bottom). Hemopexin was reliably detectable in the menstrual blood sample whereas the Calpastatin peptide could not be manually located.**

At 19.5 minutes, a clear peak exists for hemopexin (which was also confirmed via a database search) whereas the data region that should contain the calpastatin peptides only revealed only noise. This result, which was observed reproducibly across all menstrual samples analyzed, suggests that the protein is either not present in the samples or that omitting the preparation steps (*i.e.*, HemogloBind™ and immunodepletion) may have made the protein undetectable in the assay. Although calpastatin may not be a reliable marker for menstrual blood, some of the vaginal-fluid associated biomarkers were also identified in menstrual blood samples. Therefore, the best option currently available may be to use these as indicators of the vaginal origin of an unknown sample.

Similar to seminal fluid, peripheral blood was relatively straightforward as all the candidate biomarkers were reliably identified by each of the comparative proteomic

assays. Accordingly, complement C3, hemoglobin subunit beta, as well as hemopexin were included in the development of the targeted assay.

#### 3.2.2.2 Peptide Selection

The next step in assay development was the selection of optimal tryptic peptides for targeting as representative/diagnostic peptides for each of the candidate biomarker proteins. Ideal peptides were those which: (1) were detected consistently across multiple experiments, with additional weight given to peptides identified in the unfractionated experiments; (2) had no post-translational modifications that would alter the mass-charge ratio; (3) were of high abundance so as to facilitate robust detection; (4) are human specific in their primary amino acid sequence.

Following the above criteria, each protein was examined in order to select two to three diagnostic peptides. In some cases, this was a relatively straightforward process. The presence of statherin, for instance, is based on the detection and identification of only one peptide. While two to three diagnostic peptides per protein are typically optimal for targeted assays, this single peptide (shown in red below) represents approximately 50% of the entire sequence of the statherin protein (MKFLVFAFILALMVSMIGADSSEEK FLRRIGRFGYGYGPYQPVPEQPLYPQPYQPQYQQYTF) and it has been consistently identified in samples from multiple individuals.

The selection of optimal peptides for other biomarker proteins, such as cystatin-SA, was relatively straightforward as there were only a select few peptide sequences that were consistently detected in the Q-TOF analysis of the unfractionated protein samples. Table 5 shows the results obtained from three independent analyses of unfractionated

salivary protein samples. The cystatin-SA derived peptides SQPNLDTCAFHEQPELQKK, QLCSFQIYEVPWEDR, and IIEGGIYDADLNDER were consistently detected in all three runs and thus represent optimal diagnostic peptides for their parent protein.

**Table 5 – Cystatin SA identification Results for Three QTOF Injections of Unfractionated Saliva Proteins.**

| Sample | Peptide | Modifications | Spectrum Intensity | Score | SPI | Z | m/z | RT |
|---|---|---|---|---|---|---|---|---|
| Unfractionated Salivary Protein Run 1 | IIEGGIYDADLNDER | | 2.30E+05 | 18.91 | 79.6 | 2 | 846.906 | 11.9 |
| | QLCSFQIYEVPWEDR | C:Carbamidomethylation | 1.74E+05 | 16.73 | 92.5 | 2 | 985.458 | 19.8 |
| | QLCSFQIYEVPWEDR | C:Carbamidomethylation | 1.57E+05 | 16.47 | 85.2 | 3 | 657.309 | 19.8 |
| | SQPNLDTCAFHEQPELQKK | C:Carbamidomethylation | 1.07E+07 | 15.69 | 74.7 | 4 | 568.282 | 10 |
| | SQPNLDTCAFHEQPELQKK | C:Carbamidomethylation | 2.11E+06 | 9.02 | 75.6 | 3 | 757.369 | 10 |
| Unfractionated Salivary Protein Run 2 | IIEGGIYDADLNDER | | 2.61E+05 | 23.06 | 91 | 2 | 846.91 | 11.7 |
| | QLCSFQIYEVPWEDR | C:Carbamidomethylation | 1.28E+05 | 13.88 | 74.6 | 2 | 985.456 | 19.7 |
| | QLCSFQIYEVPWEDR | C:Carbamidomethylation | 1.28E+05 | 14.56 | 74.6 | 2 | 985.456 | 19.8 |
| | SQPNLDTCAFHEQPELQKK | C:Carbamidomethylation | 3.62E+05 | 10.77 | 78.5 | 3 | 757.369 | 9.9 |
| Unfractionated Salivary Protein Run 3 | ALHFVISEYNK | | 1.09E+06 | 9.14 | 65.2 | 3 | 440.901 | 18.1 |
| | IIEGGIYDADLNDER | | 5.17E+05 | 12.78 | 70.7 | 2 | 846.905 | 11.8 |
| | QLCSFQIYEVPWEDR | C:Carbamidomethylation | 3.65E+05 | 16.45 | 86.5 | 3 | 657.307 | 19.8 |
| | SQPNLDTCAFHEQPELQKK | C:Carbamidomethylation | 5.12E+06 | 8.82 | 50.8 | 3 | 757.364 | 10 |
| | SQPNLDTCAFHEQPELQKK | C:Carbamidomethylation | 5.51E+06 | 10.96 | 55.3 | 3 | 757.364 | 10 |

In other cases, peptide selection was more complicated due to the sheer number of potential targets that were identified. Large and highly abundant proteins such as semenogelin and hemoglobin produced more than 30 viable peptide targets. In these cases, the most abundant/non-modified peptides was selected for the targeted assay.

### 3.2.2.3 Q-TOF Method Development

With proteins and peptides selected, an optimal analytical gradient needed to be designed for the multiplex assay. Once an optimal gradient was chosen, replicates for each body fluid were then run in order to identify the exact retention time at which each target peptide eluted into the mass spectrometer. Using the mass-to-charge ratio and retention time data, a Q-TOF inclusion list was generated that made it possible to selectively target each peptide for each biomarker for each body fluid of interest.

The original Q-TOF analysis of unfractionated protein samples employed a 44-minute analytical run. While this method was successful in identifying numerous biomarker candidates for each body fluid that was analyzed, it was postulated that the chromatographic separation could be improved to spread out the elution time of the peptides into the mass spectrometer. Using the original method as a starting point four similar gradients were tested as well as a shorter 22-minute method. (**Table 6**).

**Table 6 – Four Analytical Gradients Tested.**

| **Gradient 1** | | **Gradient 2** | | **Gradient 3** | | **Gradient 4** | |
|---|---|---|---|---|---|---|---|
| Original 40 minute | | Short 22 minute | | Modified 40 minute | | Modified 40 minute | |
| Time | % Organic | Time | % Organic | Time | % Organic | Time | % Organic |
| 0 | 3 | 0 | 3 | 0 | 3 | 0 | 3 |
| 1 | 10 | 1 | 10 | 1 | 6 | | |
| 33 | 45 | 20 | 45 | 33 | 38 | 33 | 38 |
| 35 | 80 | 20.1 | 80 | 35 | 80 | 35 | 80 |
| 40 | 80 | 23 | 80 | 40 | 80 | 40 | 80 |
| 40.1 | 3 | 23.1 | 3 | 40.1 | 3 | 40.1 | 3 |

In an attempt to speed sample throughput, a 22-minute gradient (gradient 2) was built and tested against all target fluids. For most fluids, this offered a substantial increase

in speed with no chromatographic or detection issues when identifying the target proteins. However, problems arose with saliva using the faster throughput methods. The saliva proteome consists of a considerable number (~50% of the total protein content) of small, hydrophilic, low molecular weight proteins[207]. Using reverse-phase conditions hydrophilic peptides are retained weakly whereas hydrophobic peptides are retained strongly and elute later in the run. This translated into column overloading issues (**Figure 27**) in which the majority of the salivary peptides overloaded the column and the mass spectrometer at the front end of the run.

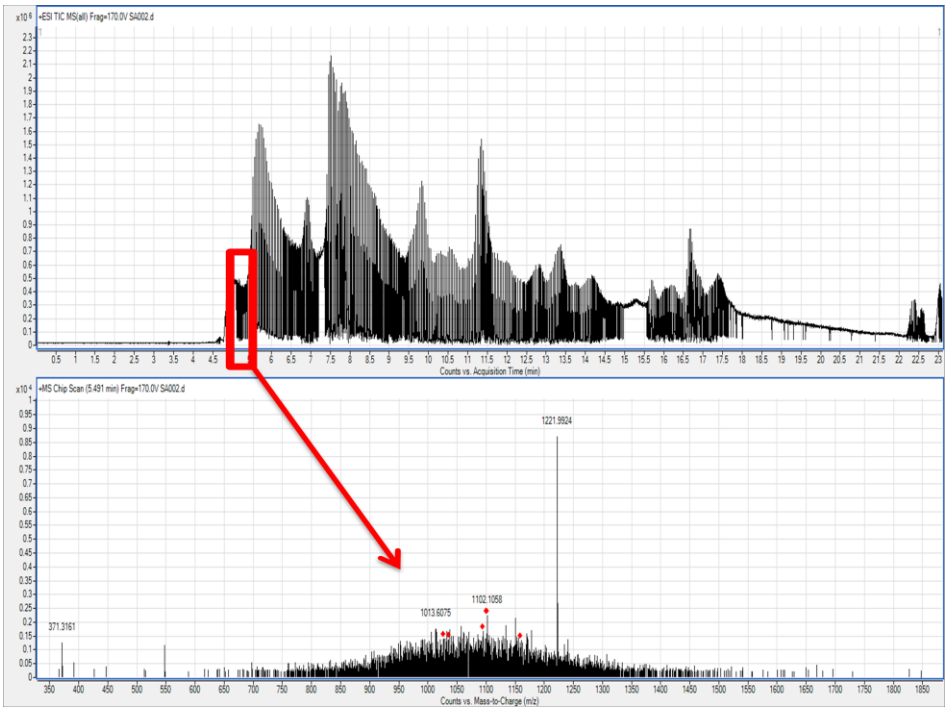



**Figure 27 – An irresolvable MS scan reflective of column overloading by small, hydrophilic, low molecular weight salivary proteins on a fast 22-minute gradient.**

Poor chromatograph and column overloading result in poor peak shapes, non-reproducible retention times, as well as ion suppression. Because of this overloading issue, the original 40-minute gradient was revisited and optimized. **Figure 28** represents

three gradients tested for replicate saliva injections. Gradient 1 is the original one used for the unfractionated experiments. It is clear that from approximately 25 minutes to 40 minutes very few peptides are eluting into the mass spectrometer essentially "wasting" instrument run time. To address this, two additional methods were created and tested to improve the overall chromatography. From the results, gradient 3 (**Center, Figure 28**) was selected as the optimal method because the overloading issues were alleviated with the greatest peptide distribution of any of the methods tested.


**Figure 28 – Three 40-minute gradients tested for chromatography optimization. Gradient 1 (top) is the original used for the unfractionated experiments where Gradient 3 (center) and Gradient 4 (bottom) represent modifications of**

this gradient. Gradient 3 (center) was selected due to optimal peptide distribution across the chromatographic separation.

With a gradient selected, two samples from each fluid were run in order to identify the exact retention time in which each peptide eluted into the system. Once located, peptide m/z, retention time, as well as charge was used to compile an inclusion list for each body fluid – a table which the instrument uses to target specific "diagnostic" peptide targets (**Tables 7A-E**). For each body fluid, the inclusion list describes the candidate proteins being assayed, their corresponding target peptides, retention times, and exact mass-to-charge ratios. The target ion inclusion lists for each "singleplex" body fluid assay were then tested before being combined into the final multiplex assay. This process resulted in the development of the final multiplex assay illustrated in **Figure 29**.

Table 7A – Peripheral Blood Peptide Inclusion List

| Protein Biomarker | Target Biomarker Peptide | Prec m/z | Z | Retention Time (min) |
|---|---|---|---|---|
| Hemoglobin subunit beta | LLVVYPWTQR | 637.8732 | 2 | 20.8 |
| | VVAGVANALAHKYH | 483.9403 | 3 | 12.4 |
| | GTFATLSELHCDK | 739.8534 | 2 | 14.4 |
| Complement C3 | TMQALPYSTVGNSNNYLHLSVLR | 860.1049 | 3 | 23.43 |
| | VYAYYNLEESCTR | 834.3748 | 2 | 14.77 |
| | VFLDCCNYITELR | 851.9012 | 2 | 21.4 |
| Hemopexin | NFPSPVDAAFR | 610.8107 | 2 | 18.3 |
| | YYCFQGNQFLR | 748.474 | 2 | 18.3 |

**Table 7B – Saliva Peptide Inclusion List**

| Protein Biomarker | Target Biomarker Peptide | Prec m/z | Z | Retention Time (min) |
|---|---|---|---|---|
| Cystatin SA | IIEGGIYDADLNDER | 846.9146 | 2 | 17.5 |
| | SQPNLDTCAFHEQPELQKK | 757.3652 | 3 | 14.5 |
| | QLCSFQIYEVPWEDR | 985.4583 | 2 | 26.3 |
| Cystatin D | SQPNLDNCPFNDQPK | 887.3956 | 2 | 13.8 |
| | TLAGGIHATDLNDK | 475.9143 | 3 | 12.2 |
| Submaxillary gland androgen regulated protein | GPYPPGPLAPPQPFGPGFVPPPPPPPYGPGR | 776.1541 | 4 | 28.8 |
| | IPPPPPAPYGPGIFPPPPPQP | 710.7205 | 3 | 25.25 |
| Statherin | FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF | 1215.8982 | 3 | 28.4 |
| Histatin-1 | EFPFYGDYGSNYLYDN | 982.4056 | 2 | 26.27 |

**Table 7C – Seminal Fluid Peptide Inclusion List**

| Protein Biomarker | Target Biomarker Peptide | Prec m/z | Z | Retention Time (min) |
|---|---|---|---|---|
| Semenogelin 1 | KQGGSQSSYVLQTEELVANK | 722.7071 | 3 | 18.5 |
| | DIFTTQDELLVYNK | 842.9264 | 2 | 23.45 |
| Semenogelin 2 | DVSQSSISFQIEK | 734.3714 | 2 | 18.31 |
| | DIFTTQDELLVYN | 842.9264 | 2 | 24.1 |
| PSA | VMDLPTQEPALGTTCYASGWGSIEPEEFLTPK | 1175.5551 | 3 | 30.5 |
| | AVCGGVLVHPQWVLTAAHCIR | 586.8051 | 4 | 24.2 |
| | LSEPAELTDAVK | 636.8399 | 2 | 16.75 |
| Prostatic Acid Phosphatase | ELSELSLLSLYGIHK | 567.9866 | 3 | 28.5 |
| | FQELESETLKSEEFQK | | | 16.3 |
| | SPIDTFPTDPIK | 665.8475 | 2 | 19.35 |
| Glycodelin | VHITSLLPTPEDNLEIVLHR | 574.8187 | 4 | 27.3 |
| | VLVEDDEIMQGFIR | 555.2807 | 3 | 25.6 |
| Epididymal secretory protein E1 | AVVHGILMGVPVPFPIPEPDGCK | 810.7632 | 3 | 32.2 |
| | EVNVSPCPTQPCQLSK | 922.4355 | 2 | 12.8 |

### Table 7D – Urine Peptide Inclusion List

| Protein Biomarker | Target Biomarker Peptide | Prec m/z | Z | Retention Time (min) |
|---|---|---|---|---|
| Osteopontin | GDSVVYGLR | 927.9524 | 2 | 13 |
| | QLYNKYPDAVATWLNPDPSQK | 816.7445 | 3 | 23.8 |
| | AIPVAQDLNAPSDWDSR | 927.9533 | 2 | 20.5 |
| Uromodulin | VLNLGPITR | 491.8078 | 2 | 18.2 |
| | STEYGEGYACDTDLR | 868.8572 | 2 | 13.1 |
| | DGPCGTVLTR | 538.2658 | 2 | 10.21 |

### Table 7E – Vaginal Fluid / Menstrual Fluid Peptide Inclusion List

| Protein Biomarker | Target Biomarker Peptide | Prec m/z | Z | Retention Time (min) |
|---|---|---|---|---|
| Mucin 5B/Cervical | GYQVCPVLADIECR | 840.3965 | 2 | 22.4 |
| | AQAQPGVPLGELGQVVECSLDFGLVCR | 967.1501 | 3 | 36.51 |
| | AAGGAVCEQPLGLECR | 844.3968 | 2 | 14.8 |
| Cornulin | ISPQIQLSGQTEQTQK | 893.4706 | 2 | 13.4 |
| | TLSESAEGACGSQESGSLHSGASQELGEGQR | 1036.125 | 3 | 21.35 |
| IgGFc-binding protein | APGWDPLCWDECR | 831.3529 | 2 | 25.3 |
| | SLAAYTAACQAAGVAVKPWR | 697.6999 | 3 | 22.2 |
| | AGCVAESTAVCR | 640.7912 | 2 | 640.79 |
| Ly6/PLAUR containing protein 3 | DGVTGPGFTLSGSCCQGSR | 971.925 | 2 | 16.02 |
| | GLDLHGLLAFIQLQQCAQDR | 766.0686 | 3 | 32.2 |
| | GCVQDEFCTR | 636.2653 | 2 | 11.24 |
| Matrix metallo-proteinase-9 | GSRPQGPFLIADKWPALPR | 527.2981 | 4 | 25.22 |
| Neutrophil gelatinase-associated lipocalin | SYPGLTSYLVR | 628.3402 | 2 | 21.2 |
| | TFVPGCQPGEFTLGNIK | 622.3172 | 3 | 22.6 |
| Suprabasin | ALDGINSGITHAGR | 461.2476 | 3 | 12.4 |
| | LGQGVNHAADQAGKEVEK | 617.652 | 3 | 9.45 |

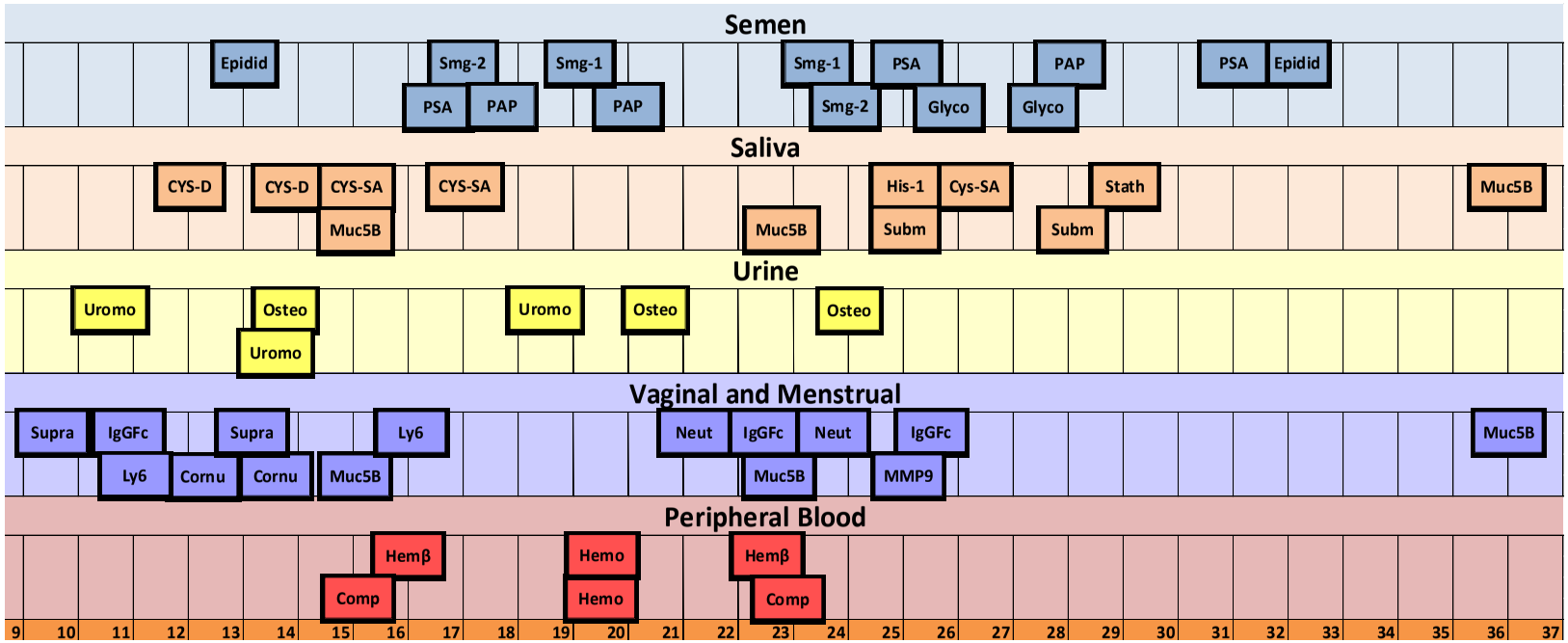



Figure 29 - Precursor ions employed in the Q-TOF body fluid multiplex assay. Ions are presented in order of elution (*i.e.*, retention time) and are segregated by body fluid specificity.

Analysis of a representative urine sample is shown in **Figure 30**. Chromatography and database search results revealed the identification of the two targeted biomarkers for urine (osteopontin and uromodulin) with no non-urine associated proteins detected.

Results obtained with samples of the other five body fluids also proved to be of high quality and specificity (**Figures 31-35**). Specifically, the analysis of a representative seminal fluid sample resulted in the detection of semenogelin I/II, prostate specific antigen, epididymal secretory protein E1, and prostatic acid phosphatase. No non-seminal fluid associated proteins were detected (**Figure 31**).

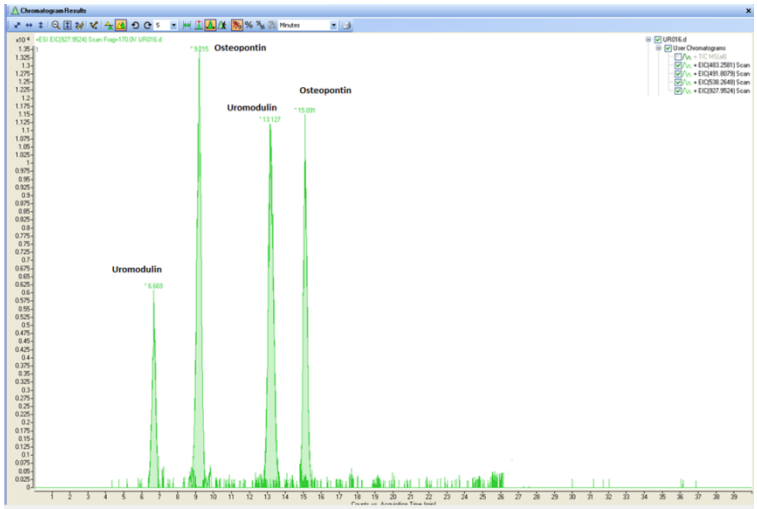
The multiplex assay of a representative saliva sample produced results that revealed the presence of mucin 5B, cystatin SA, cystatin D, submaxillary gland androgen regulated protein, statherin and histatin-1. Here again, no unexpected non-saliva associated protein was detected (**Figure 32**). It should be noted that while mucin 5B was proposed to be a vaginal-associated protein.

Analysis of a vaginal fluid sample revealed the presence of the targeted markers cornulin, IgGFc-binding protein, Ly6/PLAUR containing protein 3, neutrophil gelatinase-associated lipocalin, suprabasin, and matrix metallo-proteinase-9. No unanticipated non-vaginal associated proteins were detected (**Figure 33**). As noted above, mucin 5B is a vaginal secretion- associated protein; but was identified in saliva as well. Similarly, cornulin, also appears in menstrual blood.

The multiplex assay of a representative menstrual blood sample resulted in the detection of the target ions for complement C3, hemoglobin subunit beta, hemopexin, and

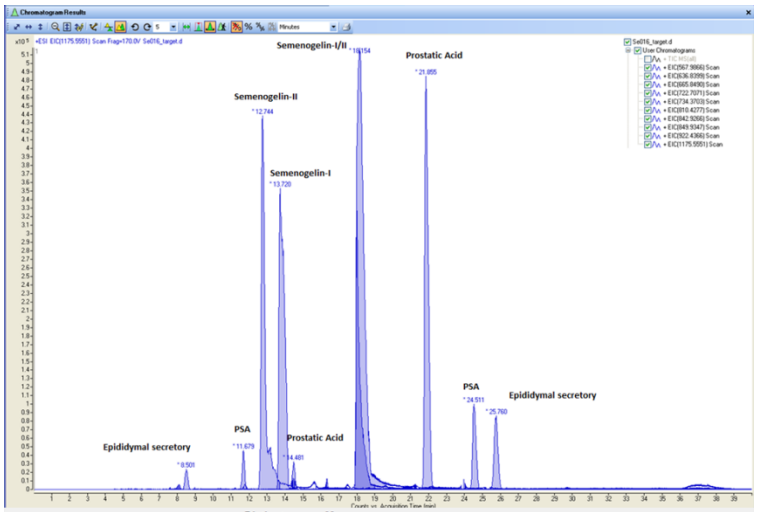
cornulin (**Figure 34**). Because menstrual blood contains peripheral blood as a major component, it was anticipated that all of the peripheral blood markers would also appear in this sample. In addition to the peripheral blood proteins, cornulin was also detected in this sample. This biomarker was initially employed as a prospective vaginal fluid-specific marker. It is not clear whether its detection in menstrual blood is the result of mixing between the menstrual blood and vaginal fluid during collection or if cornulin is also a component of menstrual blood. In either event, the ability to detect cornulin in these cases has potential utility for identifying vaginal sourced samples.

Finally, the analysis of a representative peripheral blood sample resulted in the unambiguous detection of complement C3, hemoglobin subunit beta and hemopexin (**Figure 35**). Aside from the expected occurrence of these proteins in menstrual blood, these proteins were not detected in any other body fluid.



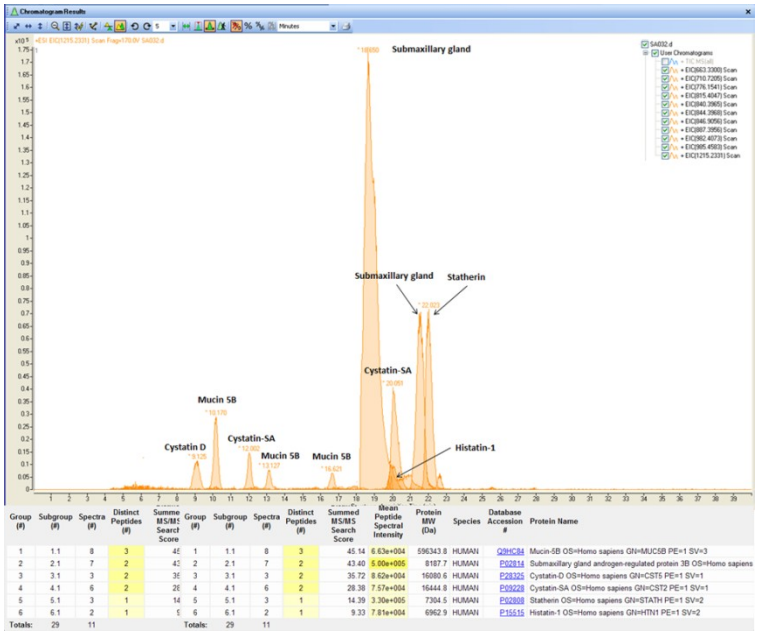

| Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Distinct Summed MS/MS Search Score | Mean Peptide Spectral Intensity | Protein MW (Da) | Species | Database Accession # | Protein Name |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | 2 | 2 | 33.30 | 1.33e+005 | 35422.9 | HUMAN | 2 | Osteopontin |
| 2 | 2.1 | 3 | 2 | 26.52 | 8.64e+004 | 69761.4 | HUMAN | 1 | Uromodulin |
| Totals: | 5 | 4 | | | | | | | |

**Figure 30 - TOP - Raw Q-TOF multiplex data indication of a urine hit. BOTTOM – Database search results indicate high-confidence confirmatory identification: Human urine. All targeted urine biomarkers were detected. No biomarkers indicative of other body fluids were detected**



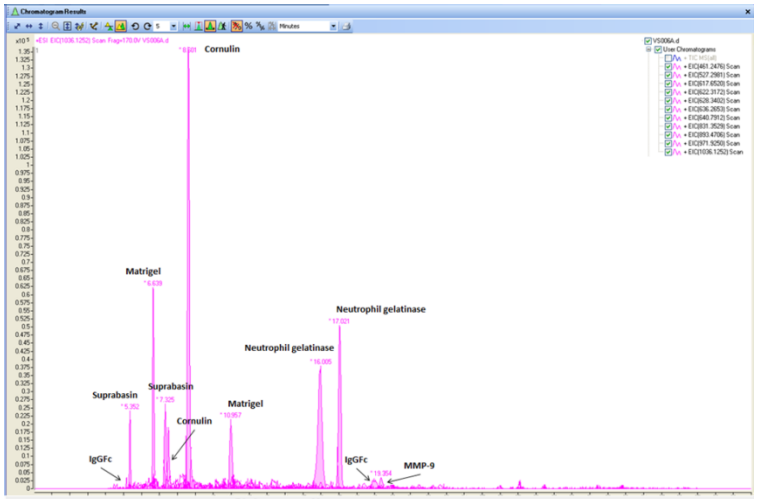

| Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Distinct Summed MS/MS Search Score | Mean Peptide Spectral Intensity | Protein MW (Da) | Species | Database Accession # | Protein Name |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | 5 | 2 | 48.50 | 1.09e+007 | 65444.7 | HUMAN | 7 | Semenogelin-2 |
| 2 | 2.1 | 6 | 2 | 42.59 | 1.04e+007 | 52131.1 | HUMAN | 6 | Semenogelin-1 |
| 3 | 3.1 | 2 | 2 | 39.91 | 3.66e+005 | 28741.5 | HUMAN | 3 | Prostate-specific antigen |
| 4 | 4.1 | 3 | 2 | 39.41 | 1.29e+006 | 16570.3 | HUMAN | 5 | Epididymal secretory protein E1 |
| 5 | 5.1 | 4 | 2 | 32.63 | 1.32e+006 | 44566.4 | HUMAN | 4 | Prostatic acid phosphatase |
| Totals: | 20 | 10 | | | | | | | |

**Figure 31 TOP – Raw Q-TOF multiplex data indication of a seminal fluid hit. BOTTOM – Database search results indicate high-confidence confirmatory identification: Human seminal fluid. All targeted seminal fluid biomarkers were detected. No biomarkers indicative of other body fluids were detected.**



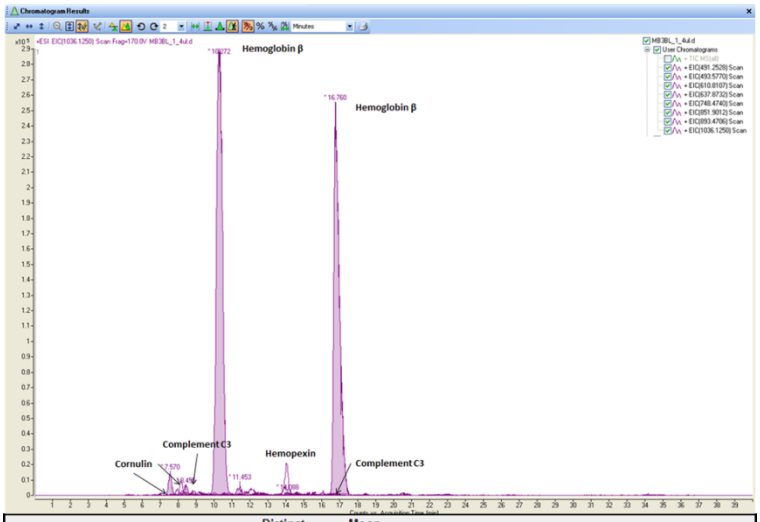

| Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Summed MS/MS Search Score | Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Summed MS/MS Search Score | Mean Peptide Spectral Intensity | Protein MW (Da) | Species | Database Accession # | Protein Name |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | 8 | 3 | 45 | 1 | 1.1 | 8 | 3 | 45.14 | 6.63e+004 | 596343.8 | HUMAN | Q9HC84 | Mucin-5B OS=Homo sapiens GN=MUC5B PE=1 SV=3 |
| 2 | 2.1 | 7 | 2 | 43 | 2 | 2.1 | 7 | 2 | 43.40 | 5.00e+005 | 8187.7 | HUMAN | P02814 | Submaxillary gland androgen-regulated protein 3B OS=Homo sapiens |
| 3 | 3.1 | 3 | 2 | 36 | 3 | 3.1 | 3 | 2 | 35.72 | 8.62e+004 | 16080.6 | HUMAN | P28325 | Cystatin-D OS=Homo sapiens GN=CST5 PE=1 SV=1 |
| 4 | 4.1 | 6 | 2 | 28 | 4 | 4.1 | 6 | 2 | 28.38 | 7.57e+004 | 16444.8 | HUMAN | P09228 | Cystatin-SA OS=Homo sapiens GN=CST2 PE=1 SV=1 |
| 5 | 5.1 | 3 | 1 | 14 | 5 | 5.1 | 3 | 1 | 14.39 | 3.30e+005 | 7304.5 | HUMAN | P02808 | Statherin OS=Homo sapiens GN=STATH PE=1 SV=2 |
| 6 | 6.1 | 2 | 1 | 9 | 6 | 6.1 | 2 | 1 | 9.33 | 7.81e+004 | 6962.9 | HUMAN | P15515 | Histatin-1 OS=Homo sapiens GN=HTN1 PE=1 SV=2 |
| Totals: | | 29 | 11 | | Totals: | | 29 | 11 | | | | | | |

**Figure 32 - TOP - Raw Q-TOF multiplex data indication of a saliva hit. BOTTOM – Database search results indicate high-confidence confirmatory identification: Human saliva. All targeted saliva biomarkers were detected. No biomarkers indicative of other body fluids were detected.**



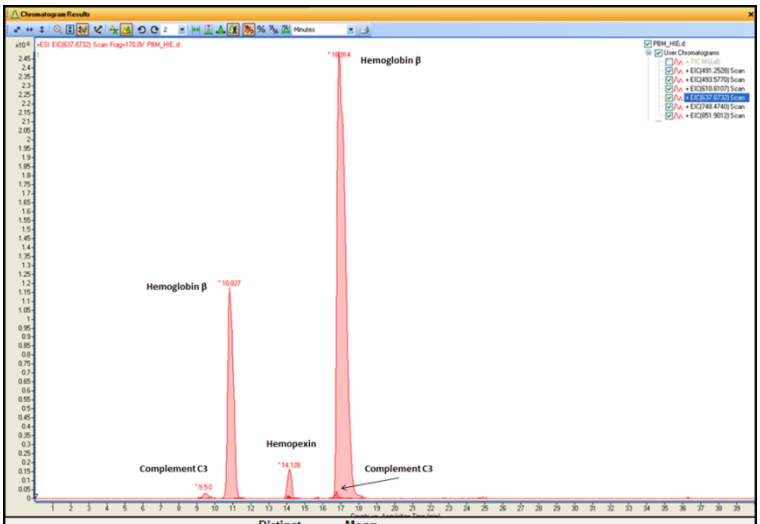

| Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Distinct Summed MS/MS Search Score | Mean Peptide Spectral Intensity | Protein MW (Da) | Species | Database Accession # | Protein Name |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | 8 | 2 | 41.29 | 2.12e+006 | 22588.2 | HUMAN | 18 | Neutrophil gelatinase-associated lipocalin |
| 2 | 2.1 | 8 | 2 | 38.34 | 3.12e+006 | 53533.5 | HUMAN | 14 | Cornulin |
| 3 | 3.1 | 8 | 2 | 32.92 | 1.28e+006 | 35971.0 | HUMAN | 16 | Ly6/PLAUR domain-containing protein 3 |
| 4 | 4.1 | 6 | 2 | 26.08 | 6.31e+005 | 25335.2 | HUMAN | 19 | Suprabasin |
| 5 | 5.1 | 2 | 1 | 16.52 | 6.54e+005 | 78458.7 | HUMAN | 17 | Matrix metalloproteinase-9 |
| Totals: | | 32 | 9 | | | | | | |

**Figure 33 - TOP - Raw Q-TOF multiplex data indication of a vaginal fluid hit. BOTTOM – Database search results indicate high-confidence confirmatory identification: Human vaginal fluid. Targeted vaginal fluid biomarkers were detected. No biomarkers indicative of other body fluids were detected.**




| Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Distinct Summed MS/MS Search Score | Mean Peptide Spectral Intensity | Protein MW (Da) | Species | Database Accession # | Protein Name |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | 2 | 2 | 46.85 | 2.37e+005 | 3042.5 | HUMAN | 2 | Complement C3 |
| 2 | 2.1 | 28 | 2 | 42.21 | 3.32e+007 | 2678.1 | HUMAN | 1 | Hemoglobin subunit beta |
| 3 | 3.1 | 7 | 2 | 39.10 | 3.41e+005 | 4818.1 | HUMAN | 4 | Cornulin |
| 4 | 4.1 | 3 | 2 | 30.62 | 7.47e+005 | 2641.0 | HUMAN | 3 | Hemopexin |
| Totals: | | 40 | 8 | | | | | | |

**Figure 34 - TOP - Raw Q-TOF multiplex data indicating a possible menstrual blood hit. BOTTOM – Database search results indicate high-confidence confirmatory identification: Human blood biomarkers and a moderate-specificity vaginal fluid biomarker. No biomarkers indicative of other body fluids were detected. Taken together, these results may indicate the presence of menstrual blood or a mixture of peripheral blood and vaginal fluid.**




| Group (#) | Subgroup (#) | Spectra (#) | Distinct Peptides (#) | Distinct Summed MS/MS Search Score | Mean Peptide Spectral Intensity | Protein MW (Da) | Species | Database Accession # | Protein Name |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | 3 | 2 | 49.16 | 6.84e+005 | 3042.5 | HUMAN | 2 | Complement C3 |
| 2 | 2.1 | 8 | 2 | 40.87 | 1.91e+007 | 2678.1 | HUMAN | 1 | Hemoglobin subunit beta |
| 3 | 3.1 | 5 | 2 | 38.44 | 1.89e+006 | 2641.0 | HUMAN | 3 | Hemopexin |
| Totals: | | 16 | 6 | | | | | | |

**Figure 35 - TOP - Raw Q-TOF multiplex data indication of a blood hit. BOTTOM – Database search results indicate high-confidence confirmatory identification: Human blood. All targeted blood biomarkers were detected. No biomarkers indicative of other body fluids were detected.**

### 3.2.3 Single-Source Population Study

The applicability of the candidate biomarkers requires thorough verification for stain specificity across a larger population sample. To assess this, multiplexed Q-TOF analyses of single-source body fluid samples from a sample population of fifty human research participants were used. This made it possible to assess the frequency at which target biomarkers may be detected in both target and non-target body fluids. The results obtained are summarized in table 8 and are described for each fluid.

**Table 8 - Results of the targeted ion Q-TOF 6-body fluid multiplex assay. Targeted biomarkers are listed across the top of the figure and the body fluids tested are listed along the right side of the figure. Green cells indicate protein biomarkers that display high-specificity for their target body fluid and which were detected in all samples tested. Yellow cells indicate protein biomarkers that were detected in less than 100% of their target body fluid samples. Red cells indicate cases where a specific biomarker was not detected in any of the samples tested.**

| | Semen | Saliva | Urine - Male | Urine - Female | Vaginal Secretions | Menstrual Blood | Peripheral Blood |
|---|---|---|---|---|---|---|---|
| Semenogelin-1 | Semen | 100.0% | 0.0% | 24.0% | 0.0% | 0.0% | 0.0% |
| Semenogelin-2 | | 100.0% | 0.0% | 20.0% | 0.0% | 0.0% | 0.0% |
| Epididymal protein E1 | | 100.0% | 0.0% | 40.0% | 29.4% | 3.2% | 0.0% |
| PSA | | 100.0% | 0.0% | 60.0% | 0.0% | 0.0% | 0.0% |
| PAP | | 100.0% | 0.0% | 80.0% | 0.0% | 0.0% | 0.0% |
| Uromodulin | Urine | 0.0% | 0.0% | 100.0% | 0.0% | 0.0% | 0.0% |
| Osteopontin | | 0.0% | 0.0% | 100.0% | 0.0% | 0.0% | 0.0% |
| Submaxillary gland | Saliva | 0.0% | 100.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Cystatin_D | | 0.0% | 76.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Cystatin_SA | | 0.0% | 94.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Statherin | | 0.0% | 90.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Histatin 1 | | 0.0% | 30.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| mucin 5b | | 0.0% | 96.0% | 4.0% | 38.0% | 20.0% | 0.0% |
| Cornulin | Vaginal and Menstrual | 0.0% | 0.0% | 4.0% | 100.0% | 20.0% | 0.0% |
| Neutrophil gelatinase-associated lipocalin | | 0.0% | 0.0% | 0.0% | 100.0% | 6.0% | 0.0% |
| Ly6/PLAUR domain-containing protein 3 | | 0.0% | 0.0% | 4.0% | 100.0% | 4.0% | 0.0% |
| IgGFc-binding protein | | 0.0% | 0.0% | 0.0% | 68.0% | 0.0% | 0.0% |
| Matrix mmp-9 | | 0.0% | 0.0% | 0.0% | 20.0% | 0.0% | 0.0% |
| Suprabasin | | 0.0% | 0.0% | 0.0% | 22.0% | 0.0% | 0.0% |
| Hemoglobin subunit beta | Peripheral Blood | 0.0% | 4.0% | 0.0% | 12.0% | 6.0% | 100.0% |
| Complement C3 | | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 100.0% |
| Hemopexin | | 0.0% | 0.0% | 0.0% | 6.0% | 4.0% | 96.0% |

### 3.2.3.1 Seminal Fluid

Seminal fluid biomarkers semenogelin I/II, epididymal secretory protein E1, prostatic acid phosphatase and prostate specific antigen were consistently detected in all semen samples. These markers were undetectable in non-target body fluids with a few exceptions. Semenogelin I/II, epididymal secretory protein E1, prostatic acid phosphatase and prostate specific antigen were observed in 20-80% of male urine samples and epididymal secretory protein E1 was observed in 4% of female urine samples.

Seminal fluid donors were instructed to refrain from ejaculation for 24 hours prior to urine collection, although compliance was based only on self-reporting. It is possible that some seminal fluid proteins may persist in the male urinary/reproductive tract. This may be due to leakage from the reproductive system or residual proteins from a prior or nocturnal ejaculation. The seminal fluid proteins employed as biomarkers in the current study have also been observed in other urinary proteomics studies involving male subjects[208]. Even though these proteins have been detected in non-target body fluids, Sg I/II, SAP and PSA/p30 were only identified in male-derived samples. Interestingly, no urine-specific markers were detected in any seminal fluid sample. This leads to the conclusion that at a minimum, these seminal fluid biomarkers may have utility in tracing source of the stain to the male genitourinary tract.

The candidate seminal fluid biomarker Epididymal secretory protein E1, however, was detected in female urine at nearly the same rate as in male urine samples. It was not possible based on the assays performed to determine if this represented an endogenous component of vaginal fluid or urine, or if the detected protein represented residual protein from a prior sexual encounter though this seems less likely. As with male participants,

females were instructed to refrain from sexual intercourse for 5 days prior to the collection of urine, vaginal fluids and menstrual blood. As a result of these findings, epididymal secretory protein E1 cannot be considered specific to seminal fluid or even males. Accordingly, it was eliminated as a marker of seminal fluid.

#### 3.2.3.2 Saliva

Saliva candidate biomarkers consisted of cystatin SA, cystatin D, submaxillary gland androgen-regulated protein, histatin-1, and statherin. Of these markers, several showed a high degree of target fluid specificity and consistency of detection across the expanded donor population. Submaxillary gland androgen-regulated protein was detected in 100% of saliva samples; cystatin SA and statherin were identified in ≥90% of the samples and; Cystatin D was identified in 76% of samples and processed on the Q-TOF. Improvements in assay sensitivity (*e.g.* use of a QQQ platform) may allow for more consistent detection of cystatin SA, cystatin D, as well as statherin. In contrast to these results, histatin 1 was only detectable in 30% of samples assayed. Based on the low identification rate achieved for histatin 1 in the samples tested, the decision was made to remove it from further consideration as a reliable biomarker of saliva.

#### 3.2.3.3 Urine

Urine candidate biomarkers uromodulin and osteopontin were detected in all male and female urine samples. These two proteins were not detected in any samples of non-target body fluids. Uromodulin and osteopontin appeared to be well-suited, reliable and high-specificity biomarkers for urine.

### 3.2.3.4 Vaginal fluid

Vaginal fluid biomarker candidates included cornulin, IgGFc-binding protein, neutrophil gelatinase-associated lipocalin, Ly6/PLAUR containing protein 3, suprabasin, and matrix metalloproteinase-9. Of these, cornulin, neutrophil gelatinase-associated lipocalin and Ly6/PLAUR containing protein 3 were detected in 100% of samples assayed. IgGFc-binding protein was detected in 68% of samples, matrix metallo-proteinase-9 in 20% of samples, and suprabasin was detected in 22% of samples analyzed on the Q-TOF. Other than the anticipated non-target identification of cornulin, neutrophil gelatinase-associated lipocalin and Ly6/PLAUR containing protein 3 in samples of menstrual blood.

Of these markers, mucin5B was detected consistently in other non-targets body fluids analyzed (*i.e.*, saliva and menstrual blood). It was initially hypothesized that mucin 5B would be specific to vaginal fluid. However, it was also detected in 20% of menstrual blood samples as well as 4% of male/female urine samples. From these results it is unlikely that mucin 5B will offer any discriminatory power with regard to body fluid identification.

### 3.2.3.5 Menstrual Blood

As discussed in the method development section, there were no reliable markers which could definitively differentiate menstrual fluid/blood from vaginal fluid. It was hypothesized that vaginal fluid markers may allow the differentiation of menstrual blood from other fluids – with differentiation from peripheral blood being particularly important. However, cornulin, neutrophil gelatinase-associated lipocalin and

Ly6/PLAUR containing protein 3, were only found in 30%, 6%, and 4% of menstrual blood samples assayed, respectively. This may be due to the complexity of the blood proteome and sensitivity limits of the Q-TOF method. Accordingly, these markers should be reevaluated for their potential utility from the perspective that a more sensitive assay may allow reliable detection of these proteins. That being said, it must be emphasized that without a 100% specific menstrual blood protein biomarker, it is impossible to differentiate between menstrual blood *vs.* a mixture of peripheral blood and vaginal fluid.

#### 3.2.3.6 Peripheral Blood

Biomarker candidates for peripheral blood included hemoglobin subunit beta, complement C3, and hemopexin. Hemoglobin and complement C3 were detectable in all peripheral blood samples while hemopexin was identified in 96% of analyzed samples.

These markers, however, were also detectable in several non-target body fluids including menstrual fluid/blood, vaginal fluid, female urine and saliva. It was expected that these candidates would be detected in menstrual fluid as peripheral blood is a major component of this fluid. Hemoglobin subunit beta, complement C3, and hemopexin were detected in 100%, 54% and 76% of menstrual blood samples. The lower rate of detection of some of these proteins in menstrual fluid is thought to be due to matrix suppression by proteins derived from cellular/endometrial tissues. Blood detection in vaginal fluid, female urine, and saliva may be due to trace amounts of these proteins in these fluids due to small abrasions from normal activities such as the use of dental floss, or feminine hygiene products. Additionally, hemoglobin, complement C3, and hemopexin have been

documented in some salivary, vaginal fluid, and urinary proteomics projects aside from the findings presented here[146, 209].

### 3.2.4 Casework-Type Samples

Laboratory samples from the completed population study were used to identify those biomarkers which demonstrated the highest degree of consistency and target fluid specificity. While these laboratory samples facilitated the selection of the most promising biomarkers for further validation, they did not assess the use of a mass spectrometry platform for the processing of samples under conditions typically encountered in a forensic lab. In order to address this gap, a series of studies were performed to evaluate the reliability of biomarker detection in a forensic context. More specifically, samples were assayed to determine whether or not the selected biomarkers could be recovered from a variety of substrates such as cotton, leather or denim and items typically associated with sexual assaults and then accurately identified. Samples were also prepared to determine the extent to which body fluids exposed to a variety of environmental and chemical contaminants could be accurately identified. Finally, mixtures containing two to three different body fluids were prepared and assayed to test the overall reliability of the multiplex assay to reliably detect multiple biomarkers in a single-pass assay. In all, 37 casework-type samples were analyzed in order to gain an initial assessment of the applicability of the method for use by forensic practitioners (**Tables 9 and 10**). These samples were divided into five specific categories including single source body fluids extracted from swabs, substrate extractions, volume-based mixtures, and contaminant extractions.

### 3.2.4.1 Single Source Swabs

For each body fluid of interest, a 50 µL aliquot of untreated sample was applied to a cotton swab which was then allowed to dry in accordance with standard practices employed by caseworking laboratories. Prior to testing, the dried sample was re-suspended in phosphate buffered saline. Summarized results can be seen in **Table 9**. Saliva, semen, vaginal fluid, peripheral blood and urine spotted onto cotton swabs were accurately identified by one or more body fluid specific proteins. Unfortunately, no vaginal-specific proteins were identified in the menstrual blood swab assayed. Inclusion of vaginal proteins in the menstrual blood sample could support an interpretation that the source of the sample had been in contact with the vaginal canal, but without an identification of any of these proteins it was impossible to differentiate menstrual blood from peripheral blood. Finally, aside from the anticipated epididymal secretory protein E1 which was detected in urine, no unexpected proteins were identified in any of these body fluids.

### 3.2.4.2 Single Source Substrate Studies

For each body fluid of interest, 50 µL of sample was added to a variety of commonly encountered substrates which were then allowed to dry in accordance with standard practices employed by caseworking laboratories. Prior to testing, the dried sample was re-suspended in phosphate buffered saline. Summarized results can be seen in **Table 9.** As was seen with single-source samples spotted on cotton swabs, saliva, semen, vaginal fluid, peripheral blood and urine were accurately identified although menstrual blood could not be differentiated from peripheral blood.

As expected, epididymal secretory protein E1 was detected in the urine sample that was tested. In addition to epididymal secretory protein E1, Ly6/PLAUR containing protein 3 was also detectable in all three saliva samples (i.e., cigarette butt, mouth of bottle, and swab of saliva stained finger). Based on prior work this protein had been considered a vaginal fluid-specific biomarker. Upon further research, it was found that Ly6/PLAUR containing protein 3 has also been identified in the squamous esophageal epithelium – though specifically in relation to esophageal cancer[210]. The detection of this protein in a non-target body fluid at this stage of the research is likely due to the modified collection protocols for the casework samples. In the population study, saliva was collected using an absorbent cotton sponge that was placed next to the salivary ducts. It is thought that this process produces a pure and cell-free saliva solution. To more realistically simulate what would be seen in an actual forensic context, saliva samples were deposited directly from the mouth into a 50 mL collection tube before application onto the cotton swab and other substrates. This modified method deposited a greater amount of cellular material and protein from the oral cavity which may account for the detectible presence of Ly6/PLAUR containing protein 3. This unexpected result highlights the importance of conducting studies on casework-type samples.

#### 3.2.4.3 Volume Mixtures

Six, two-component mixtures of human body fluids were run on the multiplex assay to evaluate potential matrix interference between the proteins of different body fluids. Summarized results can be seen in **Table 10.** In all but one case (i.e., a mixture of equivalent volumes of saliva and peripheral blood), at least one biomarker for each body

fluid present in the mixture was detected. This loss of detection is likely due to the quantitative difference in protein content between saliva and peripheral blood when mixed in equal volumes. The total protein content of saliva is approximately 1.3 mg/mL while the protein concentration in peripheral blood is approximately 120 mg/mL. Applying 25 µl of saliva to the swab results in ~32 µg of total saliva protein on the swab compared to ~3,000 µg of total blood protein onto the swab. Based on these values it is likely that there is not enough saliva protein to reliably detect it against the background of more abundant blood proteins.

#### 3.2.4.4 Contaminant Studies

A series of single-source body fluid samples were also assayed for the influence of chemical additives and/or inhibitors on biomarker detection. These chemical additives included spermicidal lubricant commonly found on condoms and sexual assault swabs; BlueStar® which is commonly employed in screens for blood, soil and chewing tobacco juice which contain potent inhibitors of some enzymes; coffee which is encountered in some beverage adulteration cases and; chewing gum. Summarized results can be seen in **Table 10.** Of the inhibitors assayed, only chewing tobacco juice appeared to prevent the identification of saliva. This was not entirely unexpected given that tobacco juice is also known to act as a potent inhibitor of DNA typing chemistries due to its action on *Taq* polymerase. Aside from saliva which has a low overall protein content, no interference of biomarker detection was seen with any of the other chemical additives that were tested.

**Table 9 - Results of the Q-TOF multiplex assays using casework-type samples. Targeted biomarkers are listed across the top of the figure and the description of the type of body fluid and substrate are provided along the right side of the figure. Green cells indicate protein biomarkers that were detected. Red cells indicate protein biomarkers that were not detected. (VF: vaginal fluid; MB: menstrual blood; PB: peripheral blood)**

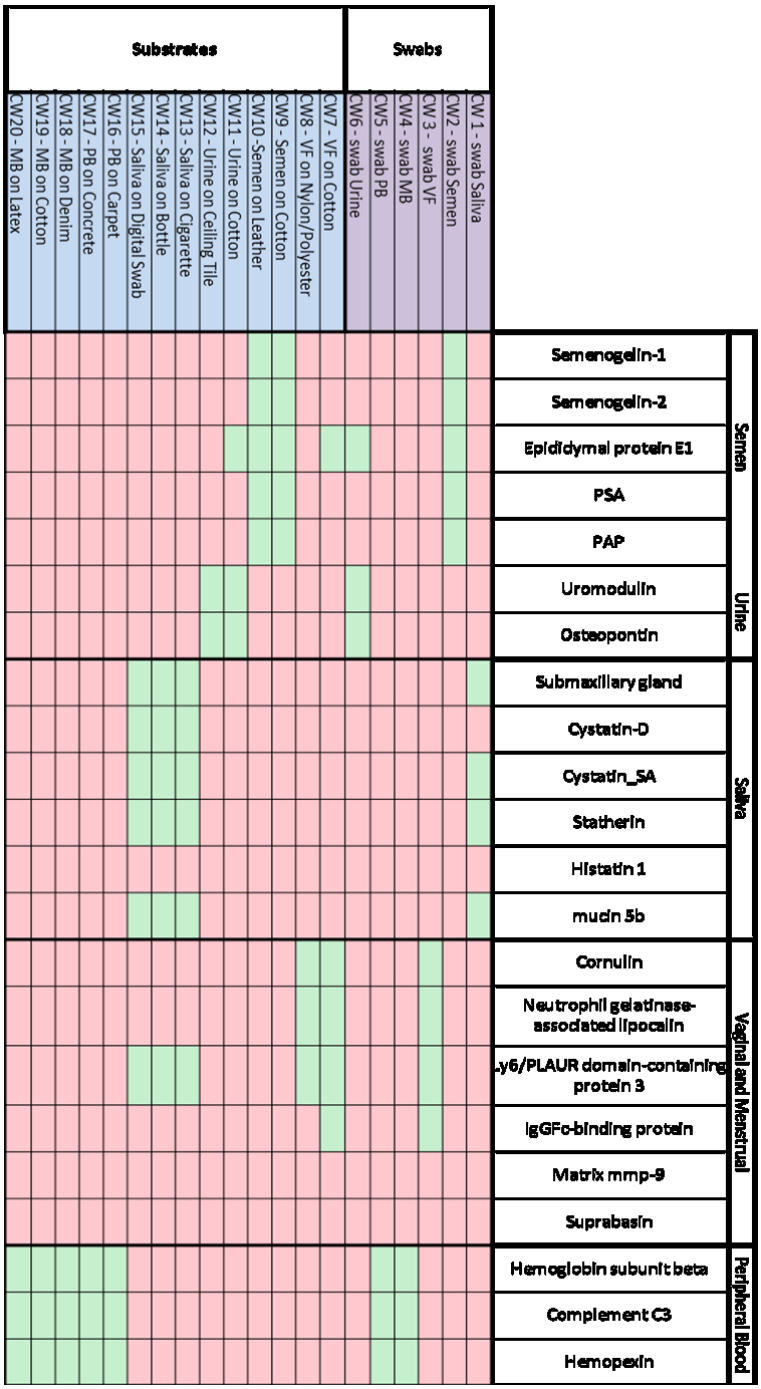
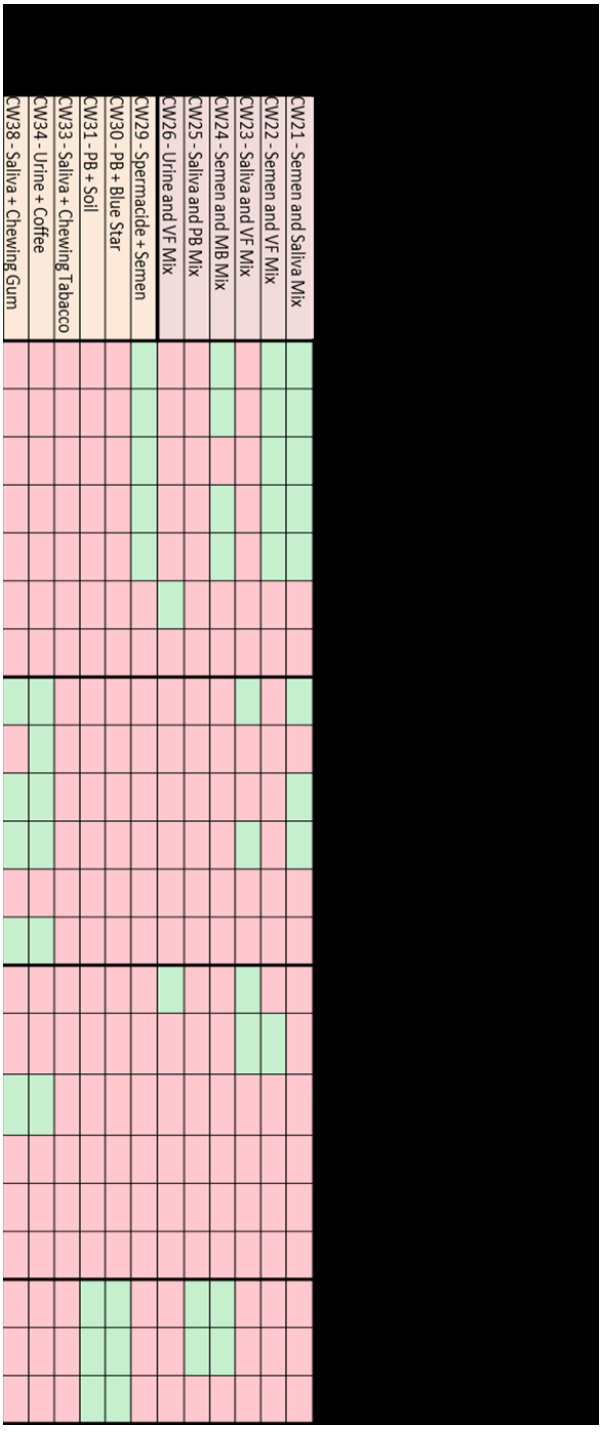


**Table 10 - Results of the Q-TOF multiplex assays using casework-type mixed-source samples and samples containing mixtures. Green cells indicate protein biomarkers that were detected. Red cells indicate protein biomarkers that were not detected. (VF: vaginal fluid; MB: menstrual blood; PB: peripheral blood)**



## 3.3 Conclusion

The second phase in the proteomics pipeline was to assess the biological applicability of candidate biomarkers to their intended use. In the case of the current research, the two major objectives were to assess whether the candidates were specific to their target body fluid as well as to determine whether the biomarkers could be reliably and consistently detected in a larger population of donors. In addition, because the overall goal of this project is to replace traditional/non-specific forensic serological tests with a confirmatory multiplex assay, it was also necessary to assess the viability of a mass spec-based assay for processing forensic-type samples.

To assess the specificity and consistency of the initial biomarker candidates, twenty one of the most promising protein markers identified during the discovery phase were selected for inclusion in a multiplex validation assay. These markers consisted of six seminal fluid candidates, two urine candidates, six saliva candidate biomarkers, seven vaginal/menstrual fluid candidates as well as three peripheral blood candidate biomarkers. Using these proteins, a 44-minute targeted ion assay was designed and tested with single-source laboratory samples collected from fifty human donors. Results from this study identified several biomarkers which were both target-fluid specific and which were consistently detected in all donor samples. In these studies, several candidate markers which were non-specific for the target body fluid or were inconsistently detected in an expanded population of human donors were also identified and subsequently eliminated from further consideration as forensic biomarkers. In addition to the population study, a series of casework-type samples were also processed in order to evaluate the performance of the prototype targeted-ion Q-TOF method for use in a

caseworking laboratory. Of these casework-type samples, detection of the target biomarkers made it possible to correctly identify the body fluid components in all but a select few challenging samples.

Overall, this series of experiments was able to successfully assess the specificity, consistency, and applicability of a mass spectrometry-based forensic assay for body fluid identification. However, even with these successes, a number of serious limitations to the instrument's performance were recognized that may act as barriers to adoption by practitioners in forensic testing laboratories. Most significantly, the instrument is not optimally engineered for multiplex assays[211] and at a top speed of 44-minutes/sample and one sample/run it is not well suited to the higher throughput needs of most forensic laboratories. With this limitation of speed in mind, a different mass spectrometry-based platform may be a superior choice to meet the throughput requirements of a typical caseworking lab. Triple quadrupole mass spectrometers performing multiple reaction monitoring (QQQ-MRM) have become the "workhorse" instrument for high-throughput clinical and toxicological screening. This is due to advantages in sensitivity, selectivity, easier operation and lower cost of these instruments compared to Q-TOFs. In addition, QQQ instruments offer faster analytical gradients which can match the needs of a caseworking/production lab. Therefore, the next step in the dissertation research was to develop a QQQ multiplex assay incorporating the optimal biomarkers identified in the initial biomarker verification phase of the current research.

# Chapter 4:

# Prototype Assay Validation

\

# 4 Introduction

The current dissertation research has followed the traditional proteomics research and development pipeline beginning with the discovery of candidate biomarkers for body fluid identification through verification of the candidate biomarker specificity and reliability. The biomarker discovery phase resulted in the generation and characterization of a panel of over thirty candidate biomarkers for the identification of biological stains. During the biomarker verification phase, a novel Q-TOF assay was designed to verify the specificity of the most promising biomarker candidates across a larger population of human donors. The Q-TOF verification phase was able to successfully confirm the specificity of many of protein biomarkers while also flagging several proteins which displayed a detectible presence in non-target biological fluids; thus eliminating them as high-specificity biomarkers. Additionally, the viability of a mass spectrometry-based serological assay for adoption was successfully assessed using casework-type samples. The data from the first two phases of this research and development effort produced a panel of body fluid-specific biomarkers which can be reliably detected using mass spectrometry. The next step in the research and development process, with an emphasis on "real-world" application, was a full developmental validation to assess the reliability, reproducibility, accuracy, sensitivity and limitations of a mass spectrometry-based assay for human stain identification. The overarching goal of this final stage of development being the implementation of an MS-based serological assay for forensic laboratory use.

While the Q-TOF assay was successfully able to confidently identify a biological stain, the instrument is not well suited for multiplex assays or the throughput required by a caseworking laboratory. Due to these critical limitations, a developmental validation

study using this instrument would be inadvisable for a laboratory implementation perspective. However, in recent years triple quadrupole mass spectrometers performing multiple reaction monitoring (QQQ-MRM) have been considered to be the new "workhorse" instrument for proteomics validation assays, clinical laboratory testing, and confirmatory testing of illicit substances in toxicology laboratories. This is due to their greater sensitivity, greater selectivity, easier operation, and lower cost of operation and ownership compared to the Q-TOF.

The use of the QQQ-MRM platform provides analysts with much greater confidence in the accuracy of the results obtained for a given stain. This is because each individual body fluid is identified based on the presence of multiple biomarker proteins (*e.g.* statherin and cystatin-SA for saliva). The presence of each biomarker protein itself is based on the detection of multiple diagnostic ions (*i.e.* precursor ions) that are derived from the original biomarker protein by trypsin digestion. The presence/identity of each precursor ion is then independently reconfirmed by detection of its breakdown products (*i.e.* product ions). The two-fold filtration mechanism in these interments leads to a much more sensitive assay when targeting specific analytes. Additionally, this internal confirmation and reconfirmation of targets stands in sharp contrast to existing commercial forensic assays where identification is typically based on the indirect detection of a single binding event between an antibody and its presumed target protein.

The following chapter will focus on the development and testing of a QQQ-MRM prototype assay for the identification of seminal fluid, saliva, and vaginal fluid (*i.e.*, a "sexual assault" targeted assay) in laboratory- and casework-type samples. Additionally, a workflow was developed in order to simultaneously process evidentiary material for

STR-based DNA analysis (for stain individualization) coupled with an MS-based serological identification.

## 4.1 Materials and Methods

All research conducted under this project was reviewed, approved by the University of Denver Institutional Review Board (IRB) for research involving human subjects and conducted in full compliance with U.S. Federal Policy for the Protection of Human Subjects. In total, 100 subjects (50 males and 50 females) were recruited from the undergraduate population at the University of Denver. Female donors were compensated $80 and males $20 for their participation in the research study. All research volunteers agreed and signed a letter of informed consent acknowledging they have received, read, and understood the protocols involved in the collection of samples. Additionally, all donated samples were given a random alphanumeric label to ensure confidentiality. Finally, males were instructed to refrain from ejaculation for 24 hours prior to sample collection and all donors were instructed to refrain from sexual intercourse for 5 days prior to the collection of urine, and vaginal fluids and menstrual blood for females.

### 4.1.1 Sample Collection and Preparation

An overview of the workflow performed from sample collection through biomarker identification is provided in **Figure 36**. Saliva, seminal fluid, and vaginal fluid, were self collected by study participants. After collection, samples were centrifuged and filtered to remove cellular material and quantified for total protein. Samples were divided into 10-15 equal volume aliquots and stored at -80°C before analysis.

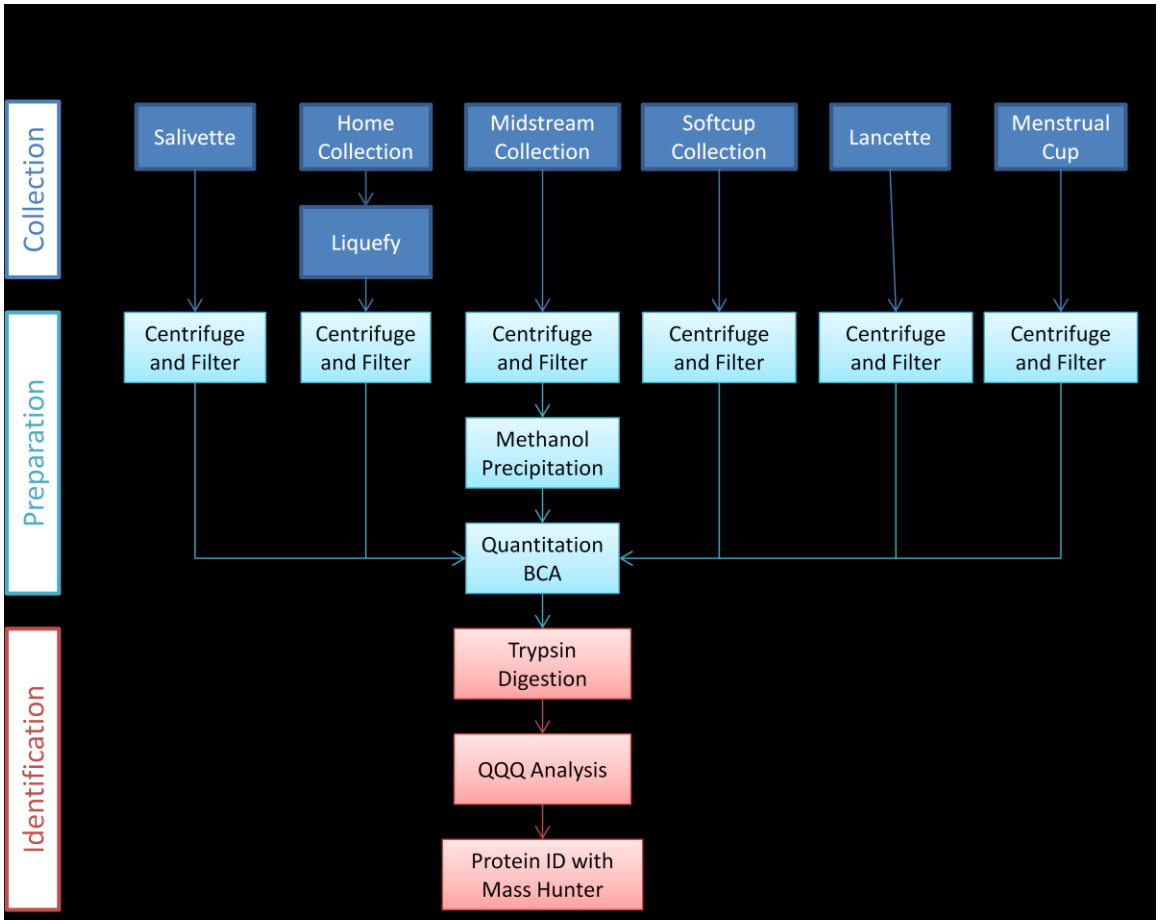



**Figure 36 - Overview of QQQ Prototype phase from sample collection to protein identification. Briefly, samples were collected and cellular debris removed via centrifugation and filtration. Samples were quantified for total protein and analyzed by mass spectrometry.**

### 4.1.2 Body Fluid Identification

#### 4.1.2.1 Trypsin Digestion and Peptide Cleanup

Trypsin digests and C-18 cleanup was performed as described in sections 2.1.3.2 and 2.1.3.3.

#### 4.1.2.2 QQQ MS/MS Acquisition

Mass spectrometry was performed on an Agilent Technologies 6410 triple quadrupole mass spectrometer. The method employed a 150 x 1.0 mm 3.5um 300SB-C18 analytical column with 12.5 x 2.1mm 5µM SB-C8 enrichment column. Solvent A

contained $H_20$ with 0.1% formic acid and Solvent B contained a 90:10 mixture of acetonitrile:$H_20$ with 0.1% formic acid. Samples were initially loaded on an enrichment column at a flow rate of 0.2 mL/min with flow through going to waste. At 1 minute, the column was switched to analysis mode with Buffers A and B passing through both the enrichment and analytical columns with a gradient of 5-35% Buffer B over 15 minutes. This was followed by 80% Buffer B from 15.01 minutes to 16 minutes and 100% Buffer B from 16.01 minutes to 17.01 minutes. A 5-minute post-gradient run was performed to re-equilibrate the column.

#### 4.1.2.3 QQQ Data Analysis

Data analysis was performed using Agilent MassHunter QQQ Quantitative software as well as the Skyline targeted proteomics environment.

### 4.1.3 Preparation and Analysis of Casework-Type Samples

All casework-type samples were prepared as described in section 3.1.3 and analyzed using the methods described in section 4.1.2.

### 4.1.4 Combined DNA Typing and Serological Analysis

Mixtures and swabs were prepared as described previously with the following minor modification. After the cellular centrifugation step, the supernatant was removed and processed for protein analysis. The cellular pellet (*i.e.*, DNA-rich material) was retained for subsequent isolation of genomic DNA suitable for STR-based typing.

#### 4.1.4.1 DNA extraction on BioRobot EZ1

Genomic DNA was extracted by paramagnetic isolation using the using the manufacturer's recommended protocol for the EZ1 DNA Tissue kit (Qiagen, Valencia, CA). Briefly, 190 µL of proprietary Buffer G2 and 10 µL of Protinase K (activity = 600 mAU/ml) were added to the pelleted material and incubated for 15 minutes at 56°C with constant shaking in and in an Eppendorf® Thermomixer to lyse the cellular material. Insoluble material, was then pelleted by centrifugation at 300 x g for 1 minute at room temperature. The DNA-containing supernatant was then transferred to a new 2 ml sample tube and processed for DNA extraction and purification on the BioRobot EZ1 using the EZ1 DNA Tissue Card.

#### 4.1.4.2 DNA Quantitation with Quantifiler™ Duo

Quantifiler™ Duo (Life Technologies/ Applied Biosystems, Foster City, CA) primer mix (4.2 µL/sample) and PCR reaction mix (5 µL/sample) were combined to form a DNA quantification master mix. In addition, eight control DNA standards were prepared at a concentration of 50, 16.7, 5.56, 1.85, 0.62, 0.21, 0.068 and 0.023 ng/µL. For each quantification reaction, 0.8 µL of DNA sample or DNA control standard was mixed with 9.2 µL of master mix in the wells of a standard optical 96-well plate. Wells were sealed with optical film and analyzed on a Life Technologies/ Applied Biosystems Prism 7900HT Sequence Detection System. DNA quantities were determined using the integrated SDS v2.1.1 software.

#### 4.1.4.3 DNA Profiling with Identifiler™ Plus

Identifiler™ Plus (Life Technologies/ Applied Biosystems, Foster City, CA) master mix (4 µL/sample) and Identifiler™ Plus primer mix (2 µL/sample) were

combined to form an master reaction mix for human DNA profiling by Short Tandem Repeats (STR) loci genotyping. For each amplification reaction, 6 µL of master mix was added to a 0.2 mL PCR tube followed by 4 µL of DNA extract at a concentration of 0.1 ng/µL. The thermal cycling conditions are shown in **Table 11**. All amplification reactions were accompanied by negative and reagent blank controls. Amplified products were electrokinetically injected for 5 sec and fractionated on an Prism™ 310 Genetic Analyzer using the GS STR POP4 (1 mL) G5 filter set module. The electrophoresis run time of 24 min allowed detection the 500 base pair size peak of the GS500 internal lane sizing standard. Data Collection Software v3.1 and Genemapper ID v3.2 (Applied Biosystems, Foster City, CA) were used for data collection and analysis.

**Table 11 – Thermocycler conditions for STR amplification with Identifiler™ Plus.**

| Initial Incubation Step | Denature | Anneal | Extend | Final Extension | Final Step |
|---|---|---|---|---|---|
| HOLD | CYCLE (standard: 28 cycles, LCN: 34 cycles) | | | HOLD | HOLD |
| 95°C<br>11 min | 94 °C<br>1 min | 59°C<br>1 min | 72°C<br>1 min | 60°C<br>60 min | 4 – 25°C<br>forever |

## 4.2 Results and Discussion

### 4.2.1 Method Development

The first objective of this phase of the current research was to identify the optimal diagnostic ions to develop an "MRM sexual assault evidence assay" that targeted the identification of human saliva, seminal fluid, and vaginal secretions. To this end, a QQQ

multiplex assay for the detection of these optimal ions was then developed. While this approach yields very dependable single pass results, the method design and development process involves a substantial amount of effort to identify biologically relevant protein targets, optimal tryptic peptides as well as the specific product ions which the QQQ-MRM assay can measure. The first two requirements were completed as part of the biomarker discovery and verification stages of the current project

From the Q-TOF verification study, several biologically relevant proteins in saliva, seminal fluid, and vaginal fluid were identified as consistently detected in an expanded sample of human donors with little to no detection in non-target body fluids. Using these data (**Table 8**) three markers were selected for the MRM assay for each body fluid including: cystatin SA, statherin, and submaxillary gland androgen-regulated protein 3B as markers for saliva; SAP and Sg-I/II for seminal fluid; and cornulin, Ly6/PLAUR domain-containing protein 3, and neutrophil gelatinase-associated lipocalin as markers for vaginal fluid. Since these markers had been examined previously, the optimal peptides for each protein were already known (**Tables 7A-7E**) and were readily ported for the development of a QQQ assay. This is because the peptides that had been selected for the Q-TOF assay were consistently detected in all donors in our sample population; displayed strong signal intensity; had high database match statistics; were unique to their parent protein; were devoid of missed cleavage sites; and did not include peptides with potential amino acid modifications including those with methionine, cystine, or N-Terminus glutamic acid residues.

The final aspect of assay design involved selection of product ions from the peptide fragmentation. Peptide exact mass as well as a unique fragmentation pattern was what the Q-TOF used to confirm the detection of a specific marker. The MRM assay used the same fragmentation mechanism but only detected specific product ions (*i.e.*, precursor-product ion pair or transition) rather than the entire product ion spectra. As a result, selection of the most abundant ions is needed to ensure that the greatest level of detection and sensitivity is achieved. **Figure 37** presents the sequence-specific fragmentation pattern generated on the Q-TOF for the peptide FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF which is diagnostic for the salivary protein marker statherin. The m/z for this peptide is 1215.2332 with a charge of +3. For product ion selection, high-abundance ions which are larger than the precursor m/z are selected. From the fragmentation in **Figure 37** the product ions 1462.6 and 1229.5 m/z were targeted for this Statherin peptide. **Figure 38** shows the QQQ raw dataset in which a much cleaner signal was obtained as compared to that obtained with the Q-TOF. Using this same approach, the optimal precursor-product ion pairs were determined for each protein biomarker. These are presented in **Table 12**.

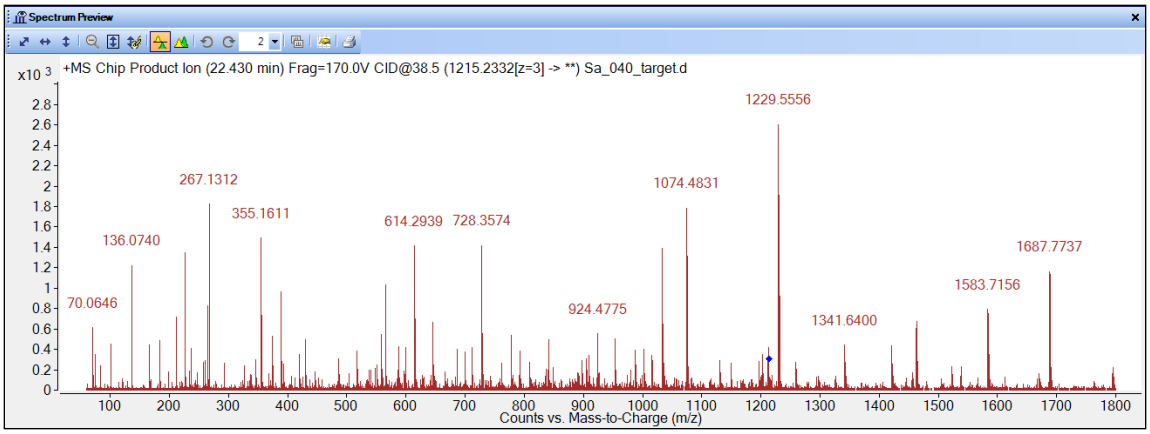



**Figure 37 - In a Q-TOF assay, inclusion list precursor ions are only isolated in the first quadrupole (Q1). After the isolated precursor ions are fragmented in the collision cell all product ions reach the detector. These fragments generate a complex fragmentation spectrum which can be analyzed to confirm the presence of a particular protein biomarker in a database.**

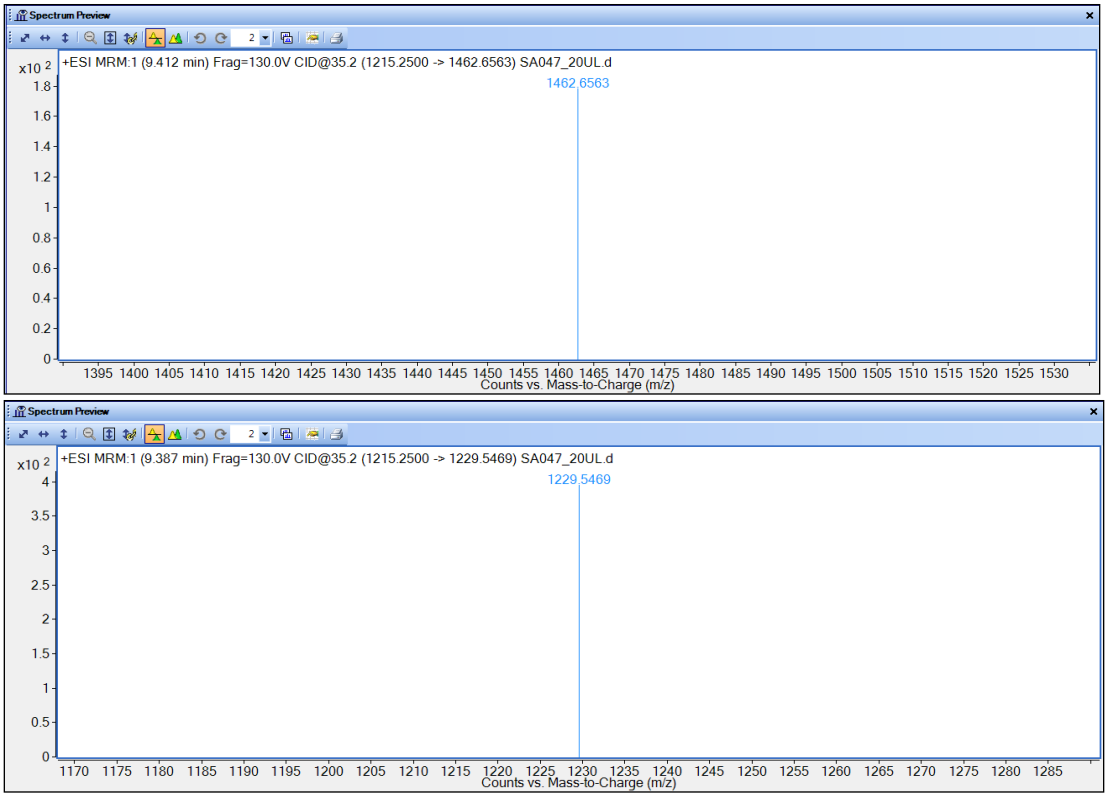



**Figure 38 - When using a QQQ in MRM mode, target-ion list precursor ions are isolated in the first quadrupole (Q1); fragmented and then sent to the second quadrupole (Q2) which acts as a second filter for the isolation of specific product ions. This two-fold filtration process greatly improves the signal-to-noise ratio for target ion detection and increases the sensitivity of the assay by 10-100 fold. Shown are two product ions (1462.6 and 1229.5 m/z) from salivary biomarker statherin peptide FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF.**

**Table 12 - QQQ-MRM Transition List Targeting Seminal Fluid, Saliva and Vaginal Fluid.**

| Compound Group | Compound Name | ISTD? | Precursor Ion | Product Ion | Dwell | Fragmentor | Collision Energy | Ion Name |
|---|---|---|---|---|---|---|---|---|
| sp\|Q9UBG3\|CRNN_HUMAN | TLSESAEGACGSQESGSLHSGASQELGEGQR | FALSE | 1036.125319 | 1131.53889 | 20 | 130 | 28.6 | y11 |
| sp\|Q9UBG3\|CRNN_HUMAN | TLSESAEGACGSQESGSLHSGASQELGEGQR | FALSE | 1036.125319 | 1259.54659 | 20 | 130 | 28.6 | y25 |
| sp\|Q9UBG3\|CRNN_HUMAN | TLSESAEGACGSQESGSLHSGASQELGEGQR | FALSE | 1036.125319 | 1195.02529 | 20 | 130 | 28.6 | y24 |
| sp\|Q9UBG3\|CRNN_HUMAN | ISPQIQLSGQTEQTQK | FALSE | 893.470697 | 1247.62262 | 20 | 130 | 30 | y11 |
| sp\|Q9UBG3\|CRNN_HUMAN | ISPQIQLSGQTEQTQK | FALSE | 893.470697 | 1006.47998 | 20 | 130 | 30 | y9 |
| sp\|Q9UBG3\|CRNN_HUMAN | ISPQIQLSGQTEQTQK | FALSE | 893.470697 | 919.44795 | 20 | 130 | 30 | y8 |
| sp\|P80188\|NGAL_HUMAN | TFVPGCQPGEFTLGNIK | FALSE | 622.313583 | 1075.57824 | 20 | 130 | 13.2 | y10 |
| sp\|P80188\|NGAL_HUMAN | TFVPGCQPGEFTLGNIK | FALSE | 622.313583 | 978.525472 | 20 | 130 | 13.2 | y9 |
| sp\|P80188\|NGAL_HUMAN | TFVPGCQPGEFTLGNIK | FALSE | 622.313583 | 759.374483 | 20 | 130 | 13.2 | y14 |
| sp\|P80188\|NGAL_HUMAN | SYPGLTSYLVR | FALSE | 628.337695 | 1005.57276 | 20 | 130 | 16.5 | y9 |
| sp\|P80188\|NGAL_HUMAN | SYPGLTSYLVR | FALSE | 628.337695 | 908.519993 | 20 | 130 | 16.5 | y8 |
| sp\|P80188\|NGAL_HUMAN | SYPGLTSYLVR | FALSE | 628.337695 | 738.414465 | 20 | 130 | 16.5 | y6 |
| sp\|O95274\|LYPD3_HUMAN | GCVQDEFCTR | FALSE | 636.260804 | 1054.46222 | 20 | 130 | 16.9 | y8 |
| sp\|O95274\|LYPD3_HUMAN | GCVQDEFCTR | FALSE | 636.260804 | 955.393806 | 20 | 130 | 16.9 | y7 |
| sp\|O95274\|LYPD3_HUMAN | GCVQDEFCTR | FALSE | 636.260804 | 827.335229 | 20 | 130 | 16.9 | y6 |
| sp\|O95274\|LYPD3_HUMAN | DGVTGPGFTLSGSCCQGSR | FALSE | 971.922731 | 1212.50958 | 20 | 130 | 34 | y11 |
| sp\|O95274\|LYPD3_HUMAN | DGVTGPGFTLSGSCCQGSR | FALSE | 971.922731 | 1111.4619 | 20 | 130 | 34 | y10 |
| sp\|O95274\|LYPD3_HUMAN | DGVTGPGFTLSGSCCQGSR | FALSE | 971.922731 | 998.377839 | 20 | 130 | 34 | y9 |
| sp\|P04279\|SEMG1_HUMAN | KQGGSQSSYVLQTEELVANK | FALSE | 722.706085 | 1031.53677 | 20 | 130 | 17 | y9 |
| sp\|P04279\|SEMG1_HUMAN | KQGGSQSSYVLQTEELVANK | FALSE | 722.706085 | 903.478188 | 20 | 130 | 17 | y8 |
| sp\|P04279\|SEMG1_HUMAN | KQGGSQSSYVLQTEELVANK | FALSE | 722.706085 | 1022.49015 | 20 | 130 | 17 | b10 |
| sp\|P04279\|SEMG1_HUMAN | DIFSTQDELLVYNK | FALSE | 842.92506 | 1309.66342 | 20 | 130 | 27.4 | y11 |
| sp\|P04279\|SEMG1_HUMAN | DIFSTQDELLVYNK | FALSE | 842.92506 | 1121.58372 | 20 | 130 | 27.4 | y9 |
| sp\|P04279\|SEMG1_HUMAN | DIFSTQDELLVYNK | FALSE | 842.92506 | 993.525138 | 20 | 130 | 27.4 | y8 |
| sp\|Q02383\|SEMG2_HUMAN | DIFTTQDELLVYNK | FALSE | 849.932885 | 1222.63139 | 20 | 130 | 27.8 | y10 |
| sp\|Q02383\|SEMG2_HUMAN | DIFTTQDELLVYNK | FALSE | 849.932885 | 1121.58372 | 20 | 130 | 27.8 | y9 |
| sp\|Q02383\|SEMG2_HUMAN | DIFTTQDELLVYNK | FALSE | 849.932885 | 993.525138 | 20 | 130 | 27.8 | y8 |
| sp\|Q02383\|SEMG2_HUMAN | DVSQSSISFQIEK | FALSE | 734.36992 | 1038.5466 | 20 | 130 | 21.9 | y9 |
| sp\|Q02383\|SEMG2_HUMAN | DVSQSSISFQIEK | FALSE | 734.36992 | 951.514573 | 20 | 130 | 21.9 | y8 |
| sp\|Q02383\|SEMG2_HUMAN | DVSQSSISFQIEK | FALSE | 734.36992 | 751.398481 | 20 | 130 | 21.9 | y6 |
| sp\|P15309\|PPAP_HUMAN | SPIDTFPTDPIK | FALSE | 665.848092 | 1033.52005 | 20 | 130 | 18.4 | y9 |
| sp\|P15309\|PPAP_HUMAN | SPIDTFPTDPIK | FALSE | 665.848092 | 918.49311 | 20 | 130 | 18.4 | y8 |
| sp\|P15309\|PPAP_HUMAN | SPIDTFPTDPIK | FALSE | 665.848092 | 817.445431 | 20 | 130 | 18.4 | y7 |
| sp\|P15309\|PPAP_HUMAN | FQELESETLKSEEFQK | FALSE | 657.98971 | 895.451973 | 20 | 130 | 14.6 | y7 |
| sp\|P15309\|PPAP_HUMAN | FQELESETLKSEEFQK | FALSE | 657.98971 | 848.917432 | 20 | 130 | 14.6 | y14 |
| sp\|P15309\|PPAP_HUMAN | FQELESETLKSEEFQK | FALSE | 657.98971 | 784.396135 | 20 | 130 | 14.6 | y13 |
| sp\|P15309\|PPAP_HUMAN | ELSELSLLSLYGIHK | FALSE | 567.985572 | 930.540728 | 20 | 130 | 11.2 | y8 |
| sp\|P15309\|PPAP_HUMAN | ELSELSLLSLYGIHK | FALSE | 567.985572 | 817.456664 | 20 | 130 | 11.2 | y7 |
| sp\|P15309\|PPAP_HUMAN | ELSELSLLSLYGIHK | FALSE | 567.985572 | 730.424636 | 20 | 130 | 11.2 | y6 |
| sp\|P02808\|STAT_HUMAN | FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF | FALSE | 1215.233017 | 1687.7751 | 20 | 130 | 35.2 | y13 |
| sp\|P02808\|STAT_HUMAN | FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF | FALSE | 1215.233017 | 1462.66376 | 20 | 130 | 35.2 | y11 |
| sp\|P02808\|STAT_HUMAN | FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF | FALSE | 1215.233017 | 1229.56259 | 20 | 130 | 35.2 | b11 |
| sp\|P02808\|STAT_HUMAN | FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF | FALSE | 1215.233017 | 1583.71652 | 20 | 130 | 35.2 | b14 |
| sp\|P09228\|CYTT_HUMAN | IIEGGIYDADLNDER | FALSE | 846.907398 | 1337.5968 | 20 | 130 | 27.6 | y12 |
| sp\|P09228\|CYTT_HUMAN | IIEGGIYDADLNDER | FALSE | 846.907398 | 1110.46981 | 20 | 130 | 27.6 | y9 |
| sp\|P09228\|CYTT_HUMAN | IIEGGIYDADLNDER | FALSE | 846.907398 | 947.406479 | 20 | 130 | 27.6 | y8 |
| sp\|P09228\|CYTT_HUMAN | QLCSFQIYEVPWEDR | FALSE | 985.459275 | 1093.4949 | 20 | 130 | 34.7 | y8 |
| sp\|P09228\|CYTT_HUMAN | QLCSFQIYEVPWEDR | FALSE | 985.459275 | 801.388979 | 20 | 130 | 34.7 | y6 |
| sp\|P09228\|CYTT_HUMAN | QLCSFQIYEVPWEDR | FALSE | 985.459275 | 702.320565 | 20 | 130 | 34.7 | y5 |
| sp\|P02814\|SMR3B_HUMAN | GPYPPGPLAPPQPFGPGFVPPPPPPPYGPGR | FALSE | 776.15636 | 1228.64732 | 20 | 130 | 18.9 | y12 |
| sp\|P02814\|SMR3B_HUMAN | GPYPPGPLAPPQPFGPGFVPPPPPPPYGPGR | FALSE | 776.15636 | 850.445765 | 20 | 130 | 18.9 | b9 |
| sp\|P02814\|SMR3B_HUMAN | GPYPPGPLAPPQPFGPGFVPPPPPPPYGPGR | FALSE | 776.15636 | 1172.60987 | 20 | 130 | 18.9 | b12 |
| sp\|P02814\|SMR3B_HUMAN | IPPPPPAPYGPGIFPPPPPQP | FALSE | 710.718937 | 729.393001 | 20 | 130 | 16.5 | y7 |
| sp\|P02814\|SMR3B_HUMAN | IPPPPPAPYGPGIFPPPPPQP | FALSE | 710.718937 | 987.529829 | 20 | 130 | 16.5 | b10 |
| sp\|P02814\|SMR3B_HUMAN | IPPPPPAPYGPGIFPPPPPQP | FALSE | 710.718937 | 1141.60406 | 20 | 130 | 16.5 | b12 |

From the transition list in Table 12, a 16-minute method was developed. The assay was first run using pooled samples of saliva, seminal fluid, and vaginal fluid to alleviate variation between individuals. The pooled data were acquired during a single time segment where the instrument cycles between all precursor-product ion pairs during

the chromatographic run. Using this unsegmented method the instrument scans for all potential biomarkers across the entire chromatographic separation without taking into account the exact elution times for any one precursor-product ion pair. Monitoring many compounds simultaneously lowered the dwell (accumulation of ions) time for each individual precursor-product ion pair causing a loss in sensitivity. While this method was not optimal, it did allow for the discovery of the retention times for each of the peptides. Using the known retention times a method was developed to divide the targets into three time segments which allowed the instrument to only target ions which elute during that specific time segment – thereby increasing the allowable dwell time. **Table 13** shows all targeted peptides and their location in the three time segments.

**Table 13 – Segmented Method Dividing the MRM Acquisition into Three Time Segments.**

| Segment | Protein | Accession | Peptide | RT (min) |
|---|---|---|---|---|
| 2 | Ly6/PLAUR domain-containing protein 3 | O95274 | GCVQDEFCTR | 5.71 |
| | Ly6/PLAUR domain-containing protein 3 | O95274 | TLSESAEGACGSQESGSLHSGASQELGEGQR | 6.61 |
| | Cornulin | Q9UBG3 | ISPQIQLSGQTEQTQK | 7.6 |
| 3 | Ly6/PLAUR domain-containing protein 3 | O95274 | DGVTGPGFTLSGSCCQGSR | 8.5 |
| | Prostatic acid phosphatase | P15309 | FQELESETLKSEEFQK | 8.57 |
| | Cystatin_SA | P09228 | IIEGGIYDADLNDER | 8.957 |
| | Semenogelin-2 | Q02383 | DVSQSSISFQIEK | 9.16 |
| | Semenogelin-1 | P04279 | KQGGSQSSYVLQTEELVANK | 9.65 |
| | Prostatic acid phosphatase | P15309 | SPIDTFPTDPIK | 10.28 |
| | Neutrophil gelatinase-associated lipocalin | P80188 | SYPGLTSYLVR | 10.9 |
| 4 | Neutrophil gelatinase-associated lipocalin | P80188 | TFVPGCQPGEFTLGNIK | 11.85 |
| | Semenogelin-1 | P04279 | DIFSTQDELLVYNK | 11.93 |
| | Semenogelin-2 | Q02383 | DIFTTQDELLVYNK | 11.93 |
| | Cystatin_SA | P09228 | QLCSFQIYEVPWEDR | 13.26 |
| | Submaxillary gland | P02814 | IPPPPPAPYGPGIFPPPPPQP | 13.47 |
| | Prostatic acid phosphatase | P15309 | ELSELSLLSLYGIHK | 14 |
| | Statherin | P02808 | FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF | 14.41 |
| | Submaxillary gland | P02814 | GPYPPGPLAPPQPFGPGFVPPPPPPPYGPGR | 14.91 |

### 4.2.2 Method Evaluation

Using the segmented method, a series of single source samples were run to evaluate the consistency of the assay on the QQQ platform. Additionally, these samples

were used to develop preliminary interpretation guidelines using the Agilent Quantitative Analysis software to evaluate the subsequent results from the mixed and casework-type samples.

A representative example of results obtained using a QQQ-MRM multiplex assay to identify human seminal fluid can be seen in **Figure 39**. Examples of assay results obtained for a known saliva sample (**Figure 40**) and a known vaginal fluid sample (**Figure 41**) illustrate the specificity and accuracy the MRM-QQQ multiplex assay.

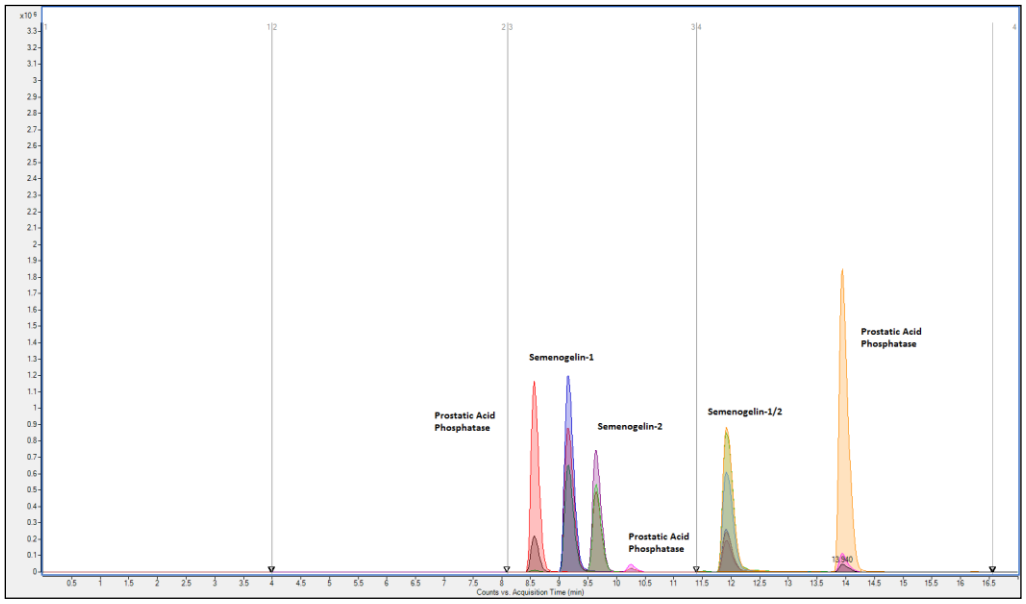



**Figure 39 - QQQ-MRM multiplex results showing seminal fluid. The assay achieved clear detection of all target ions indicating the high-confidence confirmatory identification of Human seminal fluid. No biomarkers indicative of other body fluids were detected.**

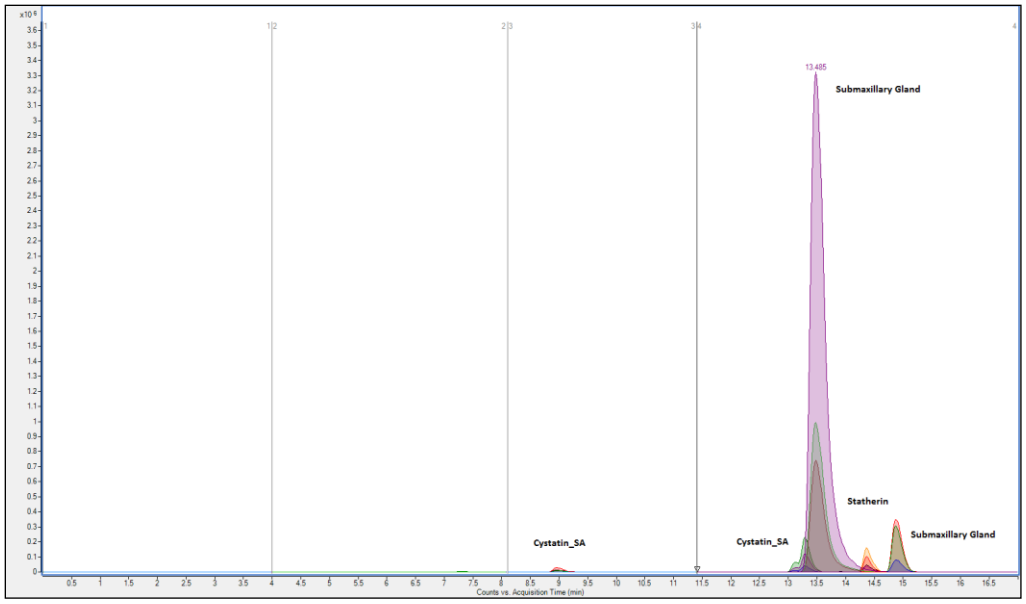



**Figure 40 - QQQ-MRM multiplex results showing a saliva hit. The assay achieved clear detection of all target ions indicating the high-confidence confirmatory identification of Human saliva. No biomarkers indicative of other body fluids were detected.**

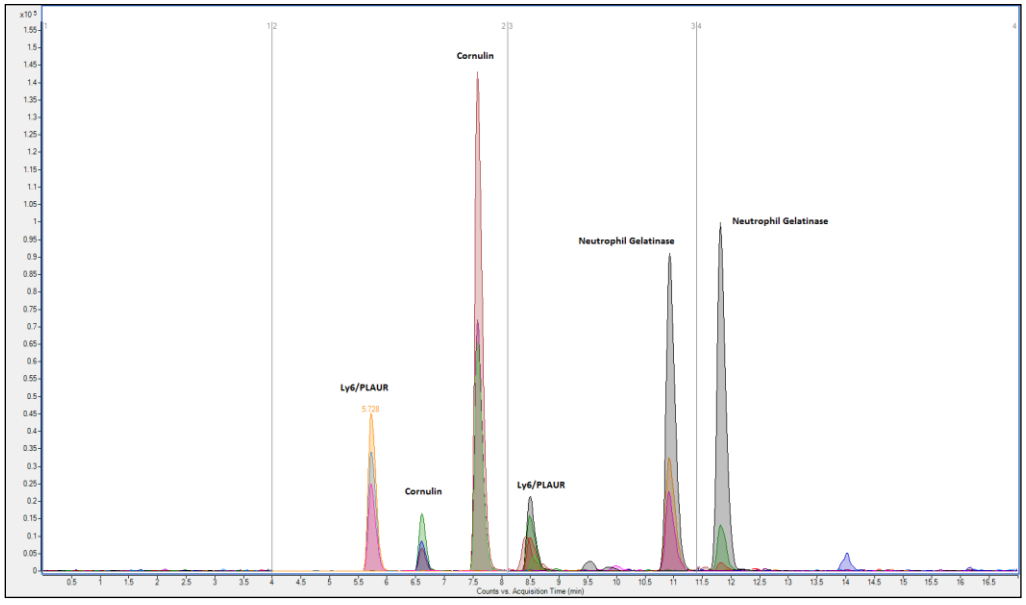



**Figure 41 - QQQ-MRM multiplex results of a three component mixture. The assay achieved clear detection of all target ions indicating the high-confidence confirmatory identification of Human vaginal fluid. No biomarkers indicative of other body fluids were detected.**

While the Q-TOF identification was confirmed through matching the entire product ion spectra to a protein database, the QQQ identification criteria was based on the response of only the selected transitions for the protein. Therefore, before analysis of the results, an analysis method was designed using Agilent's Quantitative Analysis software. To do this, a mixture of all three samples was run and loaded into the analysis software. From the mixed reference sample, the retention time as well as relative response ratios were determined and used to analyze the rest of the samples in the batch.

The outcomes of these experiments were consistent with results from previous studies conducted using a Q-TOF assay. However, because of the increased sensitivity associated with the QQQ platform, it became possible to detect protein biomarkers which might have fallen below the level of detection of the Q-TOF platform. As a result, the QQQ-MRM assay will provide a means of further confirming the specificity of the selected protein biomarkers.

**Tables 14 and 15** provide summaries of the results from the ten single-source assays using two different identification parameters. **Table 14** shows a low-stringency assay in which biomarker detection was based on the presence of at least one peptide per protein. In contrast to this, **Table 15** shows the assay results where detection was based on the confident identification of all peptides associated with that biomarker. Since each biomarker (aside from statherin) can be identified using two peptides and their associated fragments, the detection of both peptides represents a more stringent and thus more confident identification of a targeted protein biomarker.

**Table 14 - Frequency of protein identification in reference samples of saliva, seminal fluid and vaginal fluid (ten samples each). Protein identification in this table is based on at least one targeted peptide per biomarker.**

| Biomarker Identification - One or Two Peptides Detected | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Saliva | | | Seminal Fluid | | | Vaginal fluid | | |
| | | Cystatin_SA | Statherin | Submax Gland | Semenogelin-1 | Semenogelin-2 | Prostatic Acid | Cornulin | Ly6 | Neutrophil Gel |
| Samples | Saliva | 100.00% | 90.00% | 100.00% | 0.00% | 10.00% | 0.00% | 30.00% | 100.00% | 0.00% |
| | Seminal Fluid | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | 100.00% | 0.00% | 0.00% | 0.00% |
| | Vaginal fluid | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | 100.00% |

**Table 15 - Frequency of protein identification in reference samples of saliva, seminal fluid and vaginal fluid (ten samples each). Protein identification in this table is based on the detection of two targeted peptides per biomarker.**

| Biomarker Identificaion - Two Peptides Detected | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Saliva | | | Seminal Fluid | | | Vaginal fluid | | |
| | | Cystatin_SA | Statherin | Submax Gland | Semenogelin-1 | Semenogelin-2 | Prostatic Acid | Cornulin | Ly6 | Neutrophil Gel |
| Samples | Saliva | 100.00% | 90.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 80.00% | 0.00% |
| | Seminal Fluid | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | 100.00% | 0.00% | 0.00% | 0.00% |
| | Vaginal fluid | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | 100.00% |

Detection of both peptides for a protein identification (**Table 15**) results in an identification pattern similar to what was standardized using the Q-TOF assay – although the QQQ-MRM assay still achieved greater sensitivity. An example of the impact of the greater sensitivity QQQ-MRM platform was seen with the detetction of, Ly6 which had initially appeared promising as a vaginal fluid biomarker. While this protein had also been previously identified using a Q-TOF platform in a small number of casework-type samples which contained saliva, its robust and unambiguous identification using the QQQ-MRM was to be expected.

Biomarker identification based on the detection of only a single peptide results in a more complicated interpretation. Cornulin, which is involved with mucosal/epithelial immune response, was also detected with one peptide in three of the ten saliva reverence samples. While this protein had not been detected in previous experiments, it has been previously characterized in the esophagus where it is thought to have an association with oral cancer[212, 213]. Its presence in a salivary sample, therefore, is not unreasonable.

Detection of a semenogelin-2 peptide in a single saliva sample was unanticipated in an assumed “saliva only" sample and has never been seen in previous assays involving either the Q-TOF or QQQ. **Figure 42** shows the raw data comparing the semenogelin-2 peptide detection in saliva sample SA044 as well as a representative hit from seminal fluid sample SE022. By appearance alone, it is clear that three fragment ion peaks are present in both samples, but the response (peak area) of the peptide in the seminal fluid sample is approximately 1,700 times larger. While the peak height and response are significantly higher in the seminal fluid sample, other qualitative information leads to the conclusion that the peak detected in this one saliva sample is real regardless of whether it is an endogenous salivary protein or not.

Using quantitative analysis software (**Table 16**), the signal-to-noise ratio of the peptide was calculated to be 18.81 in SA044 compared to 255.57 in SE022. Typically, a lower limit of detection requires a signal-to-noise ratio greater than 5 to be valid. Additionally, the ratio between the responses of all three product ions must be within an accepted error limit. In this case. an uncertainly of 20% is acceptable compared to the reference samples. Lastly, the retention times must be the same. Using these qualifying parameters, there is every indication that this peptide is genuine in both the saliva and seminal fluid samples. While this result was not expected, there is some potential corroborating support in the biomedical literature. Using an RT-PCR approach, Semenogelin transcripts were detected in the trachea as well as the salivary gland[214]. Whether this represents a peptide that can be reproducibly detected in a broad range of humans at very low levels versus an unrecognized artifact in a single study remains to be

determined. In addition, it is possible that this result is merely an artifact from sample carryover. Moving forward, this must be assessed in a full developmental validation study.

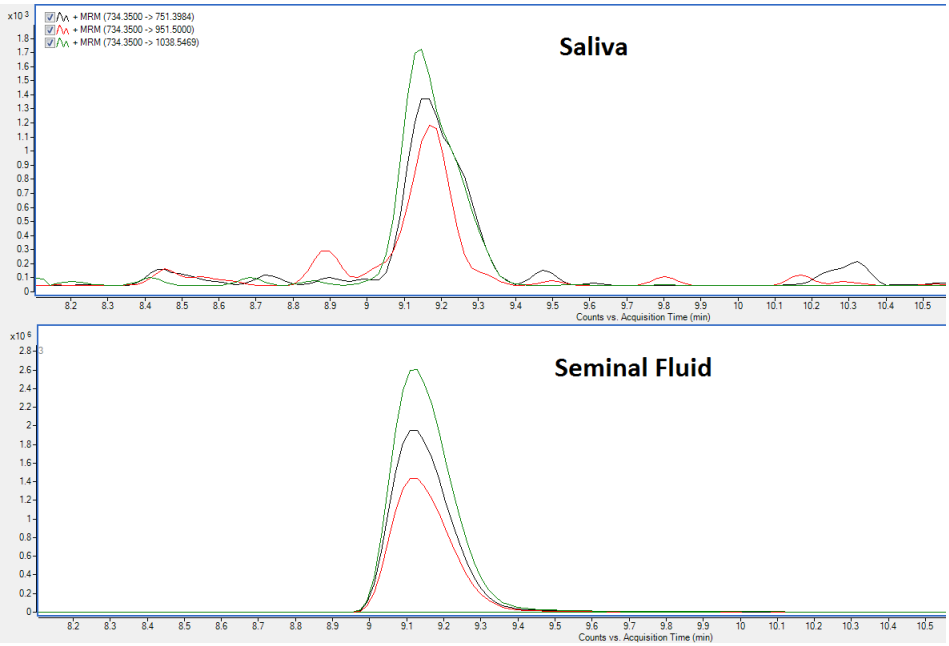



**Figure 42 - Detection of Semenogelin-2 peptide DVSQSSISFQIEK in saliva (top) and seminal fluid (bottom)**

**Table 16 - Quantitative software generated values from corresponding Semenogelin-2 peptide DVSQSSISFQIEK hit. This information is used to determine the accuracy of a peptide identification.**

| Sample | | | | | SMG_2 - DVSQSSISFQIEK Results | | | | | Qualifier (734.4 -> 951.5) Results | | | Qualifier (734.4 -> 751.4) Results | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Name | Data File | Type | Level | Acq. Date-Time | RT | Resp. | MI | Final Conc. | S/N | Ratio | MI | S/N | Ratio | MI | S/N |
| SA044_10ul | SA044_10ul.d | Sample | | 11/27/2012 4:58 PM | 9.147 | 16554 | | 0.0208 | 18.81 | 58.4 | | 19.44 | 83.3 | | 7.35 |
| SE022_10ul | SE022_10ul.d | Sample | | 11/27/2012 6:18 PM | 9.128 | 29382814 | | 1.7016 | 255.57 | 54.8 | | 278.81 | 74.5 | | 392.57 |

In order for a protein biomarker to be confidently identified, therefore, it is the opinion of the author that that good practice dictates that both peptides must be detected in a sample. It is also suggested that more samples be run in the future and that external standards be used as a quantitative means of assessing analyte concentrations. Finally, it should be noted that these samples were run on an older microflow LC-MS/MS system

and not on a nanoflow LC-MS/MS. If a nano LC-QQQ were to be employed on the same samples, sufficient sensitivity may be achieved to confidently ID the missing peptides and thus confidently demonstrate the presence of a given biomarker.

### 4.2.3 Casework-Type Samples

The QQQ-MRM multiplex body fluid identification assay has been successfully tested on a series of samples representing all targeted body fluids and a variety of casework-type biological stains. These included both single-source and mixed-source body fluid samples deposited onto a variety of substrates including denim, leather and nylon. To assess the potential impact of environmental/chemical insult on assay performance, biological stains were subjected to commonly encountered agents including gasoline, soil, laundry detergent and personal lubricants. The complete list of casework-type samples that were tested can be found in **Table 17**.

**Table 17 - QQQ-MRM Assay Results with Casework-Type Samples[a].**

| | | Biomarkers | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Saliva | | | Seminal Fluid | | | Vaginal fluid | | |
| | | Cystatin_SA | Statherin | Submaxi Gland | Semenogelin-1 | Semenogelin-2 | Prostatic Acid | Cornulin | Ly6 | Neutrophil Gel |
| Substrates | CW01 - Saliva Swab | | | | | | | | | |
| | CW02 - Semen Swab | | | | | | | | | |
| | CW03 - Vaginal Swab | | | | | | | | | |
| | CW04 - Vaginal Fluid on Cotton | | | | | | | | | |
| | CW05 - Vaginal fluid on Nylon | | | | | | | | | |
| | CW06 - Seminal Fluid on Denim | | | | | | | | | |
| | CW07 - Seminal fluid on Leather | | | | | | | | | |
| | CW08 - Saliva on Cigraette | | | | | | | | | |
| | CW09 - Saliva on Bottle | | | | | | | | | |
| Contaminants | CW10 - Semain Fluid with Spermicide | | | | | | | | | |
| | CW11 - Saliva with Chewing Tobacco | | | | | | | | | |
| Mixtures | CW12 - Mixture Swab 1SA:3VS | | | | | | | | | |
| | CW13 - Mixture Swab 1SA:1VS | | | | | | | | | |
| | CW14 - Mixture Swab 3SA:1VS | | | | | | | | | |
| | CW15 - Mixture Swab 1SE:3SA | | | | | | | | | |
| | CW16 - Mixture Swab 1SE:1SA | | | | | | | | | |
| | CW17 - Mixture Swab 3SE:1SA | | | | | | | | | |
| | CW18 - Mixture Swab 1SE:3VS | | | | | | | | | |
| | CW19 - Mixture Swab 1SE:1VS | | | | | | | | | |
| | CW20 - Mixture Swab 3SE:1VS | | | | | | | | | |
| | CW21 - 3 Component Mixture Swab | | | | | | | | | |
| Dilution | CW22 - 3 Component 2 Fold Dilution | | | | | | | | | |
| | CW23 - 3 Component 4 Fold Dilution | | | | | | | | | |
| | CW24 - 3 Component 8 Fold Dilution | | | | | | | | | |

[a]Positive protein identification requires both target peptides to be valid.

The data obtained using casework-type samples were consistent with expected results as well as the results of previous Q-TOF studies. Irrespective of substrate and/or chemical contaminant, all of the targeted protein biomarkers appeared in the appropriate sample. Mixture studies were performed using cotton swabs onto which 2-component body fluid mixtures were prepared at three volumetric different ratios (1:3, 1:1, 3:1). At every mixture ratio assayed, all expected biomarkers were detected. The only identification of a non-specific protein occurred with the 1:1 and 1:3 seminal fluid:saliva mixture swabs where the vaginal fluid marker Ly6 was detected. This has been previously observed on the Q-TOF as well with the QQQ reference samples. This marker is clearly non-specific for vaginal fluids and thus should be removed in future studies. Finally additional dilution studies to assess matrix effects on the lower limit of detection for each body fluid are clearly indicated as a focus for additional study.

### 4.2.4 Combined DNA and Serological Analysis

The current research has shown that biological stain identification using a QQQ-MRM to assay material from a variety of casework-type samples is both precise and reliable. However, in order for a mass spectrometry assay for serological analysis to be implemented into routine forensic casework it must also be possible to recover a DNA profile from the same sample. To demonstrate the feasibility of this, a modified protocol similar to a differential extraction was developed (**Figure 43)**. Similar to previous casework-type samples, unfractionated body fluid aliquots were added to cotton swabs and dried troroughly. The cotton swab containing the dried biological material was then rehydrated in PBS and place in a microfuge spin basket to separate and recover the liquid

form the cottom matrix of the swab. A high-speed centrifugation step was used to pellet cellular and any other insoluble material. The protein rich supernatant fraction was processed for QQQ-MRM analysis while the cellular (DNA rich) fraction was extracted using the EZ1 BioRobot, quantified using Quantifiler® Duo, and used to develop an STR profile with the Identifiler® Kit.

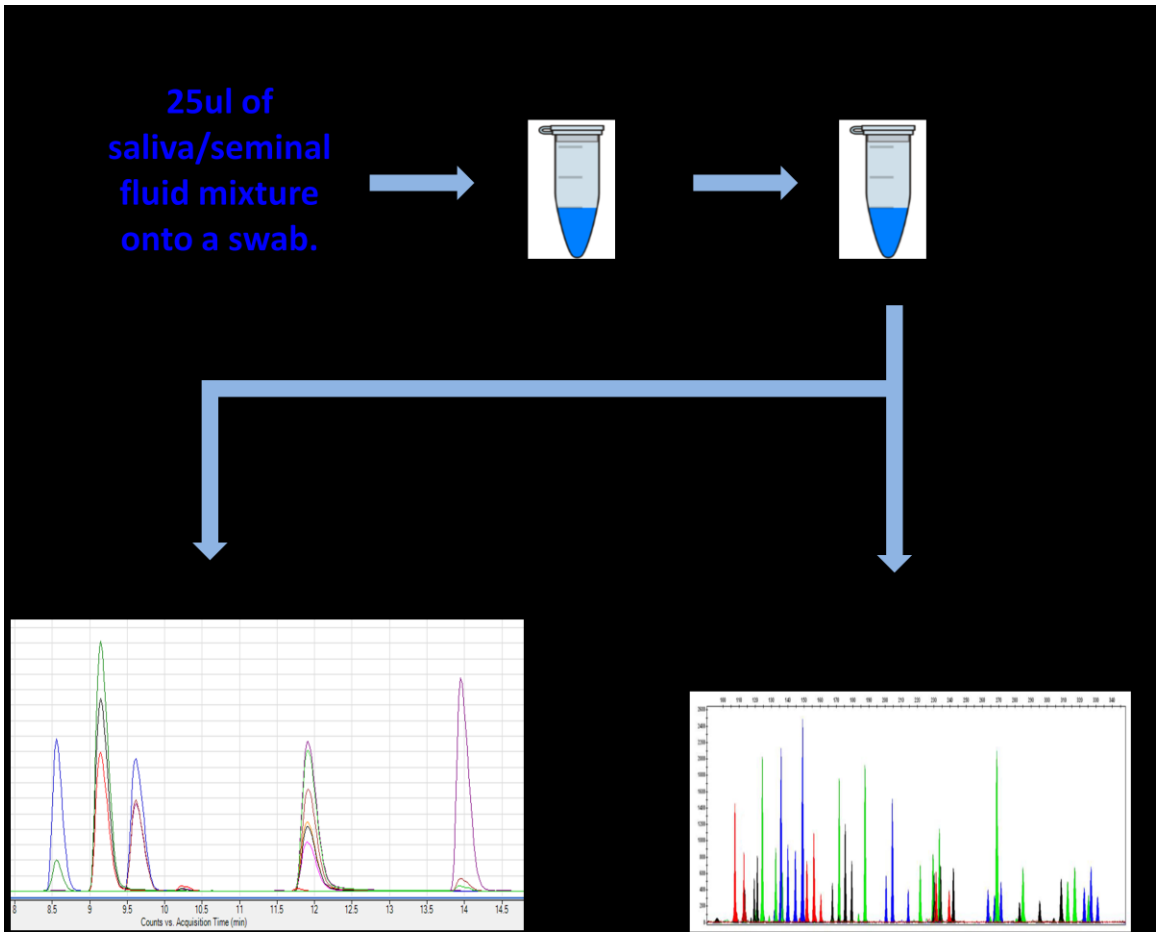



**Figure 43 – Generation of a full STR profile and stain identification for a mixed stain containing seminal fluid and saliva mixture. Swab was rehydrated in phosphate buffer saline and placed in a filter basket to remove retained fluid. Sample was then spun at high speed to pellet the cellular material for STR analysis The DNA-depleted supernatant was processed for proteomic analysis. DNA was quantified using ABI Quantifiler® duo and a full DNA profile was generated using ABI Identifiler® Plus kit. Protein concentration was determined using Pierce micro BCA assay kit, and confirmatory stain identification was achieved using the Q-TOF multiplex assay.**

In total, six swabs were processed using the modified protocol – two reference samples from semen and vaginal fluid as well as four male/female mixed samples. For the mixed samples a semen/saliva swab, semen/vaginal fluid swab, saliva/vaginal fluid swab, as well as a semen/peripheral blood swab were prepared. QQQ-MRM analyses of these swabs resulted in the successful identification of all targeted proteins from both single- and mixed-source body fluid samples (**Table 18**).

**Table 18 – Results of DNA/Protein samples processed for joint genetic and serological analysis. For each swab the fluid was confidently identified. Green cells indicate positive identification of target protein via two specific peptides. Red cells indicate no peptides were found. (M) and (F) represents male and female source, respectively.**

| | Biomarkers | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Saliva | | | Seminal Fluid | | | Vaginal fluid | | |
| | Cystatin_SA | Statherin | Submaxi Gland | Semenogelin-1 | Semenogelin-2 | Prostatic Acid | Cornulin | Ly6 | Neutrophil Gel |
| Semen (M) Saliva (F) | | | | | | | | | |
| Semen (M) Vaginal (F) | | | | | | | | | |
| Saliva (M) Vaginal (F) | | | | | | | | | |
| Semen (M) Peripheral Blood (F) | | | | | | | | | |
| Semen Refrence (M) | | | | | | | | | |
| Vaginal Refrence (F) | | | | | | | | | |

Genetic analyses (*i.e.*, STR-based DNA profiles) using the Identifiler® Plus kit were able to successfully produce the expected single- and mixed-source profiles for all samples tested (**Table 19**). In all reference and mixed-source samples, the major contributor was represented by a complete STR profile. The saliva/vaginal fluid sample also resulted in a full profile for the minor contributor. In the semen/saliva as well as the semen/vaginal fluid samples, 50% of the minor contributor's distinct alleles (*i.e.*, not counting shared alleles) were present. In the semen/peripheral blood sample 79% of the minor contributor's distinct alleles were present. The end result of this experiment provides ample support for the proposition that a mass spectrometry-based serological assay is compatible with existing STR profiling methodologies.

**Table 19 – Summary Table of STR DNA Results using the Identifiler kit. "NR"= No results, "NT"= Not tested, "(M)"=Major contributor, "(m)"= Minor contributor.**

| *EXHIBIT NUMBER* | D8S1179 | D21S11 | D7S820 | CSF1PO | D3S1358 | TH01 | D13S317 | D16S539 | D2S1338 | D19S433 | vWA | TPOX | D18S51 | AMEL | D5S818 | FGA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Male Donor (Semen) | 10,13 | 31.2, 32.2 | 10,12 | 10,11 | 15,17 | 6,7 | 10,12 | 11,12 | 17,24 | 13,14 | 17,18 | 8,11 | 11,15 | X,Y | 11,12 | 20,24 |
| Female Donor (Vaginal) | 14,15 | 29,30 | 10,11 | 10,11 | 15 | 6,9.3 | 11,12 | 9,12 | 19,21 | 13,15 | 17,18 | 8 | 16 | X | 11 | 25,27 |
| Semen/Saliva (M) | 10,13 | 31.2, 32.2 | 10,12 | 10,11 | 15,17 | 6,7 | 10,12 | 11,12 | 17,24 | 13,14 | 17,18 | 8,11 | 11,15 | X,Y | 11,12 | 20,24 |
| Semen/Saliva (m) | 14,15 | 30 | 11 | | | 9.3 | 11 | | 19 | | | | | | | |
| Semen/Vaginal (M) | 10,13 | 31.2, 32.2 | 10,12 | 10,11 | 15,17 | 6,7 | 10,12 | 11,12 | 17,24 | 13,14 | 17,18 | 8,11 | 11,15 | X,Y | 11,12 | 20,24 |
| Semen/Vaginal (m) | 14,15 | | 11 | | | | 11 | | 21 | | | | 16 | | | 25 |
| Saliva/Vaginal (M) | 10,13 | 31.2, 32.2 | 10,12 | 10,11 | 15,17 | 6,7 | 10,12 | 11,12 | 17,24 | 13,14 | 17,18 | 8,11 | 11,15 | X,Y | 11,12 | 20,24 |
| Saliva/Vaginal (m) | 14,15 | 29,30 | 11 | | | 9.3 | 11 | 9 | 19,21 | 15 | | | 16 | | | 25,27 |
| Semen/PBlood (M) | 14,15 | 29,30 | 10,11 | 10,11 | 15 | 6,9.3 | 11,12 | 9,12 | 19,21 | 13,15 | 17,18 | 8 | 16 | X | 11 | 25,27 |
| Semen/PBlood (m) | 10,13 | 31.2, 32.2 | 12 | | 17 | | 10 | 11 | 17 | 14 | | 11 | 11,15 | Y | 12 | 20,24 |

## 4.3 Conclusion

The central objective of this phase of the dissertation research was to improve upon the previous Q-TOF assay by increasing the speed of analysis and to test the feasibility of a mass-spectrometry based assay for stain identification for routine casework testing. A prototype QQQ-MRM "sexual assault assay" was designed to provide for the rapid and confirmatory identity of seminal fluid, saliva, and vaginal fluid.

Starting with the data from the Q-TOF studies, the four most promising biomarkers for the three fluids were selected for use in the prototype QQQ-MRM assay. Because these proteins (and subsequent peptides) had already been evaluated previously, the only task that remained was to identify the optimal product ions for inclusion in the QQQ-MRM assay. Once selected, the transitions were integrated into a QQQ-MRM triplex biological stain assay. The assay accurately identified seminal fluid, saliva, and vaginal secretions on the basis of 12 high-specificity protein biomarkers with an instrument run time of 18 minutes. Compared to the previous Q-TOF assay, this achieves more than a 75% reduction in total analysis time.

The QQQ-MRM three-stain assay was tested using single-source samples (7 samples each) as well as a series of casework-type samples. Data from these experiments confirmed the accuracy and consistency of the biomarkers on this high-speed platform. In addition to the single source laboratory samples, a protocol was developed to co-extract the DNA and protein fractions from a single swab. This protocol was able to identify the specific body fluid stains present as well as to generate accurate STR profiles. In summary, a high speed QQQ assay for stain identification can act as a launching point for the subsequent full developmental validation of a six-body fluid multiplex for forensic casework on a QQQ-MRM platform.

## 4.4 Future Directions and Impact on the Criminal Justice System

Forensic practitioners in the US have repeatedly stressed that the identification of biological stains can be a significant challenge for forensic serologists. Currently, there

are no means of readily or reliably identifying stains such as vaginal secretions or differentiating between peripheral vs. menstrual blood. Additionally, many tests provide only presumptive results (e.g., α-amylase assay for saliva) and can consume significant quantities of precious evidence.

A sensitive, multiplex compatible, reliable, human-specific approach for the confirmatory identification of questioned biological stains could overcome the limitations of the methods that analysts now employ. Therefore, future work should focus on the developmental validation of a mass spectrometry based multiplex assay for forensic stain identification. The obvious choice for this validation will employ a single, well-established analytical technology – targeted-ion mass spectrometry on a triple quadrupole mass spectrometer – a technology which was shown to be reliable in this dissertation research and is already employed in many forensic laboratories for toxicology and drug identification.

The implementation of this technology has the potential to modernize serological screening. As compared to traditional methodologies, results obtained are confirmatory for human stains through the identification of multiple fluid-specific protein biomarkers. Additionally, stain detection is far more reliable since identification is based on the unique chemical composition and subsequent mass spectra. Thus, assay results are not compromised by dependence on a single diagnostic protein, antibody cross reactivity, false negative results from undetected hook effects, or potentially compromised results from the visualization of chemical or enzyme reactions. The multiplex design eliminates the need to perform separate tests on an unknown stain. In addition, the entire assay from

sample extraction to data analysis is more amenable to automation and the implementation of robotics. This makes mass spectrometry-based testing better suited for batch processing.

This work has the potential to significantly improve the accuracy and sensitivity of forensic serological testing. In short, the successful completion of the developmental validation and downstream implementation of this research will provide the forensic community with a powerful tool to aid in criminal investigations.

## Literature Cited


1. Landsteiner, K., *Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe.* Zbl Bakt, 1900. **27**(10): p. 357-362.
2. *L'individualità del sangue nella biologia, nella clinica e nella medicina legale.* Journal of the American Medical Association, 1924. **82**(4): p. 329-329.
3. Bell, S. and R.C. Shaler, *Encyclopedia of Forensic Science, Revised Edition*. 2009: Facts On File, Incorporated.
4. Tilstone, W.J., K.A. Savage, and L.A. Clark, *Forensic Science: An Encyclopedia of History, Methods and Techniques*. 2006: Abc-Clio Incorporated.
5. Budowle, B., et al., *Subtyping phosphoglucomutase-1 in semen stains and bloodstains: a report on the method.* J Forensic Sci, 1986. **31**(4): p. 1341-8.
6. Budowle, B., *A method for subtyping group-specific component in bloodstains.* Forensic Sci Int, 1987. **33**(3): p. 187-96.
7. Semikhodskii, A., *Dealing with DNA Evidence: A Legal Guide*. 1 ed. 2007: Routledge-Cavendish. 200.
8. MEYER, E., *Beiträge zur Leukocytenfrage.* Münchner Medizinische Wochenschrift, 1903. **50**(35): p. 1489-1493
9. Shedd, J.H.K.a.M., *Phenolphthalin as a reagent for the oxidizing ferments.* American Chemical Journal, 1901. **26**: p. 526-539
10. R V Winchester, H.W. *BLOOD DETECTION BY CHEMICAL METHODS*
11. Werner, O., *Crime Investigation. Physical Evidence and the Police Laboratory von Paul L. Kirk. Interscience Publishers Inc., New York 1953, 1. Aufl. XXII, 784 S., 151 Abb., gebd. $ 10.–.* Angewandte Chemie, 1953. **65**(20): p. 524-524.
12. Saferstein, R., *Forensic science handbook*. 2nd ed. 2002, Upper Saddle River, NJ: Prentice Hall.
13. Berg JM, T.J., Stryer L., *Biochemistry*, in *Section 8.1, Enzymes Are Powerful and Highly Specific Catalysts*. 2002, W H Freeman: New York.
14. Bull, H., et al., *Acid phosphatases.* Mol Pathol, 2002. **55**(2): p. 65-72.
15. Cooper, J.F., A. Foti, and H. Herschman, *Combined serum and bone marrow radioimmunoassays for prostatic acid phosphatase.* J Urol, 1979. **122**(4): p. 498-502.
16. Macfarlane, G.T., J.H. Cummings, and C. Allison, *Protein degradation by human intestinal bacteria.* J Gen Microbiol, 1986. **132**(6): p. 1647-56.
17. Verbeke, P., B.F. Clark, and S.I. Rattan, *Modulating cellular aging in vitro: hormetic effects of repeated mild heat stress on protein oxidation and glycation.* Exp Gerontol, 2000. **35**(6-7): p. 787-94.
18. Basle, E., N. Joubert, and M. Pucheault, *Protein chemical modification on endogenous amino acids.* Chem Biol, 2010. **17**(3): p. 213-27.
19. Hochmeister, M.N., et al., *Evaluation of prostate-specific antigen (PSA) membrane test assays for the forensic identification of seminal fluid.* J Forensic Sci, 1999. **44**(5): p. 1057-60.

20. Pang, B.C. and B.K. Cheung, *Identification of human semenogelin in membrane strip test as an alternative method for the detection of semen.* Forensic Sci Int, 2007. **169**(1): p. 27-31.
21. Hoofnagle, A.N. and M.H. Wener, *The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry.* J Immunol Methods, 2009. **347**(1-2): p. 3-11.
22. Fischer, A.H., et al., *Hematoxylin and eosin staining of tissue and cell sections.* CSH Protoc, 2008. **2008**: p. pdb prot4986.
23. Allery, J.P., et al., *Cytological detection of spermatozoa: comparison of three staining methods.* J Forensic Sci, 2001. **46**(2): p. 349-51.
24. Miller, K.W., et al., *Developmental validation of the SPERM HY-LITER kit for the identification of human spermatozoa in forensic samples.* J Forensic Sci, 2011. **56**(4): p. 853-65.
25. De Moors, A., et al., *Validation of the fluorescence-based Sperm Hy-Liter™ kit as a means to standardize spermatozoa identification in sexual assault cases.* Forensic Science International: Genetics Supplement Series, 2011. **3**(1): p. e31-e32.
26. Mann, T. and C. Lutwak-Mann, *Male reproductive function and semen: themes and trends in physiology, biochemistry, and investigative andrology*. 1981: Springer-Verlag.
27. Andrology, A.S.o., *Handbook of andrology*. 1995: The American Society of Andrology.
28. Owen, D.H. and D.F. Katz, *A review of the physical and chemical properties of human semen and the formulation of a semen simulant.* J Androl, 2005. **26**(4): p. 459-69.
29. Mortimer, D., *Practical laboratory andrology*. 1994: Oxford University Press, Incorporated.
30. Rees, B. and T.J. Rothwell, *The identification of phosphoglucomutase isoenzymes in semen stains and its use in forensic casework investigation.* Med Sci Law, 1975. **15**(4): p. 284-93.
31. Thomas, F.a.v.H., W., *The demonstration of recent sexual intercourse in the male by the Lugol method.* . Medicine, Science and the Law 1963. **3**: p. 169-171.
32. Rothwell, T.J. and K.J. Harvey, *The limitations of the Lugol's iodine staining technique for the identification of vaginal epithelial cells.* J Forensic Sci Soc, 1978. **18**(3-4): p. 181-4.
33. Hausmann, R., C. Pregler, and B. Schellmann, *The value of the Lugol's iodine staining technique for the identification of vaginal epithelial cells.* Int J Legal Med, 1994. **106**(6): p. 298-301.
34. Hausmann, R. and B. Schellmann, *Forensic value of the Lugol's staining method: further studies on glycogenated epithelium in the male urinary tract.* Int J Legal Med, 1994. **107**(3): p. 147-51.
35. French, C.E., et al., *A novel histological technique for distinguishing between epithelial cells in forensic casework.* Forensic Sci Int, 2008. **178**(1): p. 1-6.

36. Schumacher, G.F.B. and W.-B. Schill, *Radial diffusion in gel for micro determination of enzymes: II. Plasminogen activator, elastase, and nonspecific proteases.* Analytical Biochemistry, 1972. **48**(1): p. 9-26.
37. Gaensslen, R.E. and N.I.o. Justice, *Sourcebook in forensic serology, immunology, and biochemistry*. 1983: U.S. Dept. of Justice, National Institute of Justice.
38. L.J. Blum, P.E., S. Rocquefelte, *A new high-performance reagent and procedure for latent bloodstain detection based on luminol chemiluminescence.* Can. Soc. Forensic Sci. J., 2006. **39** p. 81–100.
39. Wecht, C.H. and J.T. Rago, *Forensic Science And the Law: Investigative Applications in Criminal, Civil, And Family Justice*. 2006: CRC/Taylor & Francis.
40. Schlenk, D., J.L. Gollon, and B.R. Griffin, *Efficacy of Copper Sulfate for the Treatment of Ichthyophthiriasis in Channel Catfish.* Journal of Aquatic Animal Health, 1998. **10**(4): p. 390-396.
41. Schweers, B.A., et al., *Developmental validation of a novel lateral flow strip test for rapid identification of human blood (Rapid Stain Identification--Blood).* Forensic Sci Int Genet, 2008. **2**(3): p. 243-7.
42. Chasis, J.A. and N. Mohandas, *Red blood cell glycophorins.* Blood, 1992. **80**(8): p. 1869-1879.
43. Sikirzhytski, V., A. Sikirzhytskaya, and I.K. Lednev, *Multidimensional Raman spectroscopic signature of sweat and its potential application to forensic body fluid identification.* Analytica chimica acta, 2012. **718**: p. 78-83.
44. Asano, M., M. Oya, and M. Hayakawa, *Identification of menstrual blood stains by the electrophoretic pattern of lactate dehydrogenase isozymes.* Forensic Science, 1972. **1**(3): p. 327-332.
45. Divall, G.B. and M. Ismail, *Lactate dehydrogenase isozymes in vaginal swab extracts: A problem for the identification of menstrual blood.* Forensic Science International, 1983. **21**(2): p. 139-147.
46. Altman, P.L. and D.D. Katz, *Blood and other body fluids*. 1961: Federation of American Societies for Experimental Biology.
47. Kozu, T., et al., *Identification of urine stains by a urease spray reagent.* Rep Natl Res Inst Police Sci, 1977. **30**(1): p. 18-20.
48. Sikirzhytskaya, A., V. Sikirzhytski, and I.K. Lednev, *Raman spectroscopy coupled with advanced statistics for differentiating menstrual and peripheral blood.* J Biophotonics, 2012.
49. *Biology Methods Manual*. 1978: Metropolitan Police Forensic Science Laboratory.
50. *Protocol Manual*. 1989: FBI Laboratory Serology Unit.
51. Hochmeister, M.N., et al., *Validation studies of an immunochromatographic 1-step test for the forensic identification of human blood.* J Forensic Sci, 1999. **44**(3): p. 597-602.
52. CRICK, F., *Central Dogma of Molecular Biology*. Nature. **227**(5258): p. 561-563.
53. Buel, E., Noreault-Conti, T., *Bringing tissue identification into the 21st century: mRNA analysis as the next molecular biology revolution in forensic science?*

Forensic Genetics Research Progress, ed. F. Gonzalez-Andrade. Vol. 1. 2009: Nova Publishers.
54. Juusola, J. and J. Ballantyne, *Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification.* Forensic Sci Int, 2003. **135**(2): p. 85-96.
55. Juusola, J. and J. Ballantyne, *mRNA profiling for body fluid identification by multiplex quantitative RT-PCR.* J Forensic Sci, 2007. **52**(6): p. 1252-62.
56. Setzer, M., J. Juusola, and J. Ballantyne, *Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains.* J Forensic Sci, 2008. **53**(2): p. 296-305.
57. Alvarez, M., J. Juusola, and J. Ballantyne, *An mRNA and DNA co-isolation method for forensic casework samples.* Anal Biochem, 2004. **335**(2): p. 289-98.
58. Fleming, R.I. and S. Harbison, *The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids.* Forensic Sci Int Genet, 2010. **4**(4): p. 244-56.
59. He, L. and G.J. Hannon, *MicroRNAs: small RNAs with a big role in gene regulation.* Nat Rev Genet, 2004. **5**(7): p. 522-31.
60. Hanson, E.K., H. Lubenow, and J. Ballantyne, *Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs.* Anal Biochem, 2009. **387**(2): p. 303-14.
61. Zubakov, D., et al., *MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation.* Int J Legal Med, 2010. **124**(3): p. 217-26.
62. Courts, C. and B. Madea, *Micro-RNA – A potential for forensic science?* Forensic Science International, 2010. **203**(1–3): p. 106-111.
63. Fazzari, M.J. and J.M. Greally, *Introduction to epigenomics and epigenome-wide analysis.* Methods Mol Biol, 2010. **620**: p. 243-65.
64. Virkler, K. and I.K. Lednev, *Blood species identification for forensic purposes using Raman spectroscopy combined with advanced statistical analysis.* Anal Chem, 2009. **81**(18): p. 7773-7.
65. Ng, H.H. and A. Bird, *DNA methylation and chromatin modification.* Curr Opin Genet Dev, 1999. **9**(2): p. 158-63.
66. DeAngelis, J.T., W.J. Farrington, and T.O. Tollefsbol, *An overview of epigenetic assays.* Mol Biotechnol, 2008. **38**(2): p. 179-83.
67. Lee, H.Y., et al., *Potential forensic application of DNA methylation profiling to body fluid identification.* International journal of legal medicine, 2012. **126**(1): p. 55-62.
68. Frumkin, D., et al., *DNA methylation-based forensic tissue identification.* Forensic science international Genetics, 2011. **5**(5): p. 517-24.
69. Wasserstrom, A., et al., *Demonstration of DSI-semen--A novel DNA methylation-based forensic semen identification assay.* Forensic Sci Int Genet, 2013. **7**(1): p. 136-42.
70. Larue, B.L., J.L. King, and B. Budowle, *A validation study of the Nucleix DSI-Semen kit-a methylation-based assay for semen identification.* Int J Legal Med, 2013. **127**(2): p. 299-308.

71. Chi-Chang Lin, M.-T.K., Hsien-Chang Chang, *Review: Raman Spectroscopy – A Novel Tool for Noninvasive Analysis of Ocular Surface Fluid* Journal of Medical and Biological Engineering, 2010. **30**(6): p. 343-354.
72. Virkler, K. and I.K. Lednev, *Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene.* Forensic Science International, 2009. **188**(1–3): p. 1-17.
73. Sikirzhytski, V., A. Sikirzhytskaya, and I.K. Lednev, *Multidimensional Raman spectroscopic signatures as a tool for forensic identification of body fluid traces: a review.* Appl Spectrosc, 2011. **65**(11): p. 1223-32.
74. TA, B., *Genomes*, in *Chapter 3 Transcriptomes and Proteomes*. 2002, Oxford: Wiley-Liss
75. Bensimon, A., A.J. Heck, and R. Aebersold, *Mass spectrometry-based proteomics and network biology.* Annu Rev Biochem, 2012. **81**: p. 379-405.
76. Rifai, N., M.A. Gillette, and S.A. Carr, *Protein biomarker discovery and validation: the long and uncertain path to clinical utility.* Nat Biotechnol, 2006. **24**(8): p. 971-83.
77. Prinz, M., et al., *Establishment of a Fast and Accurate Proteomic Method for Body Fluid/Cell Type Identification*. 2011, United States.
78. Suckau, D., et al., *A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics.* Anal Bioanal Chem, 2003. **376**(7): p. 952-65.
79. Mechthild Prinz., Y.T., Donald Siegel, Heyi Yang, Bo Zhou, Haiteng Deng. *Establishment of a Fast and Accurate Proteomic Method for Body Fluid/Cell Type Identification*. 2011 [cited 2012; Available from: https://www.ncjrs.gov/pdffiles1/nij/grants/236538.pdf.
80. Cappellini, E., et al., *Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins.* Journal of proteome research, 2012. **11**(2): p. 917-26.
81. Lindgren, J., et al., *Microspectroscopic evidence of cretaceous bone proteins.* PLoS One, 2011. **6**(4): p. e19445.
82. Johnston, N.L., et al., *Multivariate analysis of RNA levels from postmortem human brains as measured by three different methods of RT-PCR.* Journal of Neuroscience Methods, 1997. **77**(1): p. 83-92.
83. Baker, D.J., E.A. Grimes, and A.J. Hopwood, *D-dimer assays for the identification of menstrual blood.* Forensic science international, 2011. **212**(1-3): p. 210-4.
84. Carr, S.A. and L. Anderson, *Protein quantitation through targeted mass spectrometry: the way out of biomarker purgatory?* Clin Chem, 2008. **54**(11): p. 1749-52.
85. Anderson, L., *Candidate-based proteomics in the search for biomarkers of cardiovascular disease.* J Physiol, 2005. **563**(Pt 1): p. 23-60.
86. Alban, A., et al., *A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard.* Proteomics, 2003. **3**(1): p. 36-44.

87. Van den Bergh, G. and L. Arckens, *Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel-based proteomics.* Curr Opin Biotechnol, 2004. **15**(1): p. 38-43.
88. Liu, H., D. Lin, and J.R. Yates, 3rd, *Multidimensional separations for protein/peptide analysis in the post-genomic era.* Biotechniques, 2002. **32**(4): p. 898, 900, 902 passim.
89. Schlautman, J.D., et al., *Multidimensional protein fractionation using ProteomeLab PF 2D for profiling amyotrophic lateral sclerosis immunity: A preliminary report.* Proteome Sci, 2008. **6**: p. 26.
90. Ong, S.E. and M. Mann, *A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC).* Nat Protoc, 2006. **1**(6): p. 2650-60.
91. Nirmalan, N., P.F. Sims, and J.E. Hyde, *Quantitative proteomics of the human malaria parasite Plasmodium falciparum and its application to studies of development and inhibition.* Mol Microbiol, 2004. **52**(4): p. 1187-99.
92. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.
93. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.
94. Boyd, B., et al., *Trace quantitative analysis by mass spectrometry*. 2008: John Wiley & Sons.
95. Zhu, W., J.W. Smith, and C.M. Huang, *Mass spectrometry-based label-free quantitative proteomics.* J Biomed Biotechnol, 2010. **2010**: p. 840518.
96. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Anal Bioanal Chem, 2007. **389**(4): p. 1017-31.
97. Domon, B. and R. Aebersold, *Options and considerations when selecting a quantitative proteomics strategy.* Nat Biotechnol, 2010. **28**(7): p. 710-21.
98. Yocum, A.K. and A.M. Chinnaiyan, *Current affairs in quantitative targeted proteomics: multiple reaction monitoring-mass spectrometry.* Brief Funct Genomic Proteomic, 2009. **8**(2): p. 145-57.
99. Elschenbroich, S. and T. Kislinger, *Targeted proteomics by selected reaction monitoring mass spectrometry: applications to systems biology and biomarker discovery.* Mol Biosyst, 2011. **7**(2): p. 292-303.
100. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by Mass Spectrometry: Approaches, Advances, and Applications.* Annual Review of Biomedical Engineering, 2009. **11**(1): p. 49-79.
101. Bradshaw, T., *A User's Guide: Introduction to Peptide and Protein HPLC.* Phenomenex, USA, 2000.
102. Watson, J.T. and O.D. Sparkman, *Introduction to mass spectrometry: instrumentation, applications, and strategies for data interpretation*. 2007: Wiley.
103. Kebarle, P. and U.H. Verkerk, *Electrospray: From ions in solution to ions in the gas phase, what we know now.* Mass Spectrometry Reviews, 2009. **28**(6): p. 898-917.
104. Han, X., A. Aslanian, and J.R. Yates, 3rd, *Mass spectrometry for proteomics.* Curr Opin Chem Biol, 2008. **12**(5): p. 483-90.

105. Hu, Q., et al., *The Orbitrap: a new mass spectrometer.* J Mass Spectrom, 2005. **40**(4): p. 430-43.
106. Finklea, H. and R. Meyers, *Encyclopedia of analytical Chemistry.* Meyers, RA, Ed, 2000: p. 11848-11872.
107. Wong, P.S. and R. Graham Cooks, *Ion trap mass spectrometry.* Current separations, 1997. **16**: p. 85-92.
108. Domon, B. and R. Aebersold, *Mass Spectrometry and Protein Analysis.* Science, 2006. **312**(5771): p. 212-217.
109. Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, *An introduction to quadrupole–time-of-flight mass spectrometry.* Journal of Mass Spectrometry, 2001. **36**(8): p. 849-865.
110. Boyd, R.K., C. Basic, and R.A. Bethem, *Trace quantitative analysis by mass spectrometry*. 2011: Wiley.
111. Yocum, A.K. and A.M. Chinnaiyan, *Current affairs in quantitative targeted proteomics: multiple reaction monitoring–mass spectrometry.* Briefings in Functional Genomics & Proteomics, 2009. **8**(2): p. 145-157.
112. Service, R.F., *Proteomics. High-speed biologists search for gold in proteins.* Science, 2001. **294**(5549): p. 2074-7.
113. Gygi, S.P., et al., *Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.* Proc Natl Acad Sci U S A, 2000. **97**(17): p. 9390-5.
114. Lee, H., et al., *Development of a multiplexed microcapillary liquid chromatography system for high-throughput proteome analysis.* Anal Chem, 2002. **74**(17): p. 4353-60.
115. Issaq, H.J., Z. Xiao, and T.D. Veenstra, *Serum and Plasma Proteomics.* Chemical Reviews, 2007. **107**(8): p. 3601-3620.
116. Denny, P., et al., *The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions.* J Proteome Res, 2008. **7**(5): p. 1994-2006.
117. Sato, I., et al., *Applicability of Nanotrap Sg as a semen detection kit before male-specific DNA profiling in sexual assaults.* Int J Legal Med, 2007. **121**(4): p. 315-9.
118. Dasari, S., et al., *Comprehensive proteomic analysis of human cervical-vaginal fluid.* J Proteome Res, 2007. **6**(4): p. 1258-68.
119. Irving, J.A., et al., *Serpins in prokaryotes.* Mol Biol Evol, 2002. **19**(11): p. 1881-90.
120. Zimmer, D.B., et al., *The S100 protein family: history, function, and expression.* Brain Res Bull, 1995. **37**(4): p. 417-29.
121. Kirchhoff, C., C. Osterhoff, and L. Young, *Molecular cloning and characterization of HE1, a major secretory protein of the human epididymis.* Biol Reprod, 1996. **54**(4): p. 847-56.
122. Vanier, M.T. and G. Millat, *Structure and function of the NPC2 protein.* Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, 2004. **1685**(1–3): p. 14-21.

123. Infante, R.E., et al., *NPC2 facilitates bidirectional transfer of cholesterol between NPC1 and lipid bilayers, a step in cholesterol egress from lysosomes.* Proc Natl Acad Sci U S A, 2008. **105**(40): p. 15287-92.
124. Kise, H., et al., *Characterization of semenogelin II and its molecular interaction with prostate-specific antigen and protein C inhibitor.* Eur J Biochem, 1996. **238**(1): p. 88-96.
125. Malm, J., et al., *Isolation and characterization of the major gel proteins in human semen, semenogelin I and semenogelin II.* Eur J Biochem, 1996. **238**(1): p. 48-53.
126. Vitorino, R., et al., *Identification of human whole saliva protein components using proteomics.* PROTEOMICS, 2004. **4**(4): p. 1109-1115.
127. Hassan, M.I., et al., *Prolactin inducible protein in cancer, fertility and immunoregulation: structure, function and its clinical implications.* Cellular and Molecular Life Sciences, 2009. **66**(3): p. 447-459.
128. Ward, A.M., J.W. Catto, and F.C. Hamdy, *Prostate specific antigen: biology, biochemistry and available commercial assays.* Ann Clin Biochem, 2001. **38**(Pt 6): p. 633-51.
129. Suzuki, K., et al., *The interaction among protein C inhibitor, prostate-specific antigen, and the semenogelin system.* Semin Thromb Hemost, 2007. **33**(1): p. 46-52.
130. Thompson, I.M., et al., *Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter.* N Engl J Med, 2004. **350**(22): p. 2239-46.
131. Basch, E., et al., *Screening for prostate cancer with prostate-specific antigen testing: American Society of Clinical Oncology Provisional Clinical Opinion.* J Clin Oncol, 2012. **30**(24): p. 3020-5.
132. Keshishian, H., et al., *Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution.* Mol Cell Proteomics, 2007. **6**(12): p. 2212-29.
133. Yam, L.T., *Clinical significance of the human acid phosphatases: A review.* The American journal of medicine, 1974. **56**(5): p. 604-616.
134. Kontturi, M., *Is acid phosphatase (PAP) still justified in the management of prostatic cancer?* Acta Oncol, 1991. **30**(2): p. 169-70.
135. Taira, A., et al., *Reviving the acid phosphatase test for prostate cancer.* Oncology (Williston Park), 2007. **21**(8): p. 1003-10.
136. Fung, K.Y., et al., *A comprehensive characterization of the peptide and protein constituents of human seminal fluid.* Prostate, 2004. **61**(2): p. 171-81.
137. Duncan, M.W. and H.S. Thompson, *Proteomics of semen and its constituents.* Proteomics Clin Appl, 2007. **1**(8): p. 861-75.
138. Al-Hashimi, I., D.P. Dickinson, and M.J. Levine, *Purification, molecular cloning, and sequencing of salivary cystatin SA-1.* Journal of Biological Chemistry, 1988. **263**(19): p. 9381-9387.
139. Freije, J.P., et al., *Structure and expression of the gene encoding cystatin D, a novel human cysteine proteinase inhibitor.* Journal of Biological Chemistry, 1991. **266**(30): p. 20538-43.

140. Tamaki, N., et al., *Comparison of inhibitory activity on calcium phosphate precipitation by acidic proline-rich proteins, statherin, and histatin-1.* Calcif Tissue Int, 2002. **71**(1): p. 59-62.
141. Cabras, T., et al., *Tyrosine polysulfation of human salivary histatin 1. A post-translational modification specific of the submandibular gland.* J Proteome Res, 2007. **6**(7): p. 2472-80.
142. Schlesinger, D.H. and D.I. Hay, *Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.* Int J Pept Protein Res, 1986. **27**(4): p. 373-9.
143. Hay, D.I., et al., *The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).* Biochem J, 1988. **255**(1): p. 15-21.
144. Schlesinger, D.H., D.I. Hay, and M.J. Levine, *Complete primary structure of statherin, a potent inhibitor of calcium phosphate precipitation, from the saliva of the monkey, Macaca arctoides.* Int J Pept Protein Res, 1989. **34**(5): p. 374-80.
145. Hay, D.I., et al., *Relationship between concentration of human salivary statherin and inhibition of calcium phosphate precipitation in stimulated human parotid saliva.* J Dent Res, 1984. **63**(6): p. 857-63.
146. Amado, F.M.L., et al., *Analysis of the human saliva proteome.* 2005.
147. Hu, S., J.A. Loo, and D.T. Wong, *Human saliva proteome analysis.* Ann N Y Acad Sci, 2007. **1098**: p. 323-9.
148. Murphy, G.F., et al., *Involucrin expression in normal and neoplastic human skin: a marker for keratinocyte differentiation.* J Invest Dermatol, 1984. **82**(5): p. 453-7.
149. Contzler, R., et al., *Cornulin, a new member of the "fused gene" family, is expressed during epidermal differentiation.* J Invest Dermatol, 2005. **124**(5): p. 990-7.
150. Arnouk, H., et al., *Characterization of Molecular Markers Indicative of Cervical Cancer Progression.* Proteomics Clin Appl, 2009. **3**(5): p. 516-527.
151. Harada, N., et al., *Human IgGFc binding protein (FcgammaBP) in colonic epithelial cells exhibits mucin-like structure.* J Biol Chem, 1997. **272**(24): p. 15232-41.
152. Smith, B.A., et al., *Identification of genes involved in human urothelial cell-matrix interactions: implications for the progression pathways of malignant urothelium.* Cancer Res, 2001. **61**(4): p. 1678-85.
153. Birkedal-Hansen, H., et al., *Matrix metalloproteinases: a review.* Crit Rev Oral Biol Med, 1993. **4**(2): p. 197-250.
154. Nikkola, J., et al., *High serum levels of matrix metalloproteinase-9 and matrix metalloproteinase-1 are associated with rapid progression in patients with metastatic melanoma.* Clin Cancer Res, 2005. **11**(14): p. 5158-66.
155. Wu, C.Y., et al., *Plasma matrix metalloproteinase-9 level is better than serum matrix metalloproteinase-9 level to predict gastric cancer evolution.* Clin Cancer Res, 2007. **13**(7): p. 2054-60.
156. Tabak, L.A., et al., *Role of salivary mucins in the protection of the oral cavity.* Journal of Oral Pathology & Medicine, 1982. **11**(1): p. 1-17.

157. Chen, Y., et al., *Characterization of human mucin 5B gene expression in airway epithelium and the genomic clone of the amino-terminal and 5'-flanking region.* Am J Respir Cell Mol Biol, 2001. **25**(5): p. 542-53.
158. Gipson, I.K., et al., *Mucin genes expressed by human female reproductive tract epithelia.* Biol Reprod, 1997. **56**(4): p. 999-1011.
159. Audie, J.P., et al., *Mucin gene expression in the human endocervix.* Hum Reprod, 1995. **10**(1): p. 98-102.
160. Goetz, D.H., et al., *The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition.* Mol Cell, 2002. **10**(5): p. 1033-43.
161. Cowland, J.B. and N. Borregaard, *Molecular characterization and pattern of tissue expression of the gene for neutrophil gelatinase-associated lipocalin from humans.* Genomics, 1997. **45**(1): p. 17-23.
162. Ruhrberg, C., et al., *Periplakin, a novel component of cornified envelopes and desmosomes that belongs to the plakin family and forms complexes with envoplakin.* J Cell Biol, 1997. **139**(7): p. 1835-49.
163. Park, G.T., et al., *Suprabasin, a novel epidermal differentiation marker and potential cornified envelope precursor.* J Biol Chem, 2002. **277**(47): p. 45195-202.
164. Pagan, R., et al., *Vimentin filaments follow the preexisting cytokeratin network during epithelial-mesenchymal transition of cultured neonatal rat hepatocytes.* Exp Cell Res, 1996. **222**(2): p. 333-44.
165. Dasari, S., et al., *Comprehensive proteomic analysis of human cervical-vaginal fluid.* Journal of proteome research, 2007. **6**(4): p. 1258-1268.
166. Di Quinzio, M.K., et al., *Proteomic analysis and characterisation of human cervico-vaginal fluid proteins.* Australian and New Zealand journal of obstetrics and gynaecology, 2007. **47**(1): p. 9-15.
167. Klein, L.L., et al., *Shotgun proteomic analysis of vaginal fluid from women in late pregnancy.* Reproductive Sciences, 2008. **15**(3): p. 263-273.
168. Pereira, L., et al., *Identification of novel protein biomarkers of preterm birth in human cervical-vaginal fluid.* Journal of proteome research, 2007. **6**(4): p. 1269-1276.
169. Shaw, J.L., C.R. Smith, and E.P. Diamandis, *Proteomic analysis of human cervico-vaginal fluid.* Journal of proteome research, 2007. **6**(7): p. 2859-2865.
170. Tang, L.-J., et al., *Proteomic analysis of human cervical-vaginal fluids.* Journal of proteome research, 2007. **6**(7): p. 2874-2883.
171. Venkataraman, N., et al., *Cationic polypeptides are required for anti-HIV-1 activity of human vaginal fluid.* The Journal of Immunology, 2005. **175**(11): p. 7560-7567.
172. Zegels, G., et al., *Comprehensive proteomic analysis of human cervical-vaginal fluid using colposcopy samples.* Proteome Sci, 2009. **7**: p. 17.
173. Mazzali, M., et al., *Osteopontin—a molecule for all seasons.* QJM, 2002. **95**(1): p. 3-13.
174. Rangaswami, H., A. Bulbule, and G.C. Kundu, *Osteopontin: role in cell signaling and cancer progression.* Trends in Cell Biology, 2006. **16**(2): p. 79-87.

175. Braitch, M. and C.S. Constantinescu, *The role of osteopontin in experimental autoimmune encephalomyelitis (EAE) and multiple sclerosis (MS).* Inflamm Allergy Drug Targets, 2010. **9**(4): p. 249-56.
176. Rampoldi, L., et al., *The rediscovery of uromodulin (Tamm-Horsfall protein): from tubulointerstitial nephropathy to chronic kidney disease.* Kidney Int, 2011. **80**(4): p. 338-47.
177. Serafini-Cessi, F., A. Monti, and D. Cavallone, *N-Glycans carried by Tamm-Horsfall glycoprotein have a crucial role in the defense against urinary tract diseases.* Glycoconj J, 2005. **22**(7-9): p. 383-94.
178. Reinholt, F.P., et al., *Osteopontin--a possible anchor of osteoclasts to bone.* Proceedings of the National Academy of Sciences, 1990. **87**(12): p. 4473-4475.
179. Adachi, J., et al., *The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins.* Genome Biol, 2006. **7**(9): p. R80.
180. Niwa, T., *Biomarker discovery for kidney diseases by mass spectrometry.* J Chromatogr B Analyt Technol Biomed Life Sci, 2008. **870**(2): p. 148-53.
181. Squier, M.K.T., et al., *Calpain and calpastatin regulate neutrophil apoptosis.* Journal of Cellular Physiology, 1999. **178**(3): p. 311-319.
182. Geesink, G.H. and M. Koohmaraie, *Postmortem proteolysis and calpain/calpastatin activity in callipyge and normal lamb biceps femoris during extended postmortem storage.* J Anim Sci, 1999. **77**(6): p. 1490-501.
183. Kolarich, D., et al., *Comprehensive glyco-proteomic analysis of human α1-antitrypsin and its charge isoforms.* PROTEOMICS, 2006. **6**(11): p. 3369-3380.
184. Parfrey, H., R. Mahadeva, and D.A. Lomas, *Alpha(1)-antitrypsin deficiency, liver disease and emphysema.* Int J Biochem Cell Biol, 2003. **35**(7): p. 1009-14.
185. Mastellos, D., et al., *Complement: structure, functions, evolution, and viral molecular mimicry.* Immunol Res, 2003. **27**(2-3): p. 367-86.
186. Sahu, A. and J.D. Lambris, *Structure and biology of complement protein C3, a connecting link between innate and acquired immunity.* Immunol Rev, 2001. **180**: p. 35-48.
187. Berg JM, T.J., Stryer L., *Biochemistry*, in *Hemoglobin Transports Oxygen Efficiently by Binding Oxygen Cooperatively*. 2002, W H Freeman: New York.
188. Tolosano, E. and F. Altruda, *Hemopexin: structure, function, and regulation.* DNA and cell biology, 2002. **21**(4): p. 297-306.
189. Ascenzi, P., et al., *Hemoglobin and heme scavenging.* IUBMB Life, 2005. **57**(11): p. 749-759.
190. Deutsch, E.W., et al., *Human Plasma PeptideAtlas.* Proteomics, 2005. **5**(13): p. 3497-500.
191. Wu, C.C. and M.J. MacCoss, *Shotgun proteomics: tools for the analysis of complex biological systems.* Curr Opin Mol Ther, 2002. **4**(3): p. 242-50.
192. Ellington, A.A., et al., *Antibody-based protein multiplex platforms: technical and operational challenges.* Clin Chem, 2010. **56**(2): p. 186-93.
193. Ling, M.M., C. Ricks, and P. Lea, *Multiplexing molecular diagnostics and immunoassays using emerging microarray technologies.* Expert Rev Mol Diagn, 2007. **7**(1): p. 87-98.

194. Makawita, S. and E.P. Diamandis, *The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry-based approaches: current strategies for candidate verification.* Clin Chem, 2010. **56**(2): p. 212-22.
195. Whiteaker, J.R., et al., *A targeted proteomics-based pipeline for verification of biomarkers in plasma.* Nat Biotechnol, 2011. **29**(7): p. 625-34.
196. Addona, T.A., et al., *A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease.* Nat Biotechnol, 2011. **29**(7): p. 635-43.
197. Laffan, A., et al., *Evaluation of semen presumptive tests for use at crime scenes.* Med Sci Law, 2011. **51**(1): p. 11-7.
198. Ens, W. and K.G. Standing, *Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules.* Methods Enzymol, 2005. **402**: p. 49-78.
199. Wang, N. and L. Li, *Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis.* Anal Chem, 2008. **80**(12): p. 4696-710.
200. Huang, C.-M., *Comparative proteomic analysis of human whole saliva.* Archives of Oral Biology, 2004. **49**(12): p. 951-962.
201. Ghafouri, B., C. Tagesson, and M. Lindahl, *Mapping of proteins in human saliva using two-dimensional gel electrophoresis and peptide mass fingerprinting.* Proteomics, 2003. **3**(6): p. 1003-1015.
202. Lieden, A., et al., *Cornulin, a marker of late epidermal differentiation, is down-regulated in eczema.* Allergy, 2009. **64**(2): p. 304-11.
203. Hu, S., et al., *Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry.* Proteomics, 2005. **5**(6): p. 1714-28.
204. Wilmarth, P.A., et al., *Two-dimensional liquid chromatography study of the human whole saliva proteome.* J Proteome Res, 2004. **3**(5): p. 1017-23.
205. Xie, H., et al., *A catalogue of human saliva proteins identified by free flow electrophoresis-based peptide separation and tandem mass spectrometry.* Mol Cell Proteomics, 2005. **4**(11): p. 1826-30.
206. Hu, S., J.A. Loo, and D.T. Wong, *Human saliva proteome analysis and disease biomarker discovery.* Expert Rev Proteomics, 2007. **4**(4): p. 531-8.
207. Amado, F., et al., *Salivary peptidomics.* Expert Rev Proteomics, 2010. **7**(5): p. 709-21.
208. Sun, W., et al., *Human urine proteome analysis by three separation approaches.* Proteomics, 2005. **5**(18): p. 4994-5001.
209. Hu, S., et al., *Human Saliva Proteome and Transcriptome.* Journal of Dental Research, 2006. **85**(12): p. 1129-33.
210. Hansen, L.V., et al., *Altered expression of the urokinase receptor homologue, C4.4A, in invasive areas of human esophageal squamous cell carcinoma.* Int J Cancer, 2008. **122**(4): p. 734-41.
211. Hoffman, E., & Stroobant, V. , *Mass spectrometry principals and applications*. 2007: Wiley.

212. Chen, K., et al., *Characterization of Tumor Suppressive Function of cornulin in Esophageal Squamous Cell Carcinoma.* PLoS One, 2013. **8**(7): p. e68838.
213. Imai, F.L., et al., *Chromosome 1 open reading frame 10 (C1orf10) gene is frequently down-regulated and inhibits cell proliferation in oral squamous cell carcinoma.* Int J Biochem Cell Biol, 2005. **37**(8): p. 1641-55.
214. Lundwall, A., et al., *Semenogelin I and II, the predominant human seminal plasma proteins, are also expressed in non-genital tissues.* Mol Hum Reprod, 2002. **8**(9): p. 805-10.

## Appendix A: ProteoSep™ 2D Protein Identification Results

### Seminal Fluid 2D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 18 | 303.34 | 37.2 | 2.83E+10 | P04279 | Semenogelin-1 precursor |
| 15 | 230.45 | 27.3 | 8.80E+09 | Q02383 | Semenogelin-2 precursor |
| 6 | 109.76 | 10.5 | 2.41E+08 | P04264 | Keratin, type II cytoskeletal 1 |
| 4 | 57.15 | 6.2 | 1.18E+08 | P35908 | Keratin, type II cytoskeletal 2 epidermal |
| 6 | 95.35 | 10.2 | 1.60E+08 | P13645 | Keratin, type I cytoskeletal 10 |
| 1 | 14.42 | 1.9 | 1.14E+07 | P02533 | Keratin, type I cytoskeletal 14 |
| 1 | 14.42 | 1.9 | 1.14E+07 | P08779 | Keratin, type I cytoskeletal 16 |
| 6 | 90.52 | 9.2 | 7.12E+07 | P02788 | Lactotransferrin |
| 6 | 88.65 | 12.9 | 7.10E+07 | P10909 | Clusterin |
| 3 | 46.79 | 7.2 | 4.23E+07 | P15309 | Prostatic acid phosphatase |
| 3 | 45.76 | 5.2 | 3.86E+07 | P02768 | Serum albumin |
| 3 | 45.36 | 49.1 | 3.96E+07 | P06702 | Protein S100-A9 |
| 2 | 34.78 | 10.7 | 2.08E+07 | P07288 | Prostate-specific antigen |

## Urine 2D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession # | Protein Name |
|---|---|---|---|---|---|
| 32 | 561.94 | 63.7 | 2.79E+09 | P02768 | Serum albumin |
| 10 | 155.15 | 39.6 | 1.21E+08 | P25311 | Zinc-alpha-2-glycoprotein |
| 8 | 153.58 | 2.4 | 6.00E+08 | P98160 | Basement membrane-specific heparan sulfate proteoglycan protein |
| 6 | 111.72 | 16.7 | 1.23E+09 | P02760 | Protein AMBP |
| 6 | 104.17 | 9.8 | 1.83E+08 | Q14624 | Inter-alpha-trypsin inhibitor heavy chain H4 |
| 5 | 77.59 | 8.3 | 1.53E+08 | P01042 | Kininogen-1 |
| 4 | 59.89 | 10.6 | 1.46E+08 | P07602 | Proactivator polypeptide |
| 4 | 58.86 | 35.5 | 4.55E+07 | P05451 | Lithostathine-1-alpha |
| 2 | 32.74 | 13.8 | 3.40E+07 | P48304 | Lithostathine-1-beta |
| 3 | 57.26 | 11.7 | 1.23E+08 | P10451 | Osteopontin |
| 4 | 54.88 | 25.1 | 1.21E+10 | P00761 | Trypsin |
| 3 | 53.21 | 5.3 | 5.09E+07 | P55290 | Cadherin-13 |
| 3 | 49.91 | 6.2 | 2.55E+07 | P13645 | Keratin, type I cytoskeletal 10 |
| 3 | 48.83 | 5.4 | 4.96E+07 | P04264 | Keratin, type II cytoskeletal 1 |
| 1 | 14.86 | 2 | 3.31E+07 | P19013 | Keratin, type II cytoskeletal 4 |
| 1 | 14.86 | 2.2 | 3.31E+07 | P05787 | Keratin, type II cytoskeletal 8 |
| 1 | 14.86 | 2.3 | 3.31E+07 | P08729 | Keratin, type II cytoskeletal 7 |
| 1 | 14.86 | 2.4 | 3.31E+07 | Q6KB66 | Keratin, type II cytoskeletal 80 |
| 1 | 14.86 | 2.5 | 3.31E+07 | P14136 | Glial fibrillary acidic protein |
| 3 | 48.68 | 14.1 | 7.39E+07 | P06870 | Kallikrein-1 |
| 4 | 47.36 | 66 | 3.05E+07 | P01834 | Ig kappa chain C region |
| 2 | 45.33 | 25.8 | 3.31E+07 | Q14508 | WAP four-disulfide core domain protein 2 |
| 3 | 44.7 | 10.9 | 5.89E+07 | P28799 | Granulins |
| 2 | 36.06 | 5.1 | 2.14E+07 | P01833 | Polymeric immunoglobulin receptor |
| 2 | 34.04 | 3.3 | 7.05E+07 | P01133 | Pro-epidermal growth factor |
| 2 | 31.1 | 17.3 | 1.28E+07 | P41222 | Prostaglandin-H2 D-isomerase |
| 2 | 29.7 | 16.4 | 4.09E+07 | P02753 | Retinol-binding protein 4 |
| 2 | 28.93 | 23.8 | 3.39E+06 | P02766 | Transthyretin |
| 2 | 28.52 | 18.7 | 9.68E+07 | P13987 | CD59 glycoprotein |
| 2 | 25.8 | 7 | 2.95E+07 | P01876 | Ig alpha-1 chain C region |
| 1 | 13.92 | 2.9 | 1.91E+07 | P01877 | Ig alpha-2 chain C region |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 21.07 | 4.7 | 1.81E+07 | P04156 | Major prion protein |
| 1 | 21.01 | 2.3 | 2.26E+07 | P05060 | Secretogranin-1 O |
| 1 | 19.38 | 5.9 | 1.73E+06 | P05452 | Tetranectin |
| 1 | 17.8 | 2.9 | 1.06E+07 | Q9ULV1 | Frizzled-4 |
| 1 | 15.91 | 1.6 | 9.02E+06 | O75339 | Cartilage intermediate layer protein 1 |
| 1 | 15.23 | 6.2 | 4.84E+06 | Q7Z3B1 | Neuronal growth regulator 1 |
| 1 | 14.51 | 10.8 | 1.71E+07 | P31151 | Protein S100-A7 |
| 1 | 14.51 | 10.8 | 1.71E+07 | Q86SG5 | Protein S100-A7A |
| 1 | 14.4 | 9.5 | 8.04E+06 | P09564 | T-cell antigen CD7 |
| 1 | 14.34 | 5.4 | 4.05E+07 | Q9UNN8 | Endothelial protein C receptor |
| 1 | 14.03 | 5.1 | 1.94E+07 | Q9BT78 | COP9 signalosome complex subunit |
| 1 | 13.31 | 2.1 | 7.41E+06 | Q14574 | Desmocollin-3 |
| 1 | 13.26 | 3.3 | 1.34E+07 | Q6NXE6 | Armadillo repeat-containing protein 6 |

## Saliva 2D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 4 | 62.41 | 48.3 | 6.71E+07 | P02808 | Statherin precursor - Homo sapiens (Human) |
| 2 | 35.56 | 3.7 | 8.29E+06 | P04264 | Keratin, type II cytoskeletal 1 |
| 1 | 17.04 | 1.8 | 8.25E+06 | P35908 | Keratin, type II cytoskeletal 2 epiderma |
| 2 | 34.81 | 5.2 | 6.57E+06 | P13645 | Keratin, type I cytoskeletal 10 |
| 1 | 21.17 | 8.1 | 1.40E+06 | P61626 | Lysozyme C precursor |
| 1 | 17.68 | 10.2 | 3.13E+05 | P02810 | Salivary acidic proline-rich phosphoprotein 1/2 |
| 1 | 17.16 | 3.5 | 4.80E+07 | P14314 | Glucosidase 2 subunit beta precursor |
| 1 | 16.53 | 6 | 9.94E+06 | Q96DU7 | Inositol-trisphosphate 3-kinase C |
| 1 | 15.98 | 5.8 | 4.24E+07 | Q9UM11 | Fizzy-related protein homolog - |
| 1 | 15.44 | 17.5 | 1.98E+05 | P15515 | Histatin-1 OS=Homo sapiens |
| 1 | 14.79 | 4.2 | 2.94E+05 | P24158 | Myeloblastin OS=Homo sapiens |
| 1 | 14.4 | 9.7 | 5.68E+06 | Q9UK76 | Hematological and neurological expressed 1 protein |
| 1 | 13.77 | 0.2 | 2.72E+06 | P49792 | E3 SUMO-protein ligase RanBP2 |
| 1 | 13.77 | 0.5 | 2.72E+06 | Q53T03 | RANBP2-like and GRIP domain-containing protein 6 |
| 1 | 13.77 | 0.5 | 2.72E+06 | Q99666 | RANBP2-like and GRIP domain-containing protein 5 |
| 1 | 13.77 | 0.5 | 2.72E+06 | A6NKT7 | RANBP2-like and GRIP domain-containing protein 3 |
| 1 | 13.77 | 0.5 | 2.72E+06 | Q7Z3J3 | RANBP2-like and GRIP domain-containing protein 4 |
| 1 | 13.77 | 0.5 | 2.72E+06 | Q68DN6 | RANBP2-like and GRIP domain-containing protein 1 |
| 1 | 13.77 | 0.9 | 2.72E+06 | Q9H0B2 | RANBP2-like and GRIP domain-containing protein 7 |

## Vaginal Fluid 2D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 18 | 301.33 | 37.1 | 1.87E+09 | P02768 | Serum albumin precursor |
| 15 | 250.88 | 38.7 | 5.00E+09 | P30740 | Leukocyte elastase inhibitor |
| 1 | 17.92 | 2.4 | 3.82E+07 | O75830 | Serpin I2 |
| 1 | 17.92 | 2.6 | 3.82E+07 | P50452 | Serpin B8 |
| 1 | 17.92 | 2.6 | 3.82E+07 | P50453 | Serpin B9 |
| 13 | 205.85 | 32.1 | 1.05E+09 | P07476 | Involucrin |
| 10 | 172.07 | 29.4 | 6.32E+08 | P07355 | Annexin A2 |
| 8 | 130.66 | 24.1 | 4.01E+08 | A6NMY6 | Putative annexin A2-like protein |
| 11 | 167.88 | 20.7 | 1.20E+09 | P13646 | Keratin, type I cytoskeletal 13 |
| 5 | 83.51 | 9.2 | 4.66E+08 | P19012 | Keratin, type I cytoskeletal 15 |
| 3 | 50.89 | 4.8 | 1.55E+08 | P02533 | Keratin, type I cytoskeletal 14 |
| 3 | 50.89 | 4.8 | 1.55E+08 | P08779 | Keratin, type I cytoskeletal |
| 2 | 30.68 | 2.5 | 5.90E+07 | P35900 | Keratin, type I cytoskeletal 20 |
| 2 | 30.68 | 2.5 | 5.90E+07 | Q04695 | Keratin, type I cytoskeletal 17 |
| 2 | 25.96 | 3.8 | 8.75E+07 | Q99456 | Keratin, type I cytoskeletal 12 |
| 2 | 25.77 | 2 | 1.56E+08 | P13645 | Keratin, type I cytoskeletal 10 |
| 1 | 15.23 | 2.2 | 8.94E+06 | P08727 | Keratin, type I cytoskeletal 19 |
| 1 | 15.23 | 3 | 8.94E+06 | Q8N1A0 | Keratin-like protein KRT222 |
| 1 | 14.96 | 2 | 1.14E+08 | Q2M2I5 | Keratin, type I cytoskeletal 24 |
| 10 | 153.97 | 16.9 | 4.57E+08 | P05164 | Myeloperoxidase |
| 9 | 140.8 | 31.7 | 1.20E+09 | P29508 | Serpin B3 |
| 1 | 13.29 | 3 | 2.30E+07 | Q9UIV8 | Serpin B13 |
| 8 | 116.33 | 27.1 | 9.32E+08 | P48594 | Serpin B4 |
| 8 | 131.35 | 31.2 | 3.82E+08 | P04083 | Annexin A1 |
| 7 | 114.87 | 12.8 | 2.84E+08 | P04264 | Keratin, type II cytoskeletal 1 |
| 1 | 13.22 | 2 | 2.26E+06 | P05787 | Keratin, type II cytoskeletal 8 |
| 6 | 104.98 | 13.1 | 1.06E+09 | P02538 | Keratin, type II cytoskeletal 6A - |
| 6 | 101.44 | 13.1 | 5.51E+08 | P04259 | Keratin, type II cytoskeletal 6B - |
| 5 | 85.83 | 10.4 | 5.40E+08 | P48668 | Keratin, type II cytoskeletal 6E |
| 4 | 70.18 | 7.9 | 4.14E+08 | P13647 | Keratin, type II cytoskeletal 5 |
| 7 | 111.65 | 53.5 | 2.69E+10 | P06702 | Protein S100-A9 |
| 5 | 89.22 | 11.4 | 1.08E+09 | P19013 | Keratin, type II cytoskeletal 4 |

| | | | | | |
|---|---|---|---|---|---|
| 6 | 81.1 | 5.4 | 1.47E+08 | A8K2U0 | Alpha-2-macroglobulin-like protein 1 |
| 5 | 61.68 | 23.3 | 1.23E+08 | P01857 | Ig gamma-1 chain C region |
| 2 | 22.99 | 6.3 | 3.99E+07 | P01860 | Ig gamma-3 chain C region |
| 2 | 22.42 | 9.7 | 3.00E+07 | P01861 | Ig gamma-4 chain C region |
| 3 | 52.48 | 32.6 | 1.47E+08 | P02766 | Transthyretin |
| 3 | 50.41 | 5.7 | 1.21E+08 | P02787 | Serotransferrin |
| 3 | 49.55 | 28.4 | 6.16E+08 | Q9UBC9 | Small proline-rich protein 3 |
| 3 | 48.85 | 4.4 | 7.43E+07 | P11021 | 78 kDa glucose-regulated protein precursor |
| 3 | 43.94 | 31.1 | 1.72E+08 | P05109 | Protein |
| 3 | 39.41 | 7 | 9.41E+07 | P13796 | Plastin-2 |
| 1 | 14.47 | 1.4 | 3.31E+07 | P13797 | Plastin-3 |
| 1 | 14.47 | 1.4 | 3.31E+07 | Q14651 | Plastin-1 |
| 2 | 36.42 | 16.2 | 3.50E+07 | P61626 | Lysozyme C |
| 2 | 36.19 | 11.1 | 8.94E+07 | Q9UBG3 | Cornulin |
| 2 | 31 | 24 | 4.12E+07 | Q9HD89 | Resistin |
| 2 | 28.87 | 6.6 | 1.26E+08 | P07951 | Tropomyosin beta chain |
| 2 | 28.87 | 7.6 | 1.26E+08 | P67936 | Tropomyosin alpha-4 chain |
| 1 | 14.82 | 3.5 | 7.18E+07 | P06753 | Tropomyosin alpha-3 chain |
| 1 | 14.82 | 3.5 | 7.18E+07 | P09493 | Tropomyosin alpha-1 chain |
| 2 | 28.49 | 2.9 | 5.99E+07 | P02788 | Lactotransferrin OS |
| 2 | 27.99 | 17.7 | 5.90E+07 | Q14508 | WAP four-disulfide core domain protein |
| 2 | 27.48 | 34.8 | 9.57E+07 | P35321 | Cornifin-A |
| 2 | 27.3 | 34.8 | 7.56E+07 | P22528 | Cornifin-B |
| 2 | 25.23 | 1.4 | 2.83E+07 | O60437 | Periplakin OS |
| 2 | 25.06 | 7.5 | 9.35E+07 | P08670 | Vimentin |
| 1 | 18.22 | 3.5 | 6.32E+07 | Q96LW4 | Coiled-coil domain-containing |
| 1 | 18.12 | 1.7 | 1.14E+07 | O00391 | Sulfhydryl oxidase 1 |
| 1 | 17.15 | 1.7 | 1.82E+07 | P34931 | Heat shock 70 kDa protein 1L |
| 1 | 17.15 | 1.7 | 1.82E+07 | P08107 | Heat shock 70 kDa protein 1 |
| 1 | 17.03 | 12.3 | 8.59E+07 | P10599 | Thioredoxin OS=Homo sapiens |
| 1 | 16.53 | 2 | 4.11E+07 | Q2M2Z5 | Uncharacterized protein C20orf19 |
| 1 | 16.21 | 18 | 3.95E+07 | P35326 | Small proline-rich protein 2A |
| 1 | 16.2 | 9 | 2.77E+07 | P00441 | Superoxide dismutase [Cu-Zn] |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 16.19 | 4.5 | 2.41E+07 | O15519 | CASP8 and FADD-like apoptosis regulator |
| 1 | 15.69 | 1.2 | 9.89E+06 | Q9NQ38 | Serine protease inhibitor Kazal-type 5 |
| 1 | 15.62 | 8.5 | 4.81E+06 | P04632 | Calpain small subunit 1 |
| 1 | 14.98 | 4.1 | 1.82E+07 | O95948 | One cut domain family member 2 |
| 1 | 14.57 | 3.6 | 6.85E+05 | P35527 | Keratin, type I cytoskeletal 9 |
| 1 | 13.56 | 2.2 | 3.06E+06 | Q8IV53 | DENN domain-containing protein |
| 1 | 13.54 | 1.2 | 1.24E+07 | P06731 | Carcinoembryonic antigen-related cell adhesion molecule 5 |
| 1 | 13.54 | 1.7 | 1.24E+07 | P13688 | Carcinoembryonic antigen-related cell adhesion molecule 1 |
| 1 | 13.54 | 2.5 | 1.24E+07 | P31997 | Carcinoembryonic antigen-related cell adhesion molecule 8 |
| 1 | 13.33 | 2.2 | 5.19E+07 | P42261 | Glutamate receptor 1 |
| 1 | 13.33 | 2.2 | 5.19E+07 | P42263 | Glutamate receptor 3 |
| 1 | 13.33 | 2.2 | 5.19E+07 | P48058 | Glutamate receptor 4 |
| 1 | 13.33 | 2.2 | 5.19E+07 | P42262 | Glutamate receptor 2 |

## Menstrual Blood 2D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 38 | 647.31 | 35.2 | 5.91E+09 | P01024 | Complement C3 |
| 33 | 523.43 | 35.6 | 4.88E+09 | P01023 | Alpha-2-macroglobulin |
| 4 | 59.92 | 3.5 | 6.40E+08 | P20742 | Pregnancy zone protein |
| 22 | 373.98 | 41.2 | 4.60E+09 | P02787 | Serotransferrin |
| 16 | 271.67 | 54.6 | 1.16E+10 | P02647 | Apolipoprotein A-I |
| 16 | 263.75 | 37.6 | 2.14E+09 | P02675 | Fibrinogen beta chain |
| 17 | 261.74 | 29.4 | 1.75E+09 | P00450 | Ceruloplasmin |
| 15 | 240.5 | 46.8 | 4.26E+09 | P01009 | Alpha-1-antitrypsin |
| 15 | 234.68 | 15.1 | 6.75E+08 | P0C0L5 | Complement C4-B |
| 15 | 234.68 | 15.1 | 6.75E+08 | P0C0L4 | Complement C4-A |
| 11 | 178.42 | 37.8 | 3.40E+09 | P02790 | Hemopexin |
| 1 | 20.35 | 3.4 | 1.70E+07 | P20058 | Hemopexin |
| 10 | 167.17 | 77.5 | 1.68E+10 | P68871 | Hemoglobin subunit beta |
| 8 | 133.25 | 62.5 | 8.07E+09 | P02042 | Hemoglobin subunit delta |
| 9 | 144.91 | 31.2 | 1.31E+09 | P02774 | Vitamin D-binding protein |
| 9 | 139.82 | 28.9 | 2.16E+09 | P02679 | Fibrinogen gamma chain |
| 8 | 137.61 | 30.7 | 8.06E+08 | P00738 | Haptoglobin |
| 4 | 73.02 | 14.6 | 4.86E+08 | P00739 | Haptoglobin-related protein |
| 9 | 130.79 | 11.3 | 3.68E+08 | P08603 | Complement factor H |
| 7 | 121.94 | 64 | 6.42E+09 | P69905 | Hemoglobin subunit alpha |
| 8 | 121.8 | 3.1 | 3.00E+08 | P04114 | Apolipoprotein B-100 |
| 7 | 120.93 | 24.2 | 3.08E+07 | P30740 | Leukocyte elastase inhibitor |
| 6 | 101.04 | 32.4 | 4.84E+08 | P02749 | Beta-2-glycoprotein 1 |
| 5 | 96.27 | 17.1 | 4.60E+08 | P01019 | Angiotensinogen |
| 6 | 88.06 | 13.2 | 1.74E+08 | P00751 | Complement factor B |
| 5 | 80.28 | 17.1 | 5.16E+08 | P04217 | Alpha-1B-glycoprotein |
| 5 | 80.08 | 7.8 | 3.46E+08 | P02671 | Fibrinogen alpha chain |
| 5 | 77.37 | 27.7 | 2.89E+07 | P32119 | Peroxiredoxin-2 |
| 1 | 15.62 | 5.5 | 1.07E+07 | Q06830 | Peroxiredoxin-1 |
| 4 | 74.91 | 21 | 3.39E+07 | P00915 | Carbonic anhydrase 1 OS=Homo sapiens GN=CA1 PE=1 SV=2 |
| 4 | 67.01 | 21.5 | 1.01E+09 | P02765 | Alpha-2-HS-glycoprotein |
| 4 | 64.11 | 48 | 1.16E+09 | P02652 | Apolipoprotein A-II |

| | | | | | |
|---|---|---|---|---|---|
| 4 | 63.4 | 10.5 | 3.09E+08 | P01042 | Kininogen-1 |
| 4 | 60.11 | 6.5 | 1.88E+08 | P19823 | Inter-alpha-trypsin inhibitor heavy chain H2 |
| 4 | 55.39 | 13.4 | 9.81E+07 | P01871 | Ig mu chain C region |
| 3 | 38.64 | 11.5 | 8.85E+07 | P04220 | Ig mu heavy chain disease protein |
| 3 | 50.98 | 18.9 | 7.13E+08 | P02763 | Alpha-1-acid glycoprotein 1 |
| 2 | 32.08 | 12.9 | 3.84E+08 | P19652 | Alpha-1-acid glycoprotein 2 |
| 4 | 49.62 | 66 | 2.05E+08 | P01834 | Ig kappa chain C region |
| 3 | 47.89 | 7.2 | 1.35E+08 | P00734 | Prothrombin OS=Homo sapiens |
| 3 | 47.6 | 9.6 | 1.06E+08 | P04004 | Vitronectin OS=Homo sapiens |
| 3 | 47.03 | 5.1 | 2.99E+08 | P19827 | Inter-alpha-trypsin inhibitor heavy chain H1 |
| 3 | 41.72 | 11.5 | 1.17E+08 | P10909 | Clusterin |
| 2 | 39.68 | 4.8 | 8.11E+07 | P01833 | Polymeric immunoglobulin receptor |
| 2 | 37.96 | 23.8 | 5.87E+07 | P02766 | Transthyretin |
| 2 | 37.47 | 5 | 6.49E+06 | Q13228 | Selenium-binding protein 1 |
| 2 | 37.33 | 16.8 | 4.56E+07 | P00441 | Superoxide dismutase [Cu-Zn] |
| 2 | 35.13 | 7.8 | 2.03E+08 | P01011 | Alpha-1-antichymotrypsin |
| 2 | 30.6 | 27.2 | 4.04E+08 | P02656 | Apolipoprotein C-III |
| 2 | 29.64 | 7.9 | 5.77E+07 | P01008 | Antithrombin-III |
| 2 | 29.2 | 20.5 | 4.50E+06 | Q9NZT1 | Calmodulin-like protein 5 |
| 2 | 28.98 | 17.7 | 7.92E+07 | Q14508 | WAP four-disulfide core domain protein 2 |
| 2 | 28.39 | 6.5 | 1.05E+08 | P02760 | Protein AMBP |
| 2 | 24.16 | 7.7 | 1.19E+07 | P00761 | Trypsin |
| 1 | 21.5 | 1.8 | 1.37E+06 | P35908 | Keratin, type II cytoskeletal 2 epidermal |
| 1 | 21.5 | 2 | 1.37E+06 | P13647 | Keratin, type II cytoskeletal 5 |
| 1 | 21.5 | 2.1 | 1.37E+06 | P02538 | Keratin, type II cytoskeletal 6A |
| 1 | 21.5 | 2.1 | 1.37E+06 | P04259 | Keratin, type II cytoskeletal 6B |
| 1 | 21.5 | 2.1 | 1.37E+06 | P48668 | Keratin, type II cytoskeletal 6C |
| 1 | 21.5 | 2.1 | 1.37E+06 | O95678 | Keratin, type II cytoskeletal 75 |
| 1 | 21.5 | 2.2 | 1.37E+06 | Q5XKE5 | Keratin, type II cytoskeletal 79 |
| 1 | 20.65 | 2.3 | 6.86E+07 | Q14624 | Inter-alpha-trypsin inhibitor heavy chain |
| 1 | 20.02 | 4.5 | 3.62E+06 | P13645 | Keratin, type I cytoskeletal 10 |
| 1 | 17.58 | 1.8 | 1.47E+06 | P04264 | Keratin, type II cytoskeletal 1 |
| 1 | 17.19 | 2.7 | 1.44E+07 | Q8BMT4 | Leucine-rich repeat-containing protein 33 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 16.65 | 9 | 6.14E+07 | P03973 | Antileukoproteinase |
| 1 | 16.06 | 11.8 | 1.81E+06 | Q01469 | Fatty acid-binding protein, epidermal |
| 1 | 15.88 | 1 | 9.21E+07 | O00160 | Myosin-If |
| 1 | 15.75 | 3.2 | 1.56E+07 | P06727 | Apolipoprotein A-IV |
| 1 | 15.69 | 8.5 | 4.06E+07 | P80188 | Neutrophil gelatinase-associated |
| 1 | 15.44 | 2.7 | 1.20E+06 | P07476 | Involucrin |
| 1 | 15.37 | 4.1 | 2.53E+06 | Q61468 | Mesothelin |
| 1 | 15.21 | 6.8 | 3.89E+07 | Q6PII5 | Hydroxyacylglutathione hydrolase-like protein |
| 1 | 15 | 15.7 | 2.98E+07 | P22528 | Cornifin-B |
| 1 | 15 | 15.7 | 2.98E+07 | P35321 | Cornifin-A |
| 1 | 14.83 | 4.2 | 3.25E+07 | P68133 | Actin, alpha skeletal muscle |
| 1 | 14.83 | 4.2 | 3.25E+07 | P68032 | Actin, alpha cardiac muscle 1 |
| 1 | 14.83 | 4.2 | 3.25E+07 | P62736 | Actin, aortic smooth muscle |
| 1 | 14.83 | 4.2 | 3.25E+07 | P63267 | Actin, gamma-enteric smooth muscle |
| 1 | 14.56 | 5.3 | 3.65E+07 | P01860 | Ig gamma-3 chain C region |
| 1 | 14.47 | 18.4 | 1.49E+07 | P61769 | Beta-2-microglobulin |
| 1 | 14.38 | 1.7 | 2.12E+07 | P04196 | Histidine-rich glycoprotein |
| 1 | 14.33 | 1.1 | 9.85E+07 | Q9P1Z9 | Uncharacterized protein KIAA1529 |
| 1 | 14.26 | 0.6 | 1.02E+06 | Q9BTC0 | Death-inducer obliterator 1 |
| 1 | 14.18 | 3.2 | 3.02E+06 | Q8R4E9 | DNA replication factor Cdt1 |
| 1 | 14.14 | 3.4 | 5.55E+06 | Q9NZN3 | EH domain-containing protein 3 |
| 1 | 14.09 | 0.8 | 4.00E+07 | Q9P212 | 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase |
| 1 | 13.94 | 5.8 | 2.74E+06 | P28289 | Tropomodulin-1 |
| 1 | 13.64 | 7.1 | 9.06E+07 | P25311 | Zinc-alpha-2-glycoprotein |
| 1 | 13.56 | 0.6 | 1.80E+06 | Q13535 | Serine/threonine-protein kinase ATR |
| 1 | 13.4 | 4.3 | 2.34E+06 | Q9H0R3 | Transmembrane protein 222 |
| 1 | 13.36 | 3.9 | 5.82E+06 | Q8N807 | Protein disulfide-isomerase-like protein of the testis |
| 1 | 13.26 | 1.4 | 6.21E+06 | P01031 | Complement C5 |
| 1 | 13.17 | 3.5 | 4.82E+06 | Q86UB2 | Basic immunoglobulin-like variable motif-containing protein |
| 1 | 13.13 | 1 | 1.38E+08 | Q702N8 | Xin actin-binding repeat-containing protein 1 |
| 1 | 13.11 | 3.3 | 1.07E+07 | O14492 | SH2B adapter protein 2 |

## Peripheral Blood 2D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 32 | 584.65 | 61.5 | 2.51E+10 | P02768 | Serum albumin |
| 32 | 557.01 | 40.4 | 4.46E+09 | P08603 | Complement factor H |
| 3 | 50.51 | 13.3 | 1.48E+08 | Q03591 | Complement factor H-related protein 1 |
| 19 | 344.32 | 55 | 3.66E+09 | P02774 | Vitamin D-binding protein |
| 13 | 244.61 | 33.1 | 9.66E+08 | P04196 | Histidine-rich glycoprotein |
| 15 | 231.06 | 42.8 | 2.37E+09 | P02790 | Hemopexin |
| 12 | 214.92 | 31.6 | 1.32E+09 | P00751 | Complement factor B |
| 12 | 205.55 | 5 | 4.47E+08 | P04114 | Apolipoprotein B-100 |
| 12 | 202.1 | 24.9 | 4.57E+08 | P00747 | Plasminogen |
| 12 | 200.29 | 12.3 | 7.95E+08 | P01024 | Complement C3 |
| 8 | 143.68 | 69.3 | 2.35E+09 | P02766 | Transthyretin |
| 7 | 111.8 | 10.7 | 2.07E+08 | P00450 | Ceruloplasmin |
| 7 | 106.06 | 26.7 | 7.40E+08 | P01008 | Antithrombin-III |
| 6 | 87.77 | 12 | 2.50E+08 | Q14624 | Inter-alpha-trypsin inhibitor heavy chain H4 |
| 4 | 64.15 | 32.3 | 4.36E+08 | P02753 | Retinol-binding protein 4 |
| 3 | 54.93 | 10.1 | 7.16E+07 | P43652 | Afamin OS=Homo sapiens |
| 2 | 39.75 | 2.9 | 1.11E+08 | P0C0L5 | Complement C4-B |
| 2 | 39.75 | 2.9 | 1.11E+08 | P0C0L4 | Complement C4-A |
| 1 | 18.85 | 3.5 | 1.82E+07 | P01019 | Angiotensinogen |
| 1 | 15.85 | 1.4 | 3.84E+07 | P01023 | Alpha-2-macroglobulin |
| 1 | 15.64 | 7.2 | 8.72E+06 | P02749 | Beta-2-glycoprotein 1 |
| 1 | 15.64 | 3.4 | 2.80E+07 | P08697 | Alpha-2-antiplasmin |
| 1 | 14.62 | 10.3 | 3.76E+07 | P05452 | Tetranectin |
| 1 | 14.18 | 4.2 | 1.45E+08 | Q3MIP1 | Inositol 1,4,5-triphosphate receptor-interacting protein-like 2 |
| 1 | 14.01 | 3.9 | 1.88E+08 | P40818 | Ubiquitin carboxyl-terminal hydrolase 8 |
| 1 | 13.86 | 0.9 | 3.15E+07 | P35498 | Sodium channel protein type 1 subunit alpha |
| 1 | 13.06 | 9.7 | 6.99E+06 | O14626 | Probable G-protein coupled receptor 171 |

## Appendix B: ProteoSep™ 1D Protein Identification Results

### Seminal Fluid 1D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 16 | 281.63 | 35.6 | 6.65E+06 | P02788 | Lactotransferrin |
| 11 | 196.27 | 28.5 | 1.26E+08 | P04279 | Semenogelin-1 |
| 10 | 190.82 | 22.6 | 1.35E+08 | Q02383 | Semenogelin-2 |
| 10 | 161.67 | 31.3 | 8.09E+06 | P15309 | Prostatic acid phosphatase |
| 8 | 138.07 | 6.9 | 1.11E+06 | P02751 | Fibronectin |
| 8 | 132.47 | 20.6 | 5.82E+06 | P02768 | Serum albumin |
| 7 | 111.84 | 51.3 | 4.58E+06 | P12273 | Prolactin-inducible protein |
| 6 | 108.85 | 16.4 | 3.98E+06 | P10909 | Clusterin |
| 4 | 87.07 | 36.7 | 6.20E+06 | P07288 | Prostate-specific antigen |
| 2 | 38.88 | 6.7 | 8.03E+04 | P49221 | Protein-glutamine gamma-glutamyltransferase 4 |
| 2 | 35.24 | 25.8 | 1.25E+06 | P61916 | Epididymal secretory protein E1 |
| 1 | 22.49 | 4.6 | 9.85E+05 | P16870 | Carboxypeptidase E |
| 1 | 21.67 | 5.4 | 1.06E+06 | P25311 | Zinc-alpha-2-glycoprotein |
| 1 | 15.48 | 5 | 1.34E+05 | P80303 | Nucleobindin-2 |
| 1 | 15.33 | 7.1 | 1.06E+05 | P04406 | Glyceraldehyde-3-phosphate dehydrogenase |
| 1 | 14.87 | 18.4 | 4.47E+05 | P61769 | Beta-2-microglobulin |
| 1 | 13.96 | 8 | 7.35E+04 | P54107 | Cysteine-rich secretory protein 1 |

## Urine 1D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 17 | 322.36 | 38 | 2.78E+07 | P02768 | Serum albumin |
| 7 | 140.98 | 27.2 | 1.29E+08 | P02760 | Protein AMBP |
| 5 | 93.39 | 28.7 | 4.71E+05 | P01857 | Ig gamma-1 chain C region |
| 2 | 40.77 | 11.6 | 2.73E+05 | P01861 | Ig gamma-4 chain C region |
| 1 | 21.03 | 4.2 | 1.61E+05 | P01860 | Ig gamma-3 chain C region |
| 4 | 74.5 | 11.6 | 3.35E+05 | P01833 | Polymeric immunoglobulin receptor |
| 4 | 73.21 | 60.3 | 1.57E+07 | P01834 | Ig kappa chain C region |
| 4 | 72.3 | 19.1 | 4.00E+05 | P19961 | Alpha-amylase 2B |
| 3 | 53.5 | 13.3 | 1.03E+05 | P04745 | Alpha-amylase 1 |
| 3 | 52.98 | 14 | 3.60E+05 | P04746 | Pancreatic alpha-amylase |
| 3 | 61.55 | 6.3 | 8.98E+04 | P10253 | Lysosomal alpha-glucosidase |
| 3 | 60.07 | 7.3 | 5.36E+05 | Q14624 | Inter-alpha-trypsin inhibitor heavy chain H4 |
| 3 | 58.76 | 9.1 | 9.14E+05 | P01042 | Kininogen-1 |
| 3 | 55.54 | 12.1 | 4.82E+05 | P10451 | Osteopontin |
| 3 | 55.36 | 8.9 | 1.16E+06 | Q6EMK4 | Vasorin |
| 3 | 52.4 | 11.4 | 1.32E+05 | P05155 | Plasma protease C1 inhibitor |
| 3 | 45.54 | 6.5 | 1.86E+07 | P07911 | Uromodulin |
| 2 | 42.82 | 17.3 | 7.73E+07 | P41222 | Prostaglandin-H2 D-isomerase |
| 2 | 36.83 | 0.6 | 1.34E+05 | P98160 | Basement membrane-specific heparan sulfate proteoglycan core protein |
| 2 | 35.66 | 3.1 | 1.14E+05 | P01133 | Pro-epidermal growth factor |
| 2 | 35.28 | 18 | 1.35E+05 | P24855 | Deoxyribonuclease-1 |
| 2 | 34.83 | 9.3 | 1.83E+05 | P10153 | Non-secretory ribonuclease |
| 2 | 33.29 | 22.2 | 9.56E+04 | Q7Z5L0 | Vitelline membrane outer layer protein 1 homolog |
| 2 | 31.03 | 6.5 | 1.20E+05 | P00734 | Prothrombin |
| 2 | 30.6 | 7.8 | 5.70E+04 | P01009 | Alpha-1-antitrypsin |
| 2 | 26.74 | 10.1 | 1.34E+05 | P01859 | Ig gamma-2 chain C region |
| 2 | 25.26 | 9.2 | 8.00E+04 | Q9UNN8 | Endothelial protein C receptor |
| 1 | 21.72 | 5.3 | 6.21E+04 | Q12907 | Vesicular integral-membrane protein VIP36 |
| 1 | 19.54 | 33.8 | 1.35E+05 | P60022 | Beta-defensin 1 |
| 1 | 19.14 | 2.3 | 1.40E+05 | P14543 | Nidogen-1 |
| 1 | 18.11 | 9.5 | 4.57E+04 | P09564 | T-cell antigen CD7 |
| 1 | 17.52 | 7.6 | 1.90E+05 | Q8WVN6 | Secreted and transmembrane protein 1 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 17.21 | 3 | 2.76E+04 | P02671 | Fibrinogen alpha chain |
| 1 | 16.82 | 6.3 | 1.81E+05 | Q16769 | Glutaminyl-peptide cyclotransferase |
| 1 | 16.52 | 3.8 | 6.66E+04 | P15309 | Prostatic acid phosphatase |
| 1 | 16.28 | 7.6 | 2.03E+04 | O75594 | Peptidoglycan recognition protein |
| 1 | 16.27 | 4.6 | 1.71E+05 | P02774 | Vitamin D-binding protein |
| 1 | 16.11 | 24.2 | 1.68E+05 | P55000 | Secreted Ly-6/uPAR-related protein 1 |
| 1 | 15.73 | 18.4 | 4.54E+05 | P61769 | Beta-2-microglobulin |
| 1 | 15.1 | 15.7 | 3.48E+04 | P01593 | Ig kappa chain V-I region AG |
| 1 | 15.1 | 15.7 | 3.48E+04 | P01594 | Ig kappa chain V-I region AU |
| 1 | 15.1 | 15.7 | 3.48E+04 | P01607 | Ig kappa chain V-I region Rei |
| 1 | 15.1 | 15.7 | 3.48E+04 | P01610 | Ig kappa chain V-I region WEA |
| 1 | 15.1 | 15.7 | 3.48E+04 | P01599 | Ig kappa chain V-I region Gal |
| 1 | 15.1 | 15.7 | 3.48E+04 | P01608 | Ig kappa chain V-I region Roy |
| 1 | 15.1 | 15.7 | 3.48E+04 | P01609 | Ig kappa chain V-I region Scw |
| 1 | 15.1 | 15.7 | 3.48E+04 | P80362 | Ig kappa chain V-I region WAT |
| 1 | 15.1 | 15.7 | 3.48E+04 | P01600 | Ig kappa chain V-I region Hau |
| 1 | 14.56 | 15.5 | 2.53E+04 | P04206 | Ig kappa chain V-III region GOL |
| 1 | 14.56 | 15.5 | 2.53E+04 | P01622 | Ig kappa chain V-III region Ti |
| 1 | 14.56 | 15.5 | 2.53E+04 | P01620 | Ig kappa chain V-III region SIE |
| 1 | 14.56 | 15.5 | 2.53E+04 | P01623 | Ig kappa chain V-III region WOL |
| 1 | 14.23 | 3.7 | 2.73E+04 | P36896 | Activin receptor type-1B |
| 1 | 13.25 | 21 | 2.99E+04 | P62988 | Ubiquitin |

## Saliva 1D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 16 | 277.08 | 35.3 | 1.89E+09 | P02768 | Serum albumin |
| 15 | 246.71 | 40.7 | 3.29E+09 | P04745 | Alpha-amylase 1 |
| 11 | 173.74 | 31.7 | 1.96E+09 | P04746 | Pancreatic alpha-amylase |
| 13 | 207.25 | 40.5 | 2.42E+09 | P19961 | Alpha-amylase 2B |
| 11 | 180.57 | 21.7 | 9.85E+08 | P01833 | Polymeric immunoglobulin receptor |
| 11 | 166.48 | 19.7 | 3.50E+08 | P02787 | Serotransferrin |
| 8 | 147.41 | 63.1 | 1.60E+10 | P01036 | Cystatin-S |
| 2 | 31.47 | 20.5 | 5.75E+08 | P01037 | Cystatin-SN |
| 3 | 52.95 | 31.2 | 2.16E+08 | P09228 | Cystatin-SA |
| 9 | 143.58 | 43.3 | 4.77E+08 | P25311 | Zinc-alpha-2-glycoprotein |
| 1 | 13.24 | 4.9 | 6.41E+07 | A8MT79 | Putative zinc-alpha-2-glycoprotein-like 1 |
| 7 | 123.98 | 60.9 | 3.74E+09 | P12273 | Prolactin-inducible protein |
| 7 | 116.69 | 36.2 | 1.92E+09 | P01876 | Ig alpha-1 chain C region |
| 5 | 83.89 | 27.3 | 9.92E+08 | P01877 | Ig alpha-2 chain C region |
| 4 | 68.01 | 60.3 | 8.79E+08 | P01834 | Ig kappa chain C region |
| 4 | 50.31 | 11.7 | 6.90E+07 | P01871 | Ig mu chain C region |
| 3 | 39.91 | 10.4 | 6.41E+07 | P04220 | Ig mu heavy chain disease protein |
| 3 | 48.98 | 11.4 | 2.03E+08 | P68133 | Actin, alpha skeletal muscle |
| 3 | 48.98 | 11.4 | 2.03E+08 | P68032 | Actin, alpha cardiac muscle 1 |
| 2 | 34.68 | 8.4 | 1.99E+08 | P62736 | Actin, aortic smooth muscle |
| 2 | 34.68 | 8.5 | 1.99E+08 | P63267 | Actin, gamma-enteric smooth muscle |
| 2 | 33.23 | 2.5 | 1.00E+08 | A5A3E0 | POTE ankyrin domain family member F |
| 2 | 33.23 | 2.5 | 1.00E+08 | Q6S8J3 | POTE ankyrin domain family member E |
| 2 | 33.23 | 7.2 | 1.00E+08 | Q9BYX7 | Beta-actin-like protein 3 |
| 2 | 33.23 | 7.2 | 1.00E+08 | P63261 | Actin, cytoplasmic 2 |
| 2 | 33.23 | 7.2 | 1.00E+08 | P60709 | Actin, cytoplasmic 1 |
| 1 | 18.93 | 4.2 | 9.56E+07 | Q562R1 | Beta-actin-like protein 2 |
| 3 | 48.65 | 51.4 | 2.30E+08 | P01842 | Ig lambda chain C regions |
| 3 | 38.1 | 28.4 | 6.94E+07 | P01591 | Immunoglobulin J chain |
| 2 | 36.67 | 11.5 | 2.05E+07 | Q6P5S2 | Uncharacterized protein C6orf58 |
| 2 | 35.37 | 6.6 | 7.89E+07 | P06733 | Alpha-enolase |
| 2 | 33.81 | 7.8 | 6.97E+07 | P20061 | Transcobalamin-1 |

| | | | | | |
|---|---|---|---|---|---|
| 2 | 29.06 | 10.3 | 3.61E+07 | P06870 | Kallikrein-1 |
| 2 | 27.57 | 7.1 | 2.38E+07 | P02790 | Hemopexin |
| 2 | 25.62 | 7.8 | 1.40E+07 | P01857 | Ig gamma-1 chain C region |
| 1 | 18.53 | 10.7 | 1.10E+07 | P18510 | Interleukin-1 receptor antagonist protein |
| 1 | 17.65 | 10 | 4.12E+07 | Q96DA0 | Zymogen granule protein 16 homolog B |
| 1 | 17.1 | 15.7 | 3.11E+07 | P01593 | Ig kappa chain V-I region AG |
| 1 | 17.1 | 15.7 | 3.11E+07 | P01594 | Ig kappa chain V-I region AU |
| 1 | 17.1 | 15.7 | 3.11E+07 | P01607 | Ig kappa chain V-I region Rei |
| 1 | 17.1 | 15.7 | 3.11E+07 | P01610 | Ig kappa chain V-I region WEA |
| 1 | 17.1 | 15.7 | 3.11E+07 | P01599 | Ig kappa chain V-I region Gal |
| 1 | 17.1 | 15.7 | 3.11E+07 | P01608 | Ig kappa chain V-I region Roy |
| 1 | 17.1 | 15.7 | 3.11E+07 | P01609 | Ig kappa chain V-I region Scw |
| 1 | 17.1 | 15.7 | 3.11E+07 | P80362 | Ig kappa chain V-I region WAT |
| 1 | 17.1 | 15.7 | 3.11E+07 | P01600 | Ig kappa chain V-I region Hau |
| 1 | 16.67 | 13.1 | 2.26E+07 | P06702 | Protein S100-A9 |
| 1 | 16.48 | 15 | 2.84E+07 | P01766 | Ig heavy chain V-III region BRO |
| 1 | 16.48 | 15.1 | 2.84E+07 | P01777 | Ig heavy chain V-III region TEI |
| 1 | 16.4 | 12.3 | 5.09E+07 | P10599 | Thioredoxin |
| 1 | 15.76 | 1.9 | 9.35E+06 | Q02487 | Desmocollin-2 |
| 1 | 15.41 | 6.3 | 6.50E+06 | Q9Y5Z4 | Heme-binding protein 2 |
| 1 | 15.27 | 12.2 | 2.39E+07 | P04080 | Cystatin-B OS=Homo sapiens |
| 1 | 14.43 | 4.5 | 1.48E+08 | P23280 | Carbonic anhydrase 6 |
| 1 | 13.87 | 1.5 | 5.71E+06 | P02671 | Fibrinogen alpha chain |
| 1 | 13.16 | 11.8 | 8.05E+06 | P05109 | Protein S100-A8 |
| 1 | 13.09 | 15.3 | 1.57E+07 | P01776 | Ig heavy chain V-III region WAS |
| 1 | 13.09 | 15.1 | 1.57E+07 | P01774 | Ig heavy chain V-III region POM |
| 1 | 13.09 | 15.5 | 1.57E+07 | P01779 | Ig heavy chain V-III region TUR |
| 1 | 13.09 | 15.6 | 1.57E+07 | P01765 | Ig heavy chain V-III region TIL |
| 1 | 13.06 | 2.7 | 7.01E+06 | Q9BR84 | Zinc finger protein 559 |

## Vaginal Fluid 1D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 22 | 403.3 | 89.9 | 5.74E+10 | Q9UBC9 | Small proline-rich protein 3 |
| 16 | 304.09 | 39.5 | 9.17E+09 | P02768 | Serum albumin |
| 16 | 285.9 | 45.3 | 8.96E+09 | P29508 | Serpin B3 |
| 7 | 128.07 | 17.6 | 5.01E+09 | P48594 | Serpin B4 |
| 18 | 269.92 | 22 | 4.13E+09 | A8K2U0 | Alpha-2-macroglobulin-like protein 1 |
| 15 | 245.67 | 4.6 | 2.76E+09 | Q09666 | Neuroblast differentiation-associated protein AHNAK |
| 13 | 244.73 | 33.1 | 5.31E+09 | P07476 | Involucrin OS=Homo sapiens |
| 13 | 239.57 | 30.3 | 8.02E+09 | P19013 | Keratin, type II cytoskeletal 4 |
| 1 | 21.92 | 2.2 | 2.71E+08 | Q86Y46 | Keratin, type II cytoskeletal 73 |
| 1 | 15.99 | 2.3 | 6.07E+08 | P08729 | Keratin, type II cytoskeletal 7 |
| 1 | 15.99 | 2.4 | 6.07E+08 | Q6KB66 | Keratin, type II cytoskeletal 80 |
| 1 | 15.99 | 2.5 | 6.07E+08 | P14136 | Glial fibrillary acidic protein |
| 12 | 212.68 | 43.3 | 2.54E+09 | P07355 | Annexin A2 |
| 8 | 138.64 | 26.8 | 1.95E+09 | A6NMY6 | Putative annexin A2-like protein |
| 10 | 184 | 48.6 | 3.76E+09 | Q9UBG3 | Cornulin |
| 9 | 157.66 | 48.4 | 8.12E+09 | P01857 | Ig gamma-1 chain C region |
| 4 | 61.96 | 19.5 | 8.33E+08 | P01861 | Ig gamma-4 chain C region |
| 4 | 59.82 | 15.3 | 1.17E+09 | P01860 | Ig gamma-3 chain C region |
| 8 | 131.52 | 50.5 | 3.14E+09 | P80188 | Neutrophil gelatinase-associated lipocalin |
| 6 | 111.71 | 23.6 | 1.87E+09 | P04083 | Annexin A1 |
| 7 | 104.24 | 14.5 | 4.00E+08 | P01833 | Polymeric immunoglobulin receptor |
| 6 | 92.53 | 52.6 | 5.02E+09 | P06702 | Protein S100-A9 |
| 5 | 88.78 | 79.2 | 5.17E+09 | P01834 | Ig kappa chain C region |
| 4 | 74.83 | 15.5 | 1.72E+09 | Q16610 | Extracellular matrix protein 1 |
| 4 | 67.62 | 44.4 | 3.24E+09 | Q01469 | Fatty acid-binding protein, epidermal |
| 1 | 16.43 | 20.7 | 5.17E+08 | A8MUU1 | Putative fatty acid-binding protein 5-like protein 3 |
| 4 | 62.36 | 66.3 | 1.96E+09 | P01040 | Cystatin-A |
| 3 | 58.81 | 51.6 | 6.70E+09 | P22528 | Cornifin-B |
| 2 | 38.13 | 34.8 | 5.23E+09 | P35321 | Cornifin-A |
| 4 | 56.17 | 18 | 5.14E+08 | P01859 | Ig gamma-2 chain C region |
| 3 | 52.1 | 39.7 | 1.56E+08 | P32320 | Cytidine deaminase |
| 3 | 50.58 | 61.1 | 8.14E+08 | P35326 | Small proline-rich protein 2A |

| | | | | | |
|---|---|---|---|---|---|
| 2 | 36.01 | 43 | 6.15E+08 | P22531 | Small proline-rich protein 2E |
| 3 | 50.53 | 61.1 | 1.04E+09 | P35325 | Small proline-rich protein 2B |
| 2 | 46.86 | 11.2 | 3.75E+09 | O95274 | Ly6/PLAUR domain-containing protein |
| 3 | 46.73 | 10 | 2.30E+08 | O60235 | Transmembrane protease, serine 11D |
| 3 | 43.3 | 29.5 | 3.28E+08 | P10599 | Thioredoxin |
| 3 | 37.77 | 51.4 | 8.17E+08 | P01842 | Ig lambda chain C regions |
| 2 | 37.38 | 16.2 | 2.01E+08 | P61626 | Lysozyme C |
| 2 | 35.26 | 23.6 | 8.17E+08 | P05109 | Protein S100-A8 |
| 2 | 32.49 | 7.8 | 3.08E+07 | P13796 | Plastin-2 |
| 2 | 31.89 | 9.6 | 6.21E+07 | P31947 | 14-3-3 protein sigma |
| 2 | 30.36 | 3.5 | 5.40E+07 | P02787 | Serotransferrin |
| 2 | 28.37 | 30.6 | 1.49E+08 | P31151 | Protein S100-A7 |
| 2 | 27.88 | 43 | 4.32E+08 | P22532 | Small proline-rich protein 2D |
| 1 | 13.36 | 25 | 1.04E+07 | Q96RM1 | Small proline-rich protein 2F |
| 2 | 24.97 | 2.7 | 2.03E+07 | P12110 | Collagen alpha-2(VI) chain |
| 1 | 21.35 | 29.2 | 1.50E+07 | Q9UGL9 | Cysteine-rich C-terminal protein 1 |
| 1 | 21.21 | 23.3 | 6.07E+08 | P00441 | Superoxide dismutase [Cu-Zn] |
| 1 | 20.94 | 3.1 | 2.34E+08 | P00751 | Complement factor B |
| 1 | 20.1 | 6.5 | 1.31E+08 | P54108 | Cysteine-rich secretory protein 3 |
| 1 | 19.71 | 3.3 | 2.17E+08 | P01009 | Alpha-1-antitrypsin |
| 1 | 19.67 | 15.2 | 5.67E+07 | P31949 | Protein S100-A11 |
| 1 | 19.35 | 4.9 | 7.22E+07 | P53634 | Dipeptidyl-peptidase 1 |
| 1 | 17.69 | 7.7 | 1.65E+07 | P09488 | Glutathione S-transferase Mu 1 |
| 1 | 17.69 | 7.7 | 1.65E+07 | P46439 | Glutathione S-transferase Mu 5 |
| 1 | 17.69 | 7.7 | 1.65E+07 | Q03013 | Glutathione S-transferase Mu 4 |
| 1 | 17.04 | 7.8 | 4.45E+08 | P04792 | Heat shock protein beta-1 |
| 1 | 17.02 | 0.6 | 1.79E+08 | Q8TD57 | Dynein heavy chain 3, axonemal |
| 1 | 16.82 | 12.2 | 8.06E+07 | P04080 | Cystatin-B |
| 1 | 16.22 | 0.9 | 2.71E+07 | O60437 | Periplakin |
| 1 | 16.07 | 22.9 | 7.02E+07 | P07108 | Acyl-CoA-binding protein |
| 1 | 16.02 | 0.4 | 1.14E+08 | P20930 | Filaggrin |
| 1 | 15.91 | 10.7 | 6.72E+07 | P18510 | Interleukin-1 receptor antagonist protein |
| 1 | 14.99 | 5.7 | 3.15E+08 | O75191 | Xylulose kinase |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 14.92 | 18.8 | 3.43E+07 | P19957 | Elafin |
| 1 | 14.77 | 1.6 | 3.67E+07 | P05164 | Myeloperoxidase |
| 1 | 14.73 | 10.3 | 9.42E+06 | P17900 | Ganglioside GM2 activator |
| 1 | 14.66 | 9.4 | 4.78E+07 | P01764 | Ig heavy chain V-III region VH26 |
| 1 | 14.64 | 7.1 | 1.61E+08 | P25311 | Zinc-alpha-2-glycoprotein |
| 1 | 14.54 | 10.5 | 2.22E+08 | Q8N4C7 | Syntaxin-19 |
| 1 | 14.51 | 3.7 | 1.70E+08 | P08238 | Heat shock protein HSP 90-beta |
| 1 | 14.37 | 1.2 | 8.55E+07 | Q9BXL7 | Caspase recruitment domain-containing protein 11 |
| 1 | 14.3 | 7 | 2.15E+08 | Q8IUQ0 | Retinaldehyde-binding protein 1-like protein 1 |
| 1 | 14.12 | 4.2 | 8.12E+07 | P01876 | Ig alpha-1 chain C region |
| 1 | 13.93 | 17.5 | 3.83E+07 | P60903 | Protein S100-A10 |
| 1 | 13.88 | 3.4 | 2.86E+08 | Q02108 | Guanylate cyclase soluble subunit alpha-3 |
| 1 | 13.79 | 1.4 | 9.01E+07 | Q8WWQ8 | Stabilin-2 |
| 1 | 13.75 | 2.4 | 2.26E+07 | P10909 | Clusterin |
| 1 | 13.73 | 4.5 | 2.88E+07 | Q8N4F0 | Bactericidal/permeability-increasing protein-like 1 |
| 1 | 13.7 | 1.3 | 8.77E+07 | Q96C24 | Synaptotagmin-like protein 4 |
| 1 | 13.66 | 3 | 1.54E+08 | Q8IYU2 | E3 ubiquitin-protein ligase HACE1 |
| 1 | 13.6 | 2.1 | 5.14E+06 | P00450 | Ceruloplasmin |
| 1 | 13.56 | 9.5 | 2.72E+07 | P09211 | Glutathione S-transferase P |
| 1 | 13.51 | 3.9 | 2.42E+07 | Q9UBX7 | Kallikrein-11 |
| 1 | 13.44 | 2.7 | 8.82E+07 | Q6AI39 | Uncharacterized protein KIAA0240 |
| 1 | 13.35 | 3.4 | 1.97E+07 | Q8IYM0 | Protein FAM186B |
| 1 | 13.33 | 6.3 | 4.96E+07 | P02649 | Apolipoprotein E |
| 1 | 13.23 | 1 | 1.78E+08 | P21333 | Filamin-A |
| 1 | 13.21 | 15 | 2.83E+07 | P01766 | Ig heavy chain V-III region BRO |
| 1 | 13.21 | 15.1 | 2.83E+07 | P01777 | Ig heavy chain V-III region TEI |
| 1 | 13.12 | 4.2 | 9.18E+07 | P38646 | Stress-70 protein, mitochondrial |
| 1 | 13.09 | 1 | 1.29E+08 | Q8N3P4 | Vacuolar protein sorting-associated protein 8 homolog |

## Menstrual Blood 1D Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 48 | 911.81 | 76.1 | 1.95E+11 | P02768 | Serum albumin |
| 39 | 665.17 | 36.3 | 6.03E+09 | P01024 | Complement C3 |
| 33 | 528.17 | 35.6 | 5.75E+09 | P01023 | Alpha-2-macroglobulin |
| 4 | 59.92 | 3.5 | 7.58E+08 | P20742 | Pregnancy zone protein |
| 26 | 467.91 | 46.7 | 8.70E+09 | P02787 | Serotransferrin |
| 19 | 310.25 | 62.5 | 1.42E+10 | P02647 | Apolipoprotein A-I |
| 16 | 266.69 | 50.4 | 5.98E+09 | P01009 | Alpha-1-antitrypsin |
| 16 | 263.75 | 37.6 | 2.14E+09 | P02675 | Fibrinogen beta chain |
| 17 | 262.95 | 29.4 | 1.96E+09 | P00450 | Ceruloplasmin |
| 13 | 247.81 | 89.7 | 1.45E+11 | P68871 | Hemoglobin subunit beta |
| 1 | 16.36 | 6.8 | 1.89E+10 | P02100 | Hemoglobin subunit epsilon |
| 1 | 16.36 | 6.8 | 1.89E+10 | P69891 | Hemoglobin subunit gamma-1 |
| 1 | 16.36 | 6.8 | 1.89E+10 | P69892 | Hemoglobin subunit gamma-2 |
| 11 | 204.24 | 80.9 | 6.30E+10 | P02042 | Hemoglobin subunit delta |
| 15 | 234.68 | 15.1 | 7.06E+08 | P0C0L5 | Complement C4-B |
| 15 | 234.68 | 15.1 | 7.06E+08 | P0C0L4 | Complement C4-A |
| 10 | 186.53 | 83.8 | 7.38E+10 | P69905 | Hemoglobin subunit alpha |
| 11 | 186.08 | 44.5 | 6.53E+09 | P01857 | Ig gamma-1 chain C region |
| 8 | 123.09 | 41.1 | 2.35E+09 | P01859 | Ig gamma-2 chain C region |
| 6 | 95.06 | 20.4 | 1.44E+09 | P01860 | Ig gamma-3 chain C region |
| 6 | 86.13 | 24.1 | 1.75E+09 | P01861 | Ig gamma-4 chain C region |
| 10 | 163.96 | 34.6 | 3.44E+09 | P02790 | Hemopexin |
| 9 | 145.45 | 31.2 | 1.53E+09 | P02774 | Vitamin D-binding protein |
| 9 | 139.82 | 28.9 | 2.53E+09 | P02679 | Fibrinogen gamma chain |
| 8 | 137.61 | 30.7 | 8.06E+08 | P00738 | Haptoglobin OS=Homo sapiens |
| 4 | 73.02 | 14.6 | 4.86E+08 | P00739 | Haptoglobin-related protein |
| 9 | 130.79 | 11.3 | 3.68E+08 | P08603 | Complement factor H |
| 8 | 121.8 | 3.1 | 3.25E+08 | P04114 | Apolipoprotein B-100 |
| 6 | 101.04 | 32.4 | 4.84E+08 | P02749 | Beta-2-glycoprotein 1 |
| 7 | 98.1 | 16.4 | 1.91E+08 | P00751 | Complement factor B |
| 5 | 96.27 | 17.1 | 5.22E+08 | P01019 | Angiotensinogen |
| 5 | 80.58 | 59 | 1.47E+09 | P02652 | Apolipoprotein A-II |

| | | | | | |
|---|---|---|---|---|---|
| 5 | 80.28 | 17.1 | 5.16E+08 | P04217 | Alpha-1B-glycoprotein |
| 4 | 73.77 | 66 | 2.94E+09 | P01834 | Ig kappa chain C region |
| 5 | 70.31 | 19 | 2.28E+08 | P01871 | Ig mu chain C region |
| 3 | 38.64 | 11.5 | 1.59E+08 | P04220 | Ig mu heavy chain disease protein |
| 4 | 67.01 | 21.5 | 1.01E+09 | P02765 | Alpha-2-HS-glycoprotein |
| 4 | 64.87 | 6.5 | 3.41E+08 | P02671 | Fibrinogen alpha chain |
| 4 | 63.4 | 10.5 | 3.09E+08 | P01042 | Kininogen-1 |
| 4 | 60.11 | 6.5 | 1.88E+08 | P19823 | Inter-alpha-trypsin inhibitor heavy chain H2 |
| 3 | 54.3 | 47.1 | 5.24E+08 | P0CG04 | Ig lambda-1 chain C regions |
| 3 | 54.3 | 47.1 | 5.24E+08 | P0CG05 | Ig lambda-2 chain C regions |
| 3 | 54.3 | 47.1 | 5.24E+08 | P0CG06 | Ig lambda-3 chain C regions |
| 2 | 33.65 | 28.3 | 3.26E+08 | P0CF74 | Ig lambda-6 chain C region |
| 1 | 15.99 | 14.1 | 1.43E+08 | A0M8Q6 | Ig lambda-7 chain C region |
| 3 | 50.98 | 18.9 | 7.13E+08 | P02763 | Alpha-1-acid glycoprotein 1 |
| 2 | 32.08 | 12.9 | 3.84E+08 | P19652 | Alpha-1-acid glycoprotein 2 |
| 3 | 49.71 | 19.2 | 2.62E+08 | P01876 | Ig alpha-1 chain C region |
| 2 | 34.03 | 15.5 | 1.77E+08 | P01877 | Ig alpha-2 chain C region |
| 3 | 47.6 | 9.6 | 1.06E+08 | P04004 | Vitronectin |
| 3 | 47.03 | 5.1 | 2.99E+08 | P19827 | Inter-alpha-trypsin inhibitor heavy chain H1 |
| 3 | 44.64 | 21 | 2.01E+08 | P00915 | Carbonic anhydrase 1 |
| 3 | 43.49 | 11.5 | 1.81E+08 | P10909 | Clusterin |
| 3 | 42.4 | 12.5 | 6.99E+07 | P01008 | Antithrombin-III |
| 2 | 39.68 | 4.8 | 1.80E+08 | P01833 | Polymeric immunoglobulin receptor |
| 2 | 35.13 | 7.8 | 2.03E+08 | P01011 | Alpha-1-antichymotrypsin |
| 2 | 31.36 | 5.6 | 1.30E+08 | P00734 | Prothrombin |
| 2 | 30.6 | 27.2 | 4.04E+08 | P02656 | Apolipoprotein C-III |
| 2 | 30 | 19.5 | 7.72E+07 | Q9UBC9 | Small proline-rich protein 3 |
| 2 | 28.62 | 16.1 | 8.17E+07 | P80188 | Neutrophil gelatinase-associated lipocalin |
| 2 | 28.39 | 6.5 | 1.05E+08 | P02760 | Protein AMBP |
| 2 | 25.2 | 19.1 | 6.42E+07 | P59666 | Neutrophil defensin 3 |
| 2 | 25.2 | 19.1 | 6.42E+07 | P59665 | Neutrophil defensin 1 |
| 2 | 24 | 1.9 | 2.58E+08 | Q7Z442 | Polycystic kidney disease protein 1-like 2 |
| 1 | 20.65 | 2.3 | 6.86E+07 | Q14624 | Inter-alpha-trypsin inhibitor heavy chain H4 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 19.59 | 14.9 | 4.97E+07 | P02766 | Transthyretin O |
| 1 | 18.06 | 0.9 | 2.62E+07 | Q86VH2 | Kinesin-like protein KIF27 |
| 1 | 16.65 | 9 | 6.14E+07 | P03973 | Antileukoproteinase |
| 1 | 15.75 | 3.2 | 1.56E+07 | P06727 | Apolipoprotein A-IV |
| 1 | 15 | 15.7 | 2.98E+07 | P22528 | Cornifin-B |
| 1 | 15 | 15.7 | 2.98E+07 | P35321 | Cornifin-A |
| 1 | 14.83 | 4.2 | 3.25E+07 | P68133 | Actin, alpha skeletal muscle |
| 1 | 14.83 | 4.2 | 3.25E+07 | P68032 | Actin, alpha cardiac muscle 1 |
| 1 | 14.83 | 4.2 | 3.25E+07 | P62736 | Actin, aortic smooth muscle |
| 1 | 14.83 | 4.2 | 3.25E+07 | P63267 | Actin, gamma-enteric smooth muscle |
| 1 | 14.47 | 18.4 | 1.49E+07 | P61769 | Beta-2-microglobulin |
| 1 | 14.38 | 1.7 | 2.12E+07 | P04196 | Histidine-rich glycoprotein |
| 1 | 13.64 | 7.1 | 9.06E+07 | P25311 | Zinc-alpha-2-glycoprotein |
| 1 | 13.49 | 3.3 | 2.92E+08 | Q8NB16 | Mixed lineage kinase domain-like protein |
| 1 | 13.26 | 1.4 | 6.21E+06 | P01031 | Complement C5 |
| 1 | 13.06 | 6.5 | 1.82E+08 | P14384 | Carboxypeptidase M |

## Appendix C: Unfractionated Protein Identification Results

### Seminal Fluid Unfractionated Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 28 | 541.87 | 54.2 | 7.25E+07 | P02788 | Lactotransferrin |
| 28 | 496.3 | 57.1 | 2.23E+08 | P02768 | Serum albumin |
| 26 | 467.91 | 57.7 | 1.09E+09 | P04279 | Semenogelin-1 |
| 22 | 404.47 | 42.2 | 7.34E+08 | Q02383 | Semenogelin-2 |
| 9 | 170.5 | 29.5 | 4.74E+07 | P15309 | Prostatic acid phosphatase |
| 10 | 168.46 | 28.7 | 3.01E+07 | P10909 | Clusterin |
| 9 | 160.45 | 7.3 | 1.00E+07 | P02751 | Fibronectin |
| 9 | 159.21 | 57.4 | 5.73E+07 | P07288 | Prostate-specific antigen |
| 9 | 151.06 | 63 | 1.07E+08 | P12273 | Prolactin-inducible protein |
| 8 | 131.22 | 7.5 | 1.07E+07 | Q6W4X9 | Mucin-6 |
| 4 | 68.39 | 21.4 | 1.18E+07 | P25311 | Zinc-alpha-2-glycoprotein |
| 1 | 14.26 | 4.9 | 2.66E+05 | A8MT79 | Putative zinc-alpha-2-glycoprotein-like 1 |
| 3 | 62.69 | 32.4 | 4.36E+07 | P61916 | Epididymal secretory protein E1 |
| 3 | 54.64 | 7.3 | 4.71E+06 | P02787 | Serotransferrin |
| 3 | 47.14 | 25.3 | 1.02E+06 | P01034 | Cystatin-C |
| 3 | 39.52 | 8.3 | 9.22E+05 | Q16610 | Extracellular matrix protein 1 |
| 2 | 35.97 | 18.8 | 2.74E+06 | P09466 | Glycodelin |
| 2 | 35.89 | 9.1 | 1.48E+06 | P07602 | Prosaposin |
| 2 | 28.62 | 38.7 | 1.99E+05 | Q14508 | WAP four-disulfide core domain protein 2 |
| 2 | 26.28 | 19.1 | 1.06E+05 | P49223 | Kunitz-type protease inhibitor 3 |
| 1 | 23.79 | 8 | 7.42E+06 | P54107 | Cysteine-rich secretory protein 1 |
| 1 | 20.93 | 4.6 | 1.80E+06 | P16870 | Carboxypeptidase E |
| 1 | 17.34 | 18.4 | 9.10E+06 | P61769 | Beta-2-microglobulin |
| 1 | 15.99 | 2.2 | 3.71E+04 | Q08380 | Galectin-3-binding protein |
| 1 | 15.08 | 4 | 5.70E+05 | P01009 | Alpha-1-antitrypsin |
| 1 | 14.62 | 15.2 | 1.97E+06 | P41222 | Prostaglandin-H2 D-isomerase |
| 1 | 14.22 | 14.7 | 8.09E+05 | P07998 | Ribonuclease pancreatic |
| 1 | 14.03 | 14.7 | 1.91E+04 | P49913 | Cathelicidin antimicrobial peptide |
| 1 | 13.38 | 0.3 | 4.88E+04 | Q8NI35 | InaD-like protein |

## Urine Unfractionated Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 33 | 609.38 | 62.5 | 1.48E+08 | P02768 | Serum albumin |
| 13 | 267.7 | 49.1 | 6.65E+07 | P02760 | Protein AMBP |
| 13 | 200.77 | 22.6 | 7.21E+06 | P04264 | Keratin, type II cytoskeletal 1 |
| 2 | 34.77 | 3.2 | 2.07E+06 | Q7Z794 | Keratin, type II cytoskeletal 1b |
| 5 | 72.43 | 8.5 | 1.40E+06 | P04259 | Keratin, type II cytoskeletal 6B |
| 12 | 186.83 | 24.6 | 9.41E+06 | P07911 | Uromodulin |
| 9 | 162.52 | 44.8 | 5.12E+06 | P01857 | Ig gamma-1 chain C region |
| 3 | 55.76 | 14.6 | 3.30E+06 | P01861 | Ig gamma-4 chain C region |
| 3 | 48.19 | 9 | 2.63E+06 | P01860 | Ig gamma-3 chain C region |
| 4 | 57.21 | 15.6 | 2.86E+06 | P01859 | Ig gamma-2 chain C region |
| 9 | 154.59 | 34.2 | 2.03E+06 | P04745 | Alpha-amylase 1 |
| 8 | 137.56 | 29.5 | 2.14E+06 | P04746 | Pancreatic alpha-amylase |
| 9 | 147.8 | 15.3 | 9.84E+06 | P01042 | Kininogen-1 |
| 8 | 144.38 | 12.2 | 8.22E+06 | Q14624 | Inter-alpha-trypsin inhibitor heavy chain H4 |
| 8 | 126.26 | 36.3 | 1.38E+07 | P10451 | Osteopontin |
| 7 | 120.82 | 2.4 | 3.48E+06 | P98160 | Basement membrane-specific heparan sulfate proteoglycan core protein |
| 6 | 119.93 | 24.6 | 3.38E+06 | P01009 | Alpha-1-antitrypsin |
| 7 | 114.17 | 16.7 | 6.05E+06 | Q6EMK4 | Vasorin |
| 7 | 110.66 | 13.3 | 2.39E+06 | P35527 | Keratin, type I cytoskeletal 9 |
| 6 | 110.11 | 15.3 | 1.75E+06 | P01833 | Polymeric immunoglobulin receptor |
| 5 | 94.22 | 79.2 | 2.30E+07 | P01834 | Ig kappa chain C region |
| 5 | 85.37 | 24.4 | 1.26E+06 | P25311 | Zinc-alpha-2-glycoprotein OS=Homo sapiens |
| 1 | 13.26 | 4.9 | 1.46E+05 | A8MT79 | Putative zinc-alpha-2-glycoprotein-like 1 |
| 5 | 79.53 | 16.6 | 4.32E+05 | P05155 | Plasma protease C1 inhibitor |
| 5 | 79.32 | 24.7 | 1.91E+06 | Q12907 | Vesicular integral-membrane protein VIP36 |
| 5 | 75.62 | 5.9 | 6.02E+05 | P01133 | Pro-epidermal growth factor |
| 4 | 73.04 | 11 | 1.97E+06 | P00734 | Prothrombin |
| 4 | 72.69 | 23.4 | 2.73E+05 | P02749 | Beta-2-glycoprotein 1 |
| 5 | 70.77 | 8.5 | 1.08E+06 | P02538 | Keratin, type II cytoskeletal 6A |
| 5 | 70.77 | 8.5 | 1.08E+06 | P48668 | Keratin, type II cytoskeletal 6C |
| 3 | 44.03 | 4.9 | 9.16E+05 | P13647 | Keratin, type II cytoskeletal 5 |

| | | | | | |
|---|---|---|---|---|---|
| 2 | 29.6 | 3.5 | 6.07E+05 | Q5XKE5 | Keratin, type II cytoskeletal 79 |
| 3 | 42.21 | 5.6 | 2.63E+05 | P19013 | Keratin, type II cytoskeletal 4 |
| 5 | 70.49 | 13.7 | 6.84E+05 | P08779 | Keratin, type I cytoskeletal 16 |
| 3 | 46.09 | 6.7 | 4.59E+05 | P02533 | Keratin, type I cytoskeletal 14 |
| 3 | 46.09 | 7 | 4.59E+05 | P19012 | Keratin, type I cytoskeletal 15 |
| 2 | 30.35 | 4.6 | 3.04E+05 | Q04695 | Keratin, type I cytoskeletal 17 |
| 2 | 27.71 | 5 | 2.62E+05 | P13646 | Keratin, type I cytoskeletal 13 |
| 1 | 18.38 | 2.2 | 1.97E+05 | P08727 | Keratin, type I cytoskeletal 19 |
| 4 | 69.37 | 36.6 | 3.25E+05 | Q7Z5L0 | Vitelline membrane outer layer protein 1 homolog |
| 4 | 67.81 | 27.7 | 1.13E+06 | P02765 | Alpha-2-HS-glycoprotein |
| 4 | 65.5 | 17.2 | 1.17E+06 | P01876 | Ig alpha-1 chain C region OS=Homo sapiens GN=IGHA1 PE=1 SV=2 |
| 3 | 48.28 | 14.4 | 7.96E+05 | P01877 | Ig alpha-2 chain C region |
| 3 | 63.27 | 32.6 | 1.67E+07 | P41222 | Prostaglandin-H2 D-isomerase |
| 3 | 61.55 | 6.3 | 1.21E+05 | P10253 | Lysosomal alpha-glucosidase |
| 3 | 59.66 | 11.6 | 1.99E+06 | P15309 | Prostatic acid phosphatase |
| 4 | 59.07 | 6.3 | 1.54E+05 | P00450 | Ceruloplasmin |
| 3 | 56.69 | 16.3 | 1.12E+06 | Q9UNN8 | Endothelial protein C receptor |
| 3 | 55.51 | 4.1 | 1.13E+06 | P16070 | CD44 antigen |
| 3 | 49.93 | 16.4 | 5.06E+05 | P02750 | Leucine-rich alpha-2-glycoprotein |
| 3 | 49.34 | 21.4 | 3.08E+05 | O75594 | Peptidoglycan recognition protein |
| 3 | 43.97 | 19.5 | 7.77E+06 | P13987 | CD59 glycoprotein |
| 2 | 41.46 | 30.2 | 1.01E+06 | P04206 | Ig kappa chain V-III region GOL |
| 2 | 41.46 | 30.2 | 1.01E+06 | P01622 | Ig kappa chain V-III region Ti |
| 2 | 41.46 | 30.2 | 1.01E+06 | P01620 | Ig kappa chain V-III region SIE |
| 2 | 41.46 | 30.2 | 1.01E+06 | P01623 | Ig kappa chain V-III region WOL |
| 1 | 22.69 | 12.4 | 8.99E+05 | P18135 | Ig kappa chain V-III region HAH |
| 1 | 22.69 | 12.4 | 8.99E+05 | P18136 | Ig kappa chain V-III region HIC |
| 2 | 39.17 | 28.3 | 1.72E+06 | P0CG04 | Ig lambda-1 chain C regions |
| 2 | 39.17 | 28.3 | 1.72E+06 | P0CG05 | Ig lambda-2 chain C regions |
| 2 | 39.17 | 28.3 | 1.72E+06 | P0CF74 | Ig lambda-6 chain C region |
| 2 | 39.17 | 28.3 | 1.72E+06 | P0CG06 | Ig lambda-3 chain C regions |
| 1 | 17.36 | 14.1 | 2.73E+05 | A0M8Q6 | Ig lambda-7 chain C region |
| 1 | 21.26 | 14.2 | 4.83E+05 | P01842 | Ig lambda chain C regions |

| | | | | | |
|---|---|---|---|---|---|
| 2 | 38.98 | 14.5 | 6.68E+05 | Q8WVN6 | Secreted and transmembrane protein 1 |
| 2 | 38.92 | 21.8 | 3.00E+05 | P19652 | Alpha-1-acid glycoprotein 2 |
| 2 | 37.92 | 9.6 | 1.35E+05 | P08571 | Monocyte differentiation antigen CD14 |
| 2 | 37.2 | 13.8 | 3.99E+05 | Q16270 | Insulin-like growth factor-binding protein 7 |
| 2 | 37.12 | 14.2 | 2.14E+05 | Q9UBC9 | Small proline-rich protein 3 |
| 2 | 36.8 | 5 | 3.68E+05 | P02671 | Fibrinogen alpha chain |
| 2 | 35.97 | 18 | 2.17E+05 | P24855 | Deoxyribonuclease-1 |
| 2 | 34.83 | 9.3 | 2.09E+05 | P10153 | Non-secretory ribonuclease |
| 2 | 34.28 | 31.8 | 2.29E+05 | P01617 | Ig kappa chain V-II region TEW |
| 1 | 15.86 | 9.7 | 1.52E+05 | P06310 | Ig kappa chain V-II region RPMI 6410 |
| 1 | 15.86 | 11.3 | 1.52E+05 | P01614 | Ig kappa chain V-II region Cum |
| 1 | 15.86 | 11.1 | 1.52E+05 | P06309 | Ig kappa chain V-II region GM607 (Fragment) |
| 2 | 33.11 | 10.2 | 6.24E+05 | Q16769 | Glutaminyl-peptide cyclotransferase |
| 2 | 32.82 | 3.4 | 8.09E+05 | P14543 | Nidogen-1 |
| 2 | 32.44 | 13.2 | 5.00E+05 | P05090 | Apolipoprotein D |
| 2 | 32.41 | 37.8 | 7.26E+05 | P55000 | Secreted Ly-6/uPAR-related protein 1 |
| 2 | 31.79 | 5.8 | 6.08E+04 | P02787 | Serotransferrin |
| 2 | 31.52 | 4.8 | 7.40E+04 | P04217 | Alpha-1B-glycoprotein |
| 2 | 28.05 | 4.9 | 4.15E+05 | P05154 | Plasma serine protease inhibitor |
| 2 | 27.98 | 30.1 | 1.75E+06 | P07998 | Ribonuclease pancreatic |
| 2 | 26.79 | 8 | 2.99E+05 | P07288 | Prostate-specific antigen |
| 1 | 21.58 | 15.7 | 1.56E+05 | P01593 | Ig kappa chain V-I region AG |
| 1 | 21.58 | 15.7 | 1.56E+05 | P01594 | Ig kappa chain V-I region AU |
| 1 | 21.58 | 15.7 | 1.56E+05 | P01607 | Ig kappa chain V-I region Rei |
| 1 | 21.58 | 15.7 | 1.56E+05 | P01610 | Ig kappa chain V-I region WEA |
| 1 | 21.58 | 15.7 | 1.56E+05 | P01599 | Ig kappa chain V-I region Gal |
| 1 | 21.58 | 15.7 | 1.56E+05 | P01608 | Ig kappa chain V-I region Roy |
| 1 | 21.58 | 15.7 | 1.56E+05 | P01609 | Ig kappa chain V-I region Scw |
| 1 | 21.58 | 15.7 | 1.56E+05 | P80362 | Ig kappa chain V-I region WAT |
| 1 | 21.58 | 15.7 | 1.56E+05 | P01600 | Ig kappa chain V-I region Hau |
| 1 | 20.23 | 14.9 | 3.51E+04 | P01625 | Ig kappa chain V-IV region Len |
| 1 | 19.71 | 3 | 4.28E+04 | P04279 | Semenogelin-1 |
| 1 | 19.67 | 9.5 | 1.56E+05 | P09564 | T-cell antigen CD7 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 19.54 | 33.8 | 1.94E+06 | P60022 | Beta-defensin 1 |
| 1 | 19.32 | 1.5 | 1.06E+05 | P55290 | Cadherin-13 |
| 1 | 18.39 | 4.6 | 4.84E+05 | P02774 | Vitamin D-binding protein |
| 1 | 18.24 | 6.9 | 2.85E+05 | P02763 | Alpha-1-acid glycoprotein 1 |
| 1 | 17.92 | 18.4 | 1.00E+06 | P61769 | Beta-2-microglobulin |
| 1 | 17.82 | 11.2 | 2.53E+05 | Q14508 | WAP four-disulfide core domain protein 2 |
| 1 | 17.23 | 2.9 | 6.61E+04 | Q08380 | Galectin-3-binding protein |
| 1 | 16.1 | 5.1 | 5.26E+04 | P10909 | Clusterin |
| 1 | 15.81 | 15.5 | 3.80E+04 | Q01469 | Fatty acid-binding protein, epidermal |
| 1 | 15.81 | 20.7 | 3.80E+04 | A8MUU1 | Putative fatty acid-binding protein 5-like protein 3 |
| 1 | 15.12 | 3.1 | 1.00E+05 | Q8WZ75 | Roundabout homolog 4 |
| 1 | 15.05 | 3.8 | 2.65E+04 | P22792 | Carboxypeptidase N subunit 2 |
| 1 | 14.51 | 13.1 | 6.25E+04 | P06702 | Protein S100-A9 |
| 1 | 14.33 | 1.4 | 1.67E+05 | Q8IWU5 | Extracellular sulfatase Sulf-2 OS=Homo sapiens |
| 1 | 14.23 | 3.7 | 2.73E+04 | P36896 | Activin receptor type-1B |
| 1 | 14.2 | 20.4 | 3.57E+04 | Q9H299 | SH3 domain-binding glutamic acid-rich-like protein 3 |
| 1 | 14.04 | 2 | 2.60E+04 | P12830 | Cadherin-1 |
| 1 | 13.87 | 29.6 | 6.38E+04 | P11684 | Uteroglobin |
| 1 | 13.82 | 3.8 | 4.82E+04 | Q9BRK3 | Matrix-remodeling-associated protein 8 |
| 1 | 13.36 | 1.7 | 6.71E+04 | P39060 | Collagen alpha-1(XVIII) chain |
| 1 | 13.25 | 21 | 2.99E+04 | P62988 | Ubiquitin |
| 1 | 13.1 | 4.5 | 3.27E+04 | P02790 | Hemopexin |

## Saliva Unfractionated Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 41 | 689.84 | 11.3 | 2.50E+07 | Q9HC84 | Mucin-5B |
| 3 | 48.8 | 0.7 | 3.84E+05 | P98088 | Mucin-5AC (Fragments) |
| 33 | 532.82 | 58.9 | 2.97E+07 | P02768 | Serum albumin |
| 25 | 507.56 | 67.1 | 3.71E+08 | P04745 | Alpha-amylase 1 |
| 18 | 363.52 | 46.7 | 2.23E+08 | P19961 | Alpha-amylase 2B |
| 17 | 344.83 | 45.4 | 2.07E+08 | P04746 | Pancreatic alpha-amylase |
| 19 | 333.39 | 35.8 | 1.41E+07 | P01833 | Polymeric immunoglobulin receptor |
| 11 | 198.38 | 61.7 | 3.00E+07 | P01876 | Ig alpha-1 chain C region |
| 8 | 139.51 | 39.4 | 2.56E+07 | P01877 | Ig alpha-2 chain C region |
| 9 | 151.85 | 39.2 | 3.66E+06 | P25311 | Zinc-alpha-2-glycoprotein |
| 3 | 39.44 | 13.2 | 1.98E+05 | A8MT79 | Putative zinc-alpha-2-glycoprotein-like 1 |
| 9 | 147.94 | 32.5 | 4.69E+06 | Q8N4F0 | Bactericidal/permeability-increasing protein-like 1 |
| 8 | 124.34 | 34.1 | 2.50E+06 | Q96DR5 | Short palate, lung and nasal epithelium carcinoma-associated protein 2 |
| 7 | 118.85 | 45.1 | 7.10E+06 | Q96DA0 | Zymogen granule protein 16 homolog |
| 6 | 115.61 | 61.7 | 4.01E+07 | P01036 | Cystatin-S |
| 6 | 115.35 | 60.2 | 2.77E+07 | P01037 | Cystatin-SN |
| 4 | 69.47 | 42.5 | 1.23E+07 | P09228 | Cystatin-SA |
| 6 | 106.13 | 43.1 | 6.81E+06 | P12273 | Prolactin-inducible protein |
| 6 | 102.13 | 22 | 3.40E+06 | P23280 | Carbonic anhydrase 6 |
| 6 | 101.26 | 79.2 | 5.33E+06 | P01834 | Ig kappa chain C region |
| 6 | 93.21 | 61.2 | 7.04E+06 | P28325 | Cystatin-D |
| 5 | 80.69 | 44.3 | 3.12E+06 | P31025 | Lipocalin-1 |
| 2 | 26.49 | 11.1 | 1.40E+06 | Q5VSP4 | Putative lipocalin 1-like protein 1 |
| 4 | 75.43 | 65 | 3.70E+06 | P0CG05 | Ig lambda-2 chain C regions |
| 4 | 75.43 | 65 | 3.70E+06 | P0CG06 | Ig lambda-3 chain C regions |
| 3 | 60.73 | 46.2 | 1.91E+06 | P0CF74 | Ig lambda-6 chain C region |
| 3 | 56.91 | 47.1 | 3.02E+06 | P0CG04 | Ig lambda-1 chain C regions |
| 2 | 38.67 | 32 | 1.47E+06 | A0M8Q6 | Ig lambda-7 chain C region |
| 4 | 73.04 | 39.8 | 3.17E+06 | P61626 | Lysozyme C |
| 4 | 70.51 | 14.4 | 2.94E+06 | P13646 | Keratin, type I cytoskeletal 13 |
| 2 | 37.45 | 3.8 | 2.24E+06 | Q99456 | Keratin, type I cytoskeletal 12 |
| 1 | 19.06 | 4 | 4.88E+05 | P08727 | Keratin, type I cytoskeletal 19 |

| | | | | | |
|---|---|---|---|---|---|
| 4 | 67.27 | 7.3 | 5.10E+05 | P02788 | Lactotransferrin |
| 3 | 60.54 | 14.8 | 9.72E+05 | P01857 | Ig gamma-1 chain C region |
| 1 | 16.99 | 4.2 | 4.50E+05 | P01860 | Ig gamma-3 chain C region |
| 1 | 16.99 | 4.8 | 4.50E+05 | P01861 | Ig gamma-4 chain C region |
| 4 | 59.36 | 29.5 | 8.11E+05 | P01591 | Immunoglobulin J chain |
| 2 | 45.17 | 65.8 | 9.07E+07 | P02814 | Submaxillary gland androgen-regulated protein 3B |
| 3 | 43.96 | 9 | 1.60E+06 | Q8TAX7 | Mucin-7 OS=Homo sapiens |
| 3 | 42.2 | 37.7 | 1.99E+05 | P06702 | Protein S100-A9 |
| 2 | 38.83 | 26.8 | 6.19E+05 | P61769 | Beta-2-microglobulin |
| 2 | 37.37 | 12.2 | 3.15E+05 | P63261 | Actin, cytoplasmic 2 |
| 2 | 37.37 | 12.2 | 3.15E+05 | P60709 | Actin, cytoplasmic 1 |
| 1 | 16.64 | 1.4 | 1.27E+05 | A5A3E0 | POTE ankyrin domain family member F |
| 1 | 16.64 | 1.4 | 1.27E+05 | Q6S8J3 | POTE ankyrin domain family member E |
| 1 | 16.64 | 4.2 | 1.27E+05 | P68133 | Actin, alpha skeletal muscle |
| 1 | 16.64 | 4.2 | 1.27E+05 | P68032 | Actin, alpha cardiac muscle 1 |
| 1 | 16.64 | 4.2 | 1.27E+05 | Q9BYX7 | Putative beta-actin-like protein 3 |
| 1 | 16.64 | 4.2 | 1.27E+05 | P62736 | Actin, aortic smooth muscle |
| 1 | 16.64 | 4.2 | 1.27E+05 | Q562R1 | Beta-actin-like protein 2 |
| 1 | 16.64 | 4.2 | 1.27E+05 | P63267 | Actin, gamma-enteric smooth muscle |
| 2 | 36.36 | 11.2 | 7.75E+05 | Q6P5S2 | UPF0762 protein C6orf58 |
| 2 | 33.2 | 19.1 | 6.37E+05 | P01034 | Cystatin-C |
| 2 | 33.02 | 9.7 | 7.40E+05 | P54108 | Cysteine-rich secretory protein 3 |
| 2 | 32.09 | 3.3 | 2.58E+05 | P22079 | Lactoperoxidase |
| 2 | 29.81 | 33.6 | 8.38E+05 | P04080 | Cystatin-B |
| 2 | 28.67 | 20.4 | 9.55E+04 | P03973 | Antileukoproteinase |
| 1 | 19.48 | 2.8 | 1.63E+05 | P19013 | Keratin, type II cytoskeletal 4 |
| 1 | 19.19 | 0.5 | 9.45E+05 | Q9UGM3 | Deleted in malignant brain tumors 1 protein |
| 1 | 18.7 | 4.7 | 1.68E+05 | P04259 | Keratin, type II cytoskeletal 6B |
| 1 | 18.7 | 4.7 | 1.68E+05 | P02538 | Keratin, type II cytoskeletal 6A |
| 1 | 18.7 | 4.7 | 1.68E+05 | P48668 | Keratin, type II cytoskeletal 6C |
| 1 | 18.34 | 5 | 1.10E+05 | P80303 | Nucleobindin-2 |
| 1 | 18.21 | 2.7 | 1.86E+05 | P20061 | Transcobalamin-1 |
| 1 | 17.17 | 12.3 | 1.10E+05 | P10599 | Thioredoxin |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 16.77 | 9.7 | 1.72E+05 | P06310 | Ig kappa chain V-II region RPMI 6410 |
| 1 | 16.77 | 11.3 | 1.72E+05 | P01614 | Ig kappa chain V-II region Cum |
| 1 | 16.77 | 11.1 | 1.72E+05 | P06309 | Ig kappa chain V-II region GM607 |
| 1 | 16.77 | 11.5 | 1.72E+05 | P01617 | Ig kappa chain V-II region TEW |
| 1 | 16.09 | 48.3 | 3.41E+07 | P02808 | Statherin |
| 1 | 15.83 | 28 | 2.29E+05 | P15515 | Histatin-1 |
| 1 | 15.7 | 15 | 1.94E+05 | P01766 | Ig heavy chain V-III region BRO |
| 1 | 15.7 | 15.1 | 1.94E+05 | P01777 | Ig heavy chain V-III region TEI |
| 1 | 15.03 | 12.4 | 7.14E+04 | P18135 | Ig kappa chain V-III region HAH |
| 1 | 15.03 | 12.4 | 7.14E+04 | P18136 | Ig kappa chain V-III region HIC |
| 1 | 15.03 | 14.6 | 7.14E+04 | P04206 | Ig kappa chain V-III region GOL |
| 1 | 15.03 | 14.6 | 7.14E+04 | P01622 | Ig kappa chain V-III region Ti |
| 1 | 15.03 | 14.6 | 7.14E+04 | P01620 | Ig kappa chain V-III region SIE |
| 1 | 15.03 | 14.6 | 7.14E+04 | P01623 | Ig kappa chain V-III region WOL |
| 1 | 13.68 | 2 | 1.34E+05 | P02787 | Serotransferrin |
| 1 | 13.37 | 2.4 | 8.73E+04 | P01871 | Ig mu chain C region |
| 1 | 13.37 | 2.8 | 8.73E+04 | P04220 | Ig mu heavy chain disease protein |

## Vaginal Fluid Unfractionated Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 52 | 847.87 | 13.3 | 4.79E+07 | Q09666 | Neuroblast differentiation-associated protein AHNAK |
| 42 | 753.16 | 70.9 | 1.05E+09 | P02768 | Serum albumin |
| 25 | 423.88 | 88.7 | 4.89E+08 | Q9UBC9 | Small proline-rich protein 3 |
| 24 | 419.66 | 68.4 | 5.78E+07 | P07355 | Annexin A2 |
| 17 | 285.75 | 46 | 3.90E+07 | A6NMY6 | Putative annexin |
| 24 | 414.07 | 58.9 | 7.78E+07 | P13646 | Keratin, type I cytoskeletal 13 |
| 3 | 57.83 | 5.4 | 1.85E+07 | Q99456 | Keratin, type I cytoskeletal 12 |
| 3 | 49.5 | 4.2 | 5.12E+06 | P35900 | Keratin, type I cytoskeletal 20 |
| 1 | 19.31 | 3 | 1.49E+06 | Q8N1A0 | Keratin-like protein KRT222 |
| 1 | 15.11 | 2 | 6.39E+05 | Q2M2I5 | Keratin, type I cytoskeletal 24 |
| 8 | 123.93 | 12.5 | 1.41E+07 | P19012 | Keratin, type I cytoskeletal 15 |
| 4 | 69.46 | 6.7 | 6.11E+06 | P02533 | Keratin, type I cytoskeletal 14 |
| 4 | 69.46 | 6.7 | 6.11E+06 | P08779 | Keratin, type I cytoskeletal 16 |
| 3 | 50.82 | 4.6 | 4.56E+06 | Q04695 | Keratin, type I cytoskeletal 17 |
| 5 | 80.89 | 9.5 | 2.32E+06 | P13645 | Keratin, type I cytoskeletal 10 |
| 23 | 412.11 | 47.1 | 2.68E+07 | P02788 | Lactotransferrin |
| 22 | 389.74 | 39.9 | 2.19E+07 | P02787 | Serotransferrin |
| 21 | 376.19 | 52.3 | 6.90E+07 | P29508 | Serpin B3 |
| 12 | 194.04 | 33.8 | 3.89E+07 | P48594 | Serpin B4 |
| 18 | 314.26 | 53.1 | 5.37E+07 | P04083 | Annexin A1 |
| 19 | 309.55 | 34.5 | 4.13E+07 | P02538 | Keratin, type II cytoskeletal 6A |
| 3 | 41.32 | 3.6 | 1.55E+06 | P12035 | Keratin, type II cytoskeletal 3 |
| 2 | 29.48 | 1.8 | 1.32E+06 | Q01546 | Keratin, type II cytoskeletal 2 oral |
| 2 | 29.48 | 2.1 | 1.32E+06 | O95678 | Keratin, type II cytoskeletal 75 |
| 2 | 29.48 | 2.2 | 1.32E+06 | Q5XKE5 | Keratin, type II cytoskeletal 79 |
| 1 | 18.81 | 2.9 | 1.08E+05 | Q14CN4 | Keratin, type II cytoskeletal 72 |
| 17 | 263.55 | 29.9 | 2.63E+07 | P04259 | Keratin, type II cytoskeletal 6B |
| 11 | 170.1 | 19.4 | 1.74E+07 | P13647 | Keratin, type II cytoskeletal 5 |
| 3 | 50.1 | 3.7 | 3.24E+06 | P35908 | Keratin, type II cytoskeletal 2 |
| 2 | 32.47 | 4.1 | 3.98E+06 | P05787 | Keratin, type II cytoskeletal 8 |
| 20 | 304.68 | 21.3 | 7.53E+06 | P01023 | Alpha-2-macroglobulin |
| 2 | 31.14 | 2 | 9.51E+05 | P20742 | Pregnancy zone protein |

| | | | | | |
|---|---|---|---|---|---|
| 17 | 283.75 | 27.3 | 3.56E+07 | P04264 | Keratin, type II cytoskeletal 1 |
| 1 | 20.62 | 2 | 1.92E+06 | Q7Z794 | Keratin, type II cytoskeletal 1b |
| 1 | 18.57 | 2.3 | 2.95E+06 | P08729 | Keratin, type II cytoskeletal 7 |
| 1 | 18.57 | 2.4 | 2.95E+06 | Q6KB66 | Keratin, type II cytoskeletal 80 |
| 1 | 18.57 | 2.5 | 2.95E+06 | P14136 | Glial fibrillary acidic protein |
| 12 | 189.75 | 23 | 2.30E+07 | P19013 | Keratin, type II cytoskeletal 4 |
| 12 | 244.23 | 88.4 | 5.64E+07 | P68871 | Hemoglobin subunit beta |
| 1 | 18.47 | 6.8 | 2.44E+06 | P02100 | Hemoglobin subunit epsilon |
| 1 | 18.47 | 6.8 | 2.44E+06 | P69891 | Hemoglobin subunit gamma-1 |
| 1 | 18.47 | 6.8 | 2.44E+06 | P69892 | Hemoglobin subunit gamma-2 |
| 7 | 146.42 | 48.9 | 3.33E+07 | P02042 | Hemoglobin subunit delta |
| 13 | 237.08 | 48.8 | 5.43E+07 | Q9UBG3 | Cornulin OS=Homo sapiens |
| 12 | 233.38 | 53 | 9.33E+07 | P01857 | Ig gamma-1 chain C region |
| 8 | 131.82 | 35.8 | 1.62E+07 | P01859 | Ig gamma-2 chain C region |
| 7 | 125.84 | 33.9 | 2.10E+07 | P01861 | Ig gamma-4 chain C region |
| 13 | 233.3 | 70.3 | 6.72E+07 | Q01469 | Fatty acid-binding protein, epidermal |
| 2 | 42.47 | 26.7 | 1.63E+07 | A8MUU1 | Putative fatty acid-binding protein 5-like protein 3 |
| 16 | 233.19 | 15.8 | 6.26E+06 | A8K2U0 | Alpha-2-macroglobulin-like protein 1 |
| 13 | 218 | 31.4 | 4.20E+06 | P13796 | Plastin-2 |
| 3 | 46.9 | 5.5 | 1.17E+06 | P13797 | Plastin-3 |
| 2 | 32.04 | 3.4 | 7.47E+05 | Q14651 | Plastin-1 |
| 15 | 207.5 | 9.8 | 2.86E+06 | P01024 | Complement C3 |
| 11 | 204.17 | 62.3 | 3.94E+07 | P31151 | Protein S100-A7 |
| 3 | 57.35 | 23.7 | 1.86E+07 | Q86SG5 | Protein S100-A7A |
| 12 | 194.91 | 4.7 | 6.12E+06 | Q9Y6R7 | IgGFc-binding protein |
| 11 | 174.31 | 27.6 | 1.50E+07 | P07476 | Involucrin |
| 10 | 168.05 | 61.1 | 3.26E+07 | P80188 | Neutrophil gelatinase-associated lipocalin |
| 7 | 124.77 | 64 | 5.57E+07 | P69905 | Hemoglobin subunit alpha |
| 8 | 123.28 | 29.3 | 1.48E+07 | P63261 | Actin, cytoplasmic 2 |
| 8 | 123.28 | 29.3 | 1.48E+07 | P60709 | Actin, cytoplasmic 1 |
| 4 | 57.89 | 9.5 | 4.11E+06 | P68133 | Actin, alpha skeletal muscle |
| 4 | 57.89 | 9.5 | 4.11E+06 | P68032 | Actin, alpha cardiac muscle 1 |
| 4 | 57.89 | 9.5 | 4.11E+06 | P62736 | Actin, aortic smooth muscle |

| | | | | | |
|---|---|---|---|---|---|
| 4 | 57.89 | 9.5 | 4.11E+06 | P63267 | Actin, gamma-enteric smooth muscle |
| 3 | 47.44 | 2.7 | 4.58E+06 | A5A3E0 | POTE ankyrin domain family member F |
| 3 | 47.44 | 2.7 | 4.58E+06 | Q6S8J3 | POTE ankyrin domain family member E |
| 3 | 47.44 | 2.7 | 4.58E+06 | P0CG38 | POTE ankyrin domain family member I |
| 2 | 27.54 | 1.8 | 1.70E+06 | P0CG39 | POTE ankyrin domain family member J |
| 1 | 17.22 | 3.4 | 1.42E+06 | Q9BYX7 | Putative beta-actin-like protein 3 |
| 7 | 116.59 | 36.2 | 1.05E+07 | P01876 | Ig alpha-1 chain C region |
| 4 | 63.78 | 21.4 | 4.12E+06 | P01877 | Ig alpha-2 chain C region |
| 7 | 111.57 | 14.3 | 2.67E+06 | P01833 | Polymeric immunoglobulin receptor |
| 7 | 106.6 | 19.9 | 1.91E+06 | Q9UIV8 | Serpin B13 |
| 5 | 103.33 | 64.1 | 6.95E+07 | P01834 | Ig kappa chain C region |
| 6 | 97.2 | 87.6 | 7.92E+07 | P22528 | Cornifin-B |
| 5 | 80.41 | 78.6 | 7.53E+07 | P35321 | Cornifin-A |
| 6 | 96.59 | 45.6 | 6.72E+07 | P06702 | Protein S100-A9 |
| 6 | 93.59 | 23.2 | 2.58E+06 | P12429 | Annexin A3 |
| 7 | 93.18 | 11.3 | 2.73E+06 | P14780 | Matrix metalloproteinase-9 |
| 5 | 88.11 | 21.6 | 2.36E+06 | P02790 | Hemopexin |
| 4 | 85.66 | 65 | 2.77E+07 | P0CG05 | Ig lambda-2 chain C regions |
| 4 | 85.66 | 65 | 2.77E+07 | P0CG06 | Ig lambda-3 chain C regions |
| 3 | 65.99 | 47.1 | 2.00E+07 | P0CG04 | Ig lambda-1 chain C regions |
| 3 | 65.55 | 46.2 | 2.49E+07 | P0CF74 | Ig lambda-6 chain C region |
| 2 | 40.77 | 32 | 1.03E+07 | A0M8Q6 | Ig lambda-7 chain C region |
| 6 | 85.62 | 18.5 | 1.63E+06 | P02675 | Fibrinogen beta chain |
| 5 | 85.06 | 44 | 2.09E+07 | P05109 | Protein S100-A8 |
| 4 | 83.58 | 64.7 | 1.07E+07 | P31949 | Protein S100-A11 |
| 5 | 81.66 | 5.1 | 1.73E+06 | Q9NQ38 | Serine protease inhibitor Kazal-type 5 |
| 4 | 77.65 | 21.9 | 1.10E+07 | O95274 | Ly6/PLAUR domain-containing protein |
| 5 | 76.99 | 24.3 | 1.94E+06 | P02647 | Apolipoprotein A-I |
| 4 | 74.65 | 32.3 | 3.95E+06 | P02763 | Alpha-1-acid glycoprotein 1 |
| 2 | 33.8 | 12.9 | 1.94E+06 | P19652 | Alpha-1-acid glycoprotein 2 |
| 4 | 72.92 | 66.6 | 3.61E+07 | P35326 | Small proline-rich protein 2A |
| 1 | 17.82 | 17.8 | 2.63E+07 | Q9BYE4 | Small proline-rich protein 2G |
| 4 | 66.38 | 66.6 | 3.09E+07 | P35325 | Small proline-rich protein 2B |

| | | | | | |
|---|---|---|---|---|---|
| 3 | 48.95 | 54.1 | 3.04E+07 | P22532 | Small proline-rich protein 2D |
| 3 | 51.55 | 66.6 | 3.28E+07 | P22531 | Small proline-rich protein 2E |
| 4 | 65.82 | 75.5 | 6.95E+06 | P01040 | Cystatin-A |
| 4 | 62.89 | 6.3 | 1.36E+06 | P05164 | Myeloperoxidase |
| 4 | 62.47 | 39.7 | 3.70E+06 | P04080 | Cystatin-B |
| 3 | 61.15 | 38 | 6.01E+06 | P10599 | Thioredoxin |
| 3 | 60.69 | 35.1 | 7.39E+06 | P61626 | Lysozyme C OS=Homo sapiens |
| 4 | 56.7 | 14 | 1.85E+06 | P00738 | Haptoglobin OS=Homo sapiens |
| 2 | 28.75 | 6.8 | 3.41E+05 | P00739 | Haptoglobin-related protein |
| 3 | 55.79 | 13.7 | 1.95E+06 | P04406 | Glyceraldehyde-3-phosphate dehydrogenase |
| 3 | 55.66 | 22.4 | 4.64E+05 | P30086 | Phosphatidylethanolamine-binding protein 1 |
| 3 | 51.45 | 20.5 | 2.98E+06 | P08246 | Neutrophil elastase |
| 2 | 50.87 | 17.9 | 1.62E+06 | P54108 | Cysteine-rich secretory protein 3 |
| 3 | 49.79 | 24.2 | 7.85E+05 | P18510 | Interleukin-1 receptor antagonist protein |
| 3 | 49.27 | 4.1 | 1.67E+05 | P00450 | Ceruloplasmin |
| 3 | 48.99 | 17.1 | 4.44E+05 | P32119 | Peroxiredoxin-2 |
| 3 | 48.92 | 9.2 | 4.89E+05 | P01871 | Ig mu chain C region |
| 2 | 26.68 | 6.6 | 1.83E+05 | P04220 | Ig mu heavy chain disease protein |
| 3 | 48.41 | 11.9 | 8.05E+05 | Q9UKR3 | Kallikrein-13 |
| 3 | 46.49 | 52.8 | 2.51E+06 | P07108 | Acyl-CoA-binding protein |
| 3 | 42.8 | 29 | 1.18E+06 | P19957 | Elafin |
| 2 | 41.75 | 30.2 | 1.82E+06 | P04206 | Ig kappa chain V-III region GOL |
| 2 | 41.75 | 30.2 | 1.82E+06 | P01622 | Ig kappa chain V-III region Ti |
| 2 | 41.75 | 30.2 | 1.82E+06 | P01620 | Ig kappa chain V-III region SIE |
| 2 | 41.75 | 30.2 | 1.82E+06 | P01623 | Ig kappa chain V-III region WOL |
| 1 | 20.34 | 12.4 | 4.86E+05 | P18135 | Ig kappa chain V-III region HAH |
| 1 | 20.34 | 12.4 | 4.86E+05 | P18136 | Ig kappa chain V-III region HIC |
| 2 | 39.18 | 0.7 | 5.01E+05 | Q9HC84 | Mucin-5B |
| 2 | 37.04 | 13.1 | 3.13E+05 | Q92876 | Kallikrein-6 |
| 2 | 36.24 | 24.1 | 3.13E+06 | P01766 | Ig heavy chain V-III region BRO |
| 1 | 21.57 | 15.1 | 2.96E+06 | P01777 | Ig heavy chain V-III region TEI |
| 1 | 14.67 | 9.5 | 1.77E+05 | P01767 | Ig heavy chain V-III region BUT |
| 1 | 14.67 | 9.6 | 1.77E+05 | P01763 | Ig heavy chain V-III region WEA |

| | | | | | |
|---|---|---|---|---|---|
| 2 | 36.2 | 19.1 | 1.07E+07 | P59666 | Neutrophil defensin 3 |
| 2 | 36.2 | 19.1 | 1.07E+07 | P59665 | Neutrophil defensin 1 |
| 2 | 33.97 | 22.9 | 7.76E+05 | P04792 | Heat shock protein beta-1 |
| 2 | 33.78 | 5.9 | 6.52E+05 | O95171 | Sciellin |
| 2 | 33.78 | 11.7 | 6.76E+05 | P08311 | Cathepsin G |
| 2 | 32.92 | 36.1 | 6.55E+06 | Q96RM1 | Small proline-rich protein 2F |
| 2 | 32.56 | 4.6 | 1.25E+06 | Q16610 | Extracellular matrix protein 1 |
| 2 | 32.25 | 4.9 | 3.17E+05 | P29401 | Transketolase |
| 2 | 31.65 | 5.9 | 1.95E+05 | P22894 | Neutrophil collagenase |
| 2 | 31.62 | 4.5 | 4.08E+05 | P14618 | Pyruvate kinase isozymes M1/M2 |
| 2 | 30.83 | 7.1 | 5.92E+05 | P37837 | Transaldolase |
| 2 | 29.54 | 20.4 | 2.69E+06 | P03973 | Antileukoproteinase |
| 2 | 28.69 | 7 | 6.36E+05 | P07951 | Tropomyosin beta chain |
| 2 | 28.69 | 8 | 6.36E+05 | P67936 | Tropomyosin alpha-4 chain |
| 1 | 14.42 | 3.5 | 5.55E+05 | P06753 | Tropomyosin alpha-3 chain |
| 1 | 14.42 | 3.5 | 5.55E+05 | P09493 | Tropomyosin alpha-1 chain |
| 2 | 28.26 | 12.9 | 6.82E+05 | Q6UWP8 | Suprabasin |
| 2 | 27.75 | 6.6 | 2.74E+05 | P24158 | Myeloblastin |
| 2 | 27.02 | 3 | 1.20E+06 | P0CG48 | Polyubiquitin-C |
| 2 | 27.02 | 9.1 | 1.20E+06 | P0CG47 | Polyubiquitin-B |
| 2 | 27.02 | 13.4 | 7.91E+05 | P62979 | Ubiquitin-40S ribosomal protein S27a |
| 2 | 27.02 | 16.4 | 7.91E+05 | P62987 | Ubiquitin-60S ribosomal protein L40 |
| 2 | 26.63 | 7.4 | 1.34E+06 | P04075 | Fructose-bisphosphate aldolase A |
| 2 | 25.38 | 10.4 | 9.06E+05 | P16402 | Histone H1.3 |
| 2 | 25.38 | 10.5 | 9.06E+05 | P10412 | Histone H1.4 |
| 2 | 25.38 | 10.7 | 9.06E+05 | P16403 | Histone H1.2 |
| 1 | 23.21 | 15.7 | 6.14E+05 | P01593 | Ig kappa chain V-I region AG |
| 1 | 23.21 | 15.7 | 6.14E+05 | P01594 | Ig kappa chain V-I region AU |
| 1 | 23.21 | 15.7 | 6.14E+05 | P01607 | Ig kappa chain V-I region Rei |
| 1 | 23.21 | 15.7 | 6.14E+05 | P01610 | Ig kappa chain V-I region WEA |
| 1 | 23.21 | 15.7 | 6.14E+05 | P01599 | Ig kappa chain V-I region Gal |
| 1 | 23.21 | 15.7 | 6.14E+05 | P01608 | Ig kappa chain V-I region Roy |
| 1 | 23.21 | 15.7 | 6.14E+05 | P01609 | Ig kappa chain V-I region Scw |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 23.21 | 15.7 | 6.14E+05 | P80362 | Ig kappa chain V-I region WAT |
| 1 | 23.21 | 15.7 | 6.14E+05 | P01600 | Ig kappa chain V-I region Hau |
| 1 | 21.86 | 15.7 | 2.71E+06 | Q9NZT1 | Calmodulin-like protein 5 |
| 1 | 21.51 | 4.2 | 1.35E+05 | Q9UL52 | Transmembrane protease serine 11E |
| 1 | 21.09 | 19.5 | 5.04E+05 | P80511 | Protein S100-A12 |
| 1 | 18.83 | 29.2 | 3.42E+05 | Q9UGL9 | Cysteine-rich C-terminal protein 1 |
| 1 | 18.52 | 2.6 | 3.35E+05 | P02679 | Fibrinogen gamma chain |
| 1 | 18.36 | 1.9 | 3.29E+05 | P22735 | Protein-glutamine gamma-glutamyltransferase K |
| 1 | 18.2 | 15.3 | 5.87E+04 | P01776 | Ig heavy chain V-III region WAS |
| 1 | 18.2 | 15.1 | 5.87E+04 | P01774 | Ig heavy chain V-III region POM |
| 1 | 18.2 | 15.5 | 5.87E+04 | P01779 | Ig heavy chain V-III region TUR |
| 1 | 18.2 | 15.6 | 5.87E+04 | P01765 | Ig heavy chain V-III region TIL |
| 1 | 17.97 | 8.8 | 1.83E+05 | P02766 | Transthyretin |
| 1 | 17.54 | 11.4 | 3.54E+05 | P62158 | Calmodulin |
| 1 | 17.24 | 4.6 | 1.18E+06 | P02774 | Vitamin D-binding protein |
| 1 | 16.68 | 11.4 | 9.20E+05 | P07737 | Profilin-1 |
| 1 | 16.59 | 0.8 | 5.16E+04 | P0C0L5 | Complement C4-B |
| 1 | 16.59 | 0.8 | 5.16E+04 | P0C0L4 | Complement C4-A |
| 1 | 16.28 | 13.6 | 8.12E+05 | P32320 | Cytidine deaminase |
| 1 | 16.25 | 7.8 | 2.71E+05 | P14174 | Macrophage migration inhibitory factor |
| 1 | 16.08 | 21.4 | 2.51E+05 | P00441 | Superoxide dismutase [Cu-Zn] |
| 1 | 15.75 | 15.7 | 3.70E+05 | Q5VTM1 | Protein FAM25 |
| 1 | 15.61 | 9.7 | 2.63E+05 | P06310 | Ig kappa chain V-II region RPMI 6410 |
| 1 | 15.61 | 11.3 | 2.63E+05 | P01614 | Ig kappa chain V-II region Cum |
| 1 | 15.61 | 11.1 | 2.63E+05 | P06309 | Ig kappa chain V-II region GM607 (Fragment) |
| 1 | 15.61 | 11.5 | 2.63E+05 | P01617 | Ig kappa chain V-II region TEW |
| 1 | 15.48 | 7.8 | 2.93E+05 | P22392 | Nucleoside diphosphate kinase B |
| 1 | 15.48 | 7.8 | 2.93E+05 | P15531 | Nucleoside diphosphate kinase A |
| 1 | 14.9 | 12.2 | 7.46E+04 | P35754 | Glutaredoxin-1 |
| 1 | 14.82 | 9.4 | 9.36E+04 | P47914 | 60S ribosomal protein L29 |
| 1 | 14.45 | 3 | 1.37E+04 | Q6XPR3 | Repetin |
| 1 | 14.43 | 43.1 | 2.55E+05 | Q08EQ4 | Putative thymosin beta-4-like protein |
| 1 | 14.43 | 43.1 | 2.55E+05 | A8MW06 | Thymosin beta-4-like protein 3 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 14.43 | 43.1 | 2.55E+05 | P62328 | Thymosin beta-4 |
| 1 | 14.43 | 43.1 | 2.55E+05 | A9Z1Y9 | Putative thymosin beta-4-like protein 6 |
| 1 | 14.4 | 4.6 | 6.36E+04 | Q9BW04 | Specifically androgen-regulated gene protein |
| 1 | 14.27 | 0.6 | 2.15E+05 | P48552 | Nuclear receptor-interacting protein 1 |
| 1 | 14.18 | 6.4 | 1.36E+05 | P49913 | Cathelicidin antimicrobial peptide |
| 1 | 14.09 | 9.7 | 4.38E+05 | P62805 | Histone H4 |
| 1 | 13.86 | 2.3 | 1.28E+05 | Q05639 | Elongation factor 1-alpha 2 |
| 1 | 13.86 | 2.3 | 1.28E+05 | Q5VTE0 | Putative elongation factor 1-alpha-like 3 |
| 1 | 13.86 | 2.3 | 1.28E+05 | P68104 | Elongation factor 1-alpha 1 |
| 1 | 13.67 | 4.3 | 4.41E+05 | P52209 | 6-phosphogluconate dehydrogenase, decarboxylating |
| 1 | 13.65 | 12.2 | 3.73E+04 | P61019 | Ras-related protein Rab-2A |
| 1 | 13.64 | 27.4 | 3.10E+05 | Q14508 | WAP four-disulfide core domain protein 2 |

## Menstrual Blood Unfractionated Protein Identification Results

| Distinct Peptides | MS/MS Score | % AA Coverage | Spectral Intensity | SwissProt Accession | Protein Name |
|---|---|---|---|---|---|
| 28 | 488.77 | 54.5 | 6.82E+07 | P02768 | Serum albumin |
| 28 | 482.75 | 56.4 | 2.10E+07 | P02788 | Lactotransferrin |
| 22 | 408.31 | 56.3 | 1.32E+08 | P13646 | Keratin, type I cytoskeletal 13 |
| 6 | 108.41 | 10.5 | 6.22E+07 | P19012 | Keratin, type I cytoskeletal 15 |
| 2 | 42.5 | 3.8 | 7.40E+06 | Q99456 | Keratin, type I cytoskeletal 12 |
| 2 | 40.79 | 6.2 | 8.84E+06 | P08727 | Keratin, type I cytoskeletal 19 |
| 2 | 35.98 | 2.5 | 1.33E+07 | P35900 | Keratin, type I cytoskeletal 20 |
| 2 | 35.98 | 2.5 | 1.33E+07 | Q04695 | Keratin, type I cytoskeletal 17 |
| 1 | 18.14 | 3 | 1.55E+06 | Q8N1A0 | Keratin-like protein KRT222 |
| 1 | 17.64 | 2 | 2.26E+06 | Q2M2I5 | Keratin, type I cytoskeletal 24 |
| 9 | 147.59 | 21.2 | 7.75E+06 | P13645 | Keratin, type I cytoskeletal 10 |
| 5 | 90.22 | 10.7 | 3.07E+07 | P08779 | Keratin, type I cytoskeletal 16 |
| 4 | 75.25 | 6.7 | 3.00E+07 | P02533 | Keratin, type I cytoskeletal 14 |
| 19 | 336.92 | 35.1 | 7.22E+07 | P02538 | Keratin, type II cytoskeletal 6A |
| 5 | 84.44 | 7.4 | 1.87E+07 | O95678 | Keratin, type II cytoskeletal 75 |
| 4 | 65.21 | 4.7 | 9.41E+06 | P12035 | Keratin, type II cytoskeletal 3 |
| 4 | 64.74 | 5.7 | 1.36E+07 | Q5XKE5 | Keratin, type II cytoskeletal 79 |
| 3 | 48.6 | 2.9 | 6.21E+06 | Q01546 | Keratin, type II cytoskeletal 2 oral |
| 18 | 309.07 | 33.3 | 6.65E+07 | P48668 | Keratin, type II cytoskeletal 6C |
| 17 | 298.89 | 30.3 | 6.09E+07 | P04259 | Keratin, type II cytoskeletal 6B |
| 11 | 183.65 | 17.7 | 3.00E+07 | P13647 | Keratin, type II cytoskeletal 5 |
| 6 | 107.08 | 7.9 | 2.16E+07 | P35908 | Keratin, type II cytoskeletal 2 epidermal |
| 3 | 54.81 | 5.5 | 8.64E+06 | P05787 | Keratin, type II cytoskeletal 8 |
| 13 | 282.18 | 89.7 | 8.56E+08 | P68871 | Hemoglobin subunit beta |
| 1 | 18.52 | 6.8 | 7.69E+07 | P02100 | Hemoglobin subunit epsilon |
| 10 | 203.79 | 72.1 | 2.24E+08 | P02042 | Hemoglobin subunit delta |
| 3 | 57.89 | 23.8 | 7.80E+07 | P69891 | Hemoglobin subunit gamma-1 |
| 3 | 57.89 | 23.8 | 7.80E+07 | P69892 | Hemoglobin subunit gamma-2 |
| 16 | 258.57 | 16.9 | 1.75E+07 | P01023 | Alpha-2-macroglobulin |
| 1 | 16.96 | 0.8 | 1.27E+06 | P20742 | Pregnancy zone protein |
| 16 | 253.88 | 48.3 | 2.71E+07 | P02647 | Apolipoprotein A-I |
| 13 | 249.46 | 26.5 | 2.48E+07 | P04264 | Keratin, type II cytoskeletal 1 |

| | | | | | |
|---|---|---|---|---|---|
| 2 | 41.05 | 3.2 | 8.46E+06 | Q7Z794 | Keratin, type II cytoskeletal 1b |
| 1 | 22.48 | 2.4 | 1.55E+06 | Q6KB66 | Keratin, type II cytoskeletal 80 |
| 1 | 22.48 | 2.5 | 1.55E+06 | P14136 | Glial fibrillary acidic protein |
| 8 | 133.15 | 20.2 | 9.51E+06 | P19013 | Keratin, type II cytoskeletal 4 |
| 3 | 52.38 | 5.7 | 3.87E+06 | P08729 | Keratin, type II cytoskeletal 7 |
| 16 | 244.61 | 30.7 | 8.24E+06 | P05164 | Myeloperoxidase |
| 2 | 24.86 | 2.9 | 4.77E+05 | P11678 | Eosinophil peroxidase |
| 14 | 233.13 | 37.6 | 9.93E+06 | P02675 | Fibrinogen beta chain |
| 14 | 213.77 | 11.4 | 5.14E+06 | P01024 | Complement C3 |
| 10 | 197.45 | 45.8 | 2.29E+07 | P63261 | Actin, cytoplasmic 2 |
| 10 | 197.45 | 45.8 | 2.29E+07 | P60709 | Actin, cytoplasmic 1 |
| 5 | 111.44 | 6.8 | 1.21E+07 | Q6S8J3 | POTE ankyrin domain family member E |
| 5 | 91.69 | 15.1 | 1.26E+07 | P68133 | Actin, alpha skeletal muscle |
| 5 | 91.69 | 15.1 | 1.26E+07 | P68032 | Actin, alpha cardiac muscle 1 |
| 4 | 87.98 | 4.7 | 1.10E+07 | A5A3E0 | POTE ankyrin domain family member F |
| 3 | 69.95 | 10.6 | 7.68E+06 | Q9BYX7 | Putative beta-actin-like protein 3 |
| 4 | 68.59 | 12.2 | 9.30E+06 | P62736 | Actin, aortic smooth muscle |
| 4 | 68.59 | 12.2 | 9.30E+06 | P63267 | Actin, gamma-enteric smooth muscle |
| 3 | 65.99 | 3.2 | 8.02E+06 | P0CG38 | POTE ankyrin domain family member I |
| 2 | 47.96 | 2.3 | 4.71E+06 | P0CG39 | POTE ankyrin domain family member J |
| 2 | 35.17 | 6.3 | 4.97E+06 | Q562R1 | Beta-actin-like protein 2 |
| 8 | 168 | 83.8 | 4.49E+08 | P69905 | Hemoglobin subunit alpha |
| 10 | 152.37 | 26.8 | 5.61E+06 | P08670 | Vimentin OS=Homo sapiens |
| 9 | 147.8 | 20 | 9.36E+06 | P02787 | Serotransferrin |
| 8 | 133.99 | 36.7 | 3.98E+06 | P04083 | Annexin A1 |
| 7 | 125.13 | 27.8 | 1.07E+07 | Q9UBG3 | Cornulin |
| 8 | 122.74 | 6.8 | 1.04E+06 | P35579 | Myosin-9 |
| 1 | 13.54 | 0.5 | 9.82E+04 | P35580 | Myosin-10 |
| 1 | 13.4 | 0.8 | 1.75E+05 | Q7Z406 | Myosin-14 |
| 1 | 13.4 | 0.8 | 1.75E+05 | P35749 | Myosin-11 |
| 6 | 114.3 | 55.2 | 3.59E+07 | P06702 | Protein S100-A9 |
| 7 | 106.99 | 23.8 | 4.43E+06 | P02679 | Fibrinogen gamma chain |
| 7 | 105.63 | 25.1 | 2.68E+06 | P00738 | Haptoglobin OS=Homo sapiens |

| | | | | | |
|---|---|---|---|---|---|
| 3 | 45.95 | 12.3 | 7.37E+05 | P00739 | Haptoglobin-related protein |
| 6 | 100.79 | 20.5 | 1.87E+06 | Q5VTE0 | Putative elongation factor 1-alpha-like 3 |
| 6 | 100.79 | 20.5 | 1.87E+06 | P68104 | Elongation factor 1-alpha 1 |
| 3 | 49.92 | 6.6 | 7.25E+05 | Q05639 | Elongation factor 1-alpha 2 |
| 6 | 96.79 | 18.4 | 7.44E+06 | P01009 | Alpha-1-antitrypsin |
| 6 | 94.82 | 33 | 2.56E+06 | P01857 | Ig gamma-1 chain C region |
| 2 | 28.63 | 11.6 | 6.41E+05 | P01861 | Ig gamma-4 chain C region |
| 3 | 45.88 | 15.1 | 2.11E+06 | P01860 | Ig gamma-3 chain C region |
| 7 | 93.14 | 19.7 | 1.23E+06 | P68363 | Tubulin alpha-1B chain |
| 6 | 81.25 | 16.9 | 1.07E+06 | Q9BQE3 | Tubulin alpha-1C chain |
| 5 | 71.76 | 15.6 | 7.07E+05 | P68366 | Tubulin alpha-4A chain |
| 5 | 65.28 | 13.7 | 1.00E+06 | Q71U36 | Tubulin alpha-1A chain |
| 3 | 38.44 | 9.7 | 5.31E+05 | Q13748 | Tubulin alpha-3C/D chain |
| 3 | 38.44 | 9.7 | 5.31E+05 | Q6PEY2 | Tubulin alpha-3E chain |
| 2 | 30.95 | 7.1 | 2.97E+05 | Q9NY65 | Tubulin alpha-8 chain |
| 1 | 15.97 | 5.8 | 6.47E+04 | Q9H853 | Putative tubulin-like protein alpha-4B |
| 1 | 15.92 | 2 | 1.44E+05 | A6NHL2 | Tubulin alpha chain-like 3 |
| 4 | 83.3 | 37.4 | 6.36E+06 | P08246 | Neutrophil elastase |
| 5 | 80.88 | 41.7 | 1.66E+07 | P62805 | Histone H4 |
| 5 | 75.16 | 40.3 | 1.23E+06 | P60660 | Myosin light polypeptide 6 |
| 2 | 30.32 | 11.5 | 8.20E+05 | P14649 | Myosin light chain 6B |
| 5 | 74.41 | 21.5 | 1.34E+06 | Q9BVA1 | Tubulin beta-2B chain |
| 5 | 74.41 | 21.5 | 1.34E+06 | Q13885 | Tubulin beta-2A chain |
| 5 | 74.41 | 21.5 | 1.34E+06 | P68371 | Tubulin beta-2C chain |
| 5 | 74.41 | 21.6 | 1.34E+06 | P07437 | Tubulin beta chain |
| 5 | 74.41 | 21.6 | 1.34E+06 | P04350 | Tubulin beta-4 chain |
| 2 | 28.84 | 6.8 | 3.84E+05 | Q13509 | Tubulin beta-3 chain |
| 2 | 28.27 | 6.2 | 1.41E+05 | Q9BUF5 | Tubulin beta-6 chain |
| 1 | 14.58 | 3.1 | 7.63E+04 | Q3ZCM7 | Tubulin beta-8 chain |
| 1 | 14.58 | 3.1 | 7.63E+04 | A6NNZ2 | Tubulin beta-8 chain B |
| 1 | 14.58 | 3.7 | 7.63E+04 | A6NKZ8 | Putative tubulin beta chain-like protein ENSP00000290377 |
| 5 | 72.14 | 28.6 | 2.01E+06 | P01876 | Ig alpha-1 chain C region |
| 3 | 39.18 | 18.5 | 1.22E+06 | P01877 | Ig alpha-2 chain C region |

| | | | | | |
|---|---|---|---|---|---|
| 4 | 69.78 | 39.8 | 1.59E+06 | P20160 | Azurocidin |
| 5 | 69.64 | 22.8 | 1.13E+06 | P63104 | 14-3-3 protein zeta/delta |
| 2 | 21.96 | 7 | 6.17E+05 | P62258 | 14-3-3 protein epsilon |
| 2 | 21.96 | 7.2 | 6.17E+05 | P61981 | 14-3-3 protein gamma |
| 2 | 21.96 | 7.3 | 6.17E+05 | Q04917 | 14-3-3 protein eta |
| 2 | 21.96 | 7.3 | 6.17E+05 | P31946 | 14-3-3 protein beta/alpha |
| 2 | 21.96 | 7.3 | 6.17E+05 | P27348 | 14-3-3 protein theta |
| 4 | 55.74 | 16.9 | 1.26E+06 | P31947 | 14-3-3 protein sigma |
| 4 | 69.26 | 6.9 | 5.07E+05 | P02730 | Band 3 anion transport protein |
| 3 | 58.1 | 24.7 | 2.19E+07 | P05109 | Protein S100-A8 |
| 4 | 56.99 | 19.6 | 5.29E+06 | P02790 | Hemopexin |
| 3 | 55.36 | 43.8 | 2.68E+06 | P04908 | Histone H2A type 1-B/E |
| 3 | 55.36 | 43.8 | 2.68E+06 | Q7L7L0 | Histone H2A type 3 |
| 3 | 55.36 | 43.8 | 2.68E+06 | P20671 | Histone H2A type 1-D |
| 3 | 55.36 | 43.8 | 2.68E+06 | Q93077 | Histone H2A type 1-C |
| 3 | 55.36 | 43.8 | 2.68E+06 | P0C0S8 | Histone H2A type 1 |
| 3 | 55.36 | 44.1 | 2.68E+06 | Q9BTM1 | Histone H2A.J |
| 3 | 55.36 | 44.5 | 2.68E+06 | Q99878 | Histone H2A type 1-J |
| 3 | 55.36 | 44.5 | 2.68E+06 | Q96KK5 | Histone H2A type 1-H |
| 2 | 35.35 | 26.5 | 1.89E+06 | P16104 | Histone H2A.x |
| 2 | 35.35 | 29.2 | 1.89E+06 | Q8IUE6 | Histone H2A type 2-B |
| 1 | 13.09 | 6.8 | 1.05E+06 | Q96QV6 | Histone H2A type 1-A |
| 1 | 13.09 | 7 | 1.05E+06 | P0C0S5 | Histone H2A.Z |
| 1 | 13.09 | 7 | 1.05E+06 | Q71UI9 | Histone H2A.V |
| 3 | 52.39 | 43.8 | 2.14E+06 | Q6FI13 | Histone H2A type 2-A |
| 3 | 52.39 | 44.1 | 2.14E+06 | Q16777 | Histone H2A type 2-C |
| 3 | 53.14 | 8.3 | 2.17E+05 | P06576 | ATP synthase subunit beta, mitochondrial |
| 3 | 52.82 | 50.9 | 4.15E+06 | P01834 | Ig kappa chain C region |
| 3 | 51.35 | 10.1 | 6.28E+05 | P02774 | Vitamin D-binding protein |
| 4 | 50.97 | 17.9 | 8.53E+05 | P27797 | Calreticulin |
| 3 | 50.61 | 6.6 | 5.93E+05 | P00747 | Plasminogen |
| 3 | 50.57 | 27.8 | 1.93E+06 | P04792 | Heat shock protein beta-1 |
| 3 | 50.16 | 18.9 | 6.03E+06 | P02763 | Alpha-1-acid glycoprotein 1 |

| | | | | | |
|---|---|---|---|---|---|
| 3 | 47.5 | 4.5 | 4.72E+06 | P00450 | Ceruloplasmin |
| 3 | 47.49 | 20.3 | 6.51E+05 | P00915 | Carbonic anhydrase 1 |
| 4 | 47.44 | 2 | 5.83E+05 | P15924 | Desmoplakin |
| 3 | 45.57 | 9.6 | 1.06E+06 | P04004 | Vitronectin |
| 3 | 45.33 | 27.1 | 4.21E+05 | P07737 | Profilin-1 |
| 3 | 44.18 | 11.3 | 1.23E+06 | P24158 | Myeloblastin |
| 3 | 41.93 | 8.4 | 3.51E+05 | P07237 | Protein disulfide-isomerase |
| 3 | 41.01 | 45.9 | 7.97E+05 | P04080 | Cystatin-B |
| 3 | 39.68 | 32.3 | 6.20E+05 | P47929 | Galectin-7 |
| 2 | 38.53 | 19 | 1.20E+07 | Q99879 | Histone H2B type 1-M |
| 2 | 38.53 | 19 | 1.20E+07 | Q99880 | Histone H2B type 1-L |
| 2 | 38.53 | 19 | 1.20E+07 | P33778 | Histone H2B type 1-B |
| 2 | 38.53 | 19 | 1.20E+07 | P57053 | Histone H2B type F-S |
| 2 | 38.53 | 19 | 1.20E+07 | P58876 | Histone H2B type 1-D |
| 2 | 38.53 | 19 | 1.20E+07 | Q99877 | Histone H2B type 1-N |
| 2 | 38.53 | 19 | 1.20E+07 | Q16778 | Histone H2B type 2-E |
| 2 | 38.53 | 19 | 1.20E+07 | Q5QNW6 | Histone H2B type 2-F |
| 2 | 38.53 | 19 | 1.20E+07 | Q8N257 | Histone H2B type 3-B |
| 2 | 38.53 | 19 | 1.20E+07 | P23527 | Histone H2B type 1-O |
| 2 | 38.53 | 19 | 1.20E+07 | P62807 | Histone H2B type 1-C/E/F/G/I |
| 2 | 38.53 | 19 | 1.20E+07 | P06899 | Histone H2B type 1-J |
| 2 | 38.53 | 19 | 1.20E+07 | Q93079 | Histone H2B type 1-H |
| 2 | 38.53 | 19 | 1.20E+07 | O60814 | Histone H2B type 1-K |
| 1 | 17.34 | 7 | 8.48E+06 | Q96A08 | Histone H2B type 1-A |
| 3 | 38.23 | 6.5 | 3.58E+05 | P11021 | 78 kDa glucose-regulated protein |
| 2 | 36.68 | 33 | 5.46E+05 | P0CG04 | Ig lambda-1 chain C regions |
| 2 | 36.68 | 33 | 5.46E+05 | P0CG05 | Ig lambda-2 chain C regions |
| 2 | 36.68 | 33 | 5.46E+05 | P0CG06 | Ig lambda-3 chain C regions |
| 1 | 19.96 | 14.1 | 3.21E+05 | A0M8Q6 | Ig lambda-7 chain C region |
| 1 | 19.96 | 14.1 | 3.21E+05 | P0CF74 | Ig lambda-6 chain C region |
| 2 | 36.25 | 4.6 | 3.23E+05 | Q8N1N4 | Keratin, type II cytoskeletal 78 |
| 2 | 35.43 | 11.9 | 1.27E+06 | P02765 | Alpha-2-HS-glycoprotein |
| 2 | 34.24 | 27 | 1.39E+06 | P61626 | Lysozyme C |

## Peripheral Blood Unfractionated Protein Identification Results

| | | | | | |
|---|---|---|---|---|---|
| 2 | 33.75 | 4.8 | 4.12E+05 | P01042 | Kininogen-1 |
| 2 | 33.68 | 21.7 | 3.00E+05 | P29373 | Cellular retinoic acid-binding protein 2 |
| 2 | 33.67 | 5.7 | 7.04E+05 | P01871 | Ig mu chain C region OS=Homo sapiens |
| 2 | 33.67 | 6.6 | 7.04E+05 | P04220 | Ig mu heavy chain disease protein |
| 2 | 33.03 | 25.9 | 6.00E+05 | Q01469 | Fatty acid-binding protein, epidermal |
| 1 | 13.44 | 20.7 | 2.45E+05 | A8MUU1 | Putative fatty acid-binding protein 5-like protein 3 |
| 2 | 32.55 | 1.6 | 9.04E+05 | P02751 | Fibronectin |
| 2 | 32.28 | 8.6 | 3.25E+05 | P10909 | Clusterin OS=Homo sapiens |
| 2 | 29.64 | 25.7 | 4.30E+05 | P31949 | Protein S100-A11 |
| 2 | 29.23 | 19.1 | 6.54E+05 | P59666 | Neutrophil defensin 3 |
| 2 | 29.23 | 19.1 | 6.54E+05 | P59665 | Neutrophil defensin 1 |
| 2 | 29.01 | 24.8 | 3.06E+05 | P62937 | Peptidyl-prolyl cis-trans isomerase A |
| 1 | 14.45 | 8.5 | 6.49E+04 | Q9Y536 | Peptidylprolyl cis-trans isomerase A-like 4A/B/C O |
| 2 | 28.71 | 7.3 | 2.17E+05 | P14923 | Junction plakoglobin |
| 2 | 27.57 | 1.7 | 2.79E+05 | P08603 | Complement factor H |
| 2 | 27.27 | 18.9 | 9.31E+05 | Q9UBC9 | Small proline-rich protein 3 |
| 2 | 25.94 | 9.5 | 3.09E+05 | P32119 | Peroxiredoxin-2 |
| 1 | 21.94 | 5.1 | 8.46E+04 | P30101 | Protein disulfide-isomerase A3 |
| 1 | 20.67 | 19.5 | 7.80E+04 | P80511 | Protein S100-A12 |
| 1 | 20.36 | 20 | 1.43E+06 | P02652 | Apolipoprotein A-II |
| 1 | 20 | 3 | 2.65E+05 | P04217 | Alpha-1B-glycoprotein |
| 1 | 19.46 | 8.8 | 2.04E+05 | P02766 | Transthyretin |
| 1 | 19.4 | 3.4 | 1.15E+05 | P08697 | Alpha-2-antiplasmin |
| 1 | 19.2 | 3.7 | 1.01E+05 | P25705 | ATP synthase subunit alpha, mitochondrial |
| 1 | 18.67 | 8.2 | 1.73E+05 | P08311 | Cathepsin G |
| 1 | 18.27 | 6.6 | 3.63E+05 | P02749 | Beta-2-glycoprotein 1 |
| 1 | 17.87 | 1.1 | 7.94E+04 | P11215 | Integrin alpha-M |
| 1 | 17.68 | 0.6 | 4.41E+04 | P21333 | Filamin-A |
| 1 | 17.63 | 23.5 | 3.44E+05 | P68431 | Histone H3.1 |
| 1 | 17.52 | 5.3 | 5.58E+04 | P02649 | Apolipoprotein E |
| 1 | 16.05 | 1.3 | 1.07E+05 | O43707 | Alpha-actinin-4 |
| 1 | 16.05 | 1.3 | 1.07E+05 | P12814 | Alpha-actinin-1 |
| 1 | 16.02 | 5 | 2.64E+05 | P28676 | Grancalcin |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 17.32 | 4.8 | 5.15E+05 | P01861 | Ig gamma-4 chain C region |
| 2 | 34.22 | 2.1 | 4.16E+05 | P0C0L4 | Complement C4-A |
| 2 | 34.22 | 2.1 | 4.16E+05 | P0C0L5 | Complement C4-B |
| 2 | 30.5 | 0.6 | 3.68E+05 | P04114 | Apolipoprotein B-100 |
| 2 | 29.14 | 3.6 | 2.94E+05 | P00734 | Prothrombin |
| 2 | 24.93 | 7.3 | 8.85E+05 | P01011 | Alpha-1-antichymotrypsin |
| 2 | 24.9 | 5.5 | 5.18E+05 | P06727 | Apolipoprotein A-IV |
| 1 | 18.85 | 0.9 | 1.65E+05 | Q14624 | Inter-alpha-trypsin inhibitor heavy chain H4 |
| 1 | 17.77 | 3.1 | 2.92E+05 | P06396 | Gelsolin |
| 1 | 17.72 | 3.3 | 1.59E+05 | P01871 | Ig mu chain C region |
| 1 | 17.72 | 3.8 | 1.59E+05 | P04220 | Ig mu heavy chain disease protein |
| 1 | 16.73 | 3.1 | 3.56E+05 | P02760 | Protein AMBP |
| 1 | 16.42 | 3.1 | 4.24E+05 | P04004 | Vitronectin |
| 1 | 15.88 | 4.9 | 2.84E+05 | P02749 | Beta-2-glycoprotein 1 |
| 1 | 14.86 | 2 | 5.00E+05 | P05155 | Plasma protease C1 inhibitor |
| 1 | 14.53 | 5.1 | 3.01E+05 | P10909 | Clusterin |
| 1 | 14.49 | 3 | 4.87E+05 | P04217 | Alpha-1B-glycoprotein |
| 1 | 13.29 | 3.4 | 2.13E+06 | P04040 | Catalase |