Final Technical Report

REPORT TITLE:	**Isolation of Highly Specific Protein Markers for the Identification of Biological Stains: Adapting Comparative Proteomics to Forensics**
AWARD NUMBER:	**DNA Research and Development Award 2006-DN-BX-K001**
AUTHOR:	**Phillip B. Danielson, Ph.D.**

ABSTRACT

Overview – While DNA profiling allows biological stains to be individualized, the unambiguous identification of the stain itself can present forensic serologists with a significant challenge. For example, there is no reliable test for vaginal secretions and tests for blood cannot distinguish peripheral from menstrual blood even though this information can be probative to an investigation.

Both mRNA and proteomic profiling represent reasonable and complementary approaches to clearly identifying a broader range of forensically relevant body fluids. Using advanced protein separation technology, a high-resolution profile of virtually every protein in six body fluids of particular value to the forensic community has been generated. By rigorously comparing these "proteomic profiles" it was hypothesized that a panel of high-specificity candidate protein biomarkers for individual body fluids could be identified. Collaboration with practitioners has helped to guide this research to best meet the needs of the forensic community.

Project Objectives - The broad goals of this research program was to:

(1) Collect samples of six forensically relevant body fluids (saliva, semen, peripheral blood, menstrual blood, vaginal secretions, and urine).

(2) Generate quantitative high-resolution profiles of target body fluid proteomes by 2-Dimensional High-Performance Liquid Chromatography.

(3) Employ a rigorous comparative proteomics strategy to obtain high-specificity biomarkers for unambiguous biological stain identification.

Results and Conclusions - All core objectives have been achieved. Through a combination of highly-reproducible 2D HPLC proteome fractionation and implementation of customized clustering algorithms, high-resolution consensus proteomic profiles (2D pI/hydrophobicity maps) were produced for each of six different bodily fluids. By conducting quantitative pair-wise comparisons among these datasets, it was possible to distinguish those proteins that were likely to be genuinely characteristic of a specific bodily fluid versus those that reflected interindividual variability in protein expression. False negatives were also minimized by normalizing the consensus datasets for subtle variations in pH, solvent gradients and protein concentration. Based

on these analyses, a comprehensive panel of candidate biomarkers was generated and characterized by mass spectrometry.

The results have been extremely promising with the apparent specificity of some biomarkers of saliva and semen (e.g., statherin and semenogelin-1 and 2) being independently by forensic researchers working on the development of mRNA biomarkers. A thorough validation of the specificity of these candidate biomarkers in a larger population group, and using forensic casework type samples represents the next step toward the development of a practitioner-ready high-specificity test for biological stain identification.

Table of Contents

EXECUTIVE SUMMARY

The central goal of the current research project was to isolate and identify candidate protein biomarkers that are highly specific to individual types of biological stains of forensic utility (*i.e.*, saliva, semen, peripheral blood, menstrual blood, vaginal secretions, and urine). The availability of such protein biomarkers can complement the use of DNA profiling in criminal investigations by making it possible to more accurately and confidently associate a DNA sample with a specific type of biological stain. The inability to confidently identify the nature of a biological stain can introduce ambiguity and present forensic serologists with a significant challenge when interpreting the significance of biological evidence in some criminal cases.

Introduction and Statement of Problem

Blood and semen proteins, which once held promise as discriminatory instruments for individualizing biological stains, have long been supplanted by DNA markers which can be amplified from tiny amounts of biological material. While DNA analysis of an evidentiary swab may reveal the presence of a DNA profile consistent with an alleged victim, the DNA profile cannot indicate if the DNA came from saliva, vaginal secretions, urine or a host of other sources. The ability to associate a DNA extract with a specific tissue source, however, can provide criminal investigators with critical information.

Consider the case of an alleged sexual assault where a DNA profile consistent with the victim is found on the suspect's fingers. The victim states that the suspect used his fingers to penetrate her vagina. The suspect counters that the alleged victim had been licking food off of his fingers and that no sexual contact occurred. Both stories account for the presence of the victim's DNA on the suspect's fingers. The ability to reliably differentiate between saliva and vaginal secretions in this case could help to either confirm or refute these opposing claims. Other scenarios can easily be imagined where the ability to differentiate between menstrual and peripheral blood, or urine and saliva would have probative value.

While tests for the detection of blood, semen, saliva, urine and fecal matter are available[1-4], some can be laborious (*e.g.* tests for creatinine and urobilinogen as indicators of urine and fecal matter). Some tests may consume significant amounts of precious evidence (*e.g.*, the test for amylase as a marker of saliva may use half of an evidentiary swab) and still fail to provide the specificity or sensitivity needed by forensic practitioners. Finally, some serological assays employ chemicals with known health risks (*e.g.*, mercuric chloride is a highly toxic compound used in chemical tests for fecal matter).

The development of commercial forensic kits has greatly facilitated routine forensic testing for blood and seminal fluid. Based on the detection of antigen-antibody interactions, these one-step immunoassay tests have provided forensic practitioners with high specificity and excellent sensitivity. For the generic detection of blood, the protein hemoglobin has served as an excellent diagnostic marker[4, 5]. Similarly, the p30 (Prostate-Specific Antigen) protein is widely used as a fairly reliable biomarker of seminal fluid [3, 6, 7]. For a range of other biological stains, however, forensically-validated commercial kits based on body fluid specific antigens are lacking. Part of the reason for this is that unlike hemoglobin and p30 which are abundant and relatively specific protein biomarkers, less has been known about the proteins present in other forensically-relevant body fluids. The proteomic profile of a given body fluid is fundamentally a function of the

specific genes that are transcribed into mRNA and then translated into functional proteins. Among the hundreds to thousands of proteins present in any given body fluid, many are be common to several body fluids while others are be highly specific markers of a single body fluid. By rigorously comparing the full complement of expressed proteins among different body fluids, it should be possible to compile a comprehensive list of candidate protein biomarkers with potential forensic utility for reliable and highly-specific stain identification. When combined with existing technology such as that used for ABAcard and Seratec® kits, it would be possible to develop low-cost assay systems for use by practitioners.

This same logic underlies the use of differential mRNA expression as a means of identifying body fluids[8-10]. Rather than being competing technologies, the use of mRNA and proteins as specific markers for body fluid identification is complementary in nature[11]. Messenger RNA profiling offers the potential for PCR-based amplification and compatibility with existing DNA technology. The use of protein based markers with high-sensitivity antibody-based assays offers the potential for direct body fluid identification without the need for an amplification step. Protein profiling may also facilitate the identification of urine and saliva where mRNA-based strategies have had limited success. In addition, while both mRNA and proteins have been shown to persist in dried forensic stains and post mortem tissues[12, 13], proteins are generally considered to be more stable and thus low-cost, immunochromatographic assays continue to yield reliable results even with degraded forensic samples[14, 15].

Core Research Objectives – The central goal of the current research project was to identify protein biomarkers with high specificity for the identification of biological stains of forensic utility (*i.e.*, saliva, semen, peripheral blood, menstrual blood, vaginal secretions, and urine). To achieve this goal, a comparative proteomics strategy was developed with three major research objectives. These were to:

1) Collect samples of six forensically relevant body fluids from a minimum of five individuals each in accordance with established National Institutes of Health protocols.
2) Perform high-resolution fractionation and quantitative mapping of the complete proteomic profile for each body fluid by 2-Dimensional High-Performance Liquid Chromatography (2-D HPLC).
3) Identify unique protein markers for each of the body fluids by conducting rigorous quantitative pair-wise comparisons of the proteomic profiles obtained for each body fluid and from each individual donor.

The successful completion of these objectives would aid forensic analysts by proving tools for the development of high-throughput multiplex assays for reliable biological stain identification.

Methods

This research and all the methods presented here have been approved by the University of Denver's Institutional Review Board for Research Involving Human Subjects. Furthermore, all samples were coded to ensure the anonymity of the donor.

Samples of six forensically-significant body fluids (*i.e.*, peripheral and menstrual blood, semen, saliva, vaginal secretions and urine) were collected from human research volunteers. The choice of these specific body fluids was based on discussions with forensic serologists at the Colorado Bureau of Investigation. For each body fluid, at least five individuals were recruited to donate a sample. This redundancy helped to discriminate (under objective #3) between proteins that were characteristic of a specific body fluid versus those that reflect inter-individual variability. Samples were collected using IRB approved standard operating procedures. Briefly, peripheral blood was obtained by venipuncture at the University of Denver Heath Center. Saliva was collected using the Sarstedt Salivette™ sponge. Morning urine samples and semen samples were self-collected. Vaginal secretions were self-collected using standard tampons based on National Institutes of Health protocols and menstrual blood was self-collected using an FDA-approved over-the-counter latex-free, hypoallergenic menstrual cup. All samples were transported to the laboratory on cold packs and then either processed immediately or stored until needed in a locked freezer.

After protein extraction and (as warranted) sample concentration, the proteome of each body fluid sample was fractionated and mapped by 2-Dimensional HPLC. This employed the commercial ProteomeLab™ PF2D System (Beckman Coulter, Fullerton, CA). This instrument, which was specifically designed for comparative proteomics, employs high-resolution 2-dimensional chromatography to fractionate the complex mixtures of proteins present biological samples. A high-performance chromatofocusing column is used for the first dimension to separate proteins based on their isoelectric point and a high-performance non-porous silica reverse-phase column is used for the second dimension to further fractionate proteins on the basis of hydrophobicity. All fractions from the second dimension output to a series of 96-well plates from which specific fractions of potential interest can be recovered for further analysis and protein identification/characterization (*e.g.*, by mass spectrometry)[16]. The reproducibility of the system ensures that different proteome maps can be readily and quantitatively compared to identify differences in protein expression.

An in-house bioinformatics application (ProteinMiner™) was developed to manage and analyze the average of 500,000+ data points that comprised each proteome map. This custom application provided a robust means of comparing different proteomes to tag potential biomarkers while taking into account subtle run-to-run and interindividual differences.

Proteins of interest as potential biomarkers were then identified by mass spectrometry on an Agilent 6300 series ion trap mass spectrometer coupled to a 1200 series HPLC-Chip/MS system. Data analysis was performed using Spectrum Mill software suite by Agilent Technologies. The Swiss-Prot database was used to match MS/MS spectrum generated on mass spectrometer. Typically proteins identified with 2+ peptides and scores >16 were considered confident matches.

Results and Discussion

Through the use of advanced 2-D HPLC protein separation technology, the complex mixtures of the hundreds to thousands of proteins present in any given body fluid were rigorously fractionated and quantified. This process made it possible to produced high-resolution proteomic profiles (*i.e.*, 2D pI/hydrophobicity maps) representing different bodily fluids and/or different individuals. For each bodily fluid, five individuals were recruited to provide a sample.

This redundancy helped The decision to initially analyze samples of each body fluid from this relatively small sample of individuals reflected a balance between maximizing the amount of comparative proteomic information within a reasonable budget and timeframe.

In the course of the current project, a series of comprehensive two-dimensional proteome profiles have been generated. These represent virtually every protein expressed in six targeted body fluids of value to the forensic community (*i.e.*, peripheral and menstrual blood, vaginal secretions, semen, urine, and saliva). While it is reasonable to expect that some proteins are found in multiple body fluids, the primary objective of the current research was to identify a definable subset of proteins that were unique to - and thus diagnostic for – a single body fluid.

A major advantage of this type of comparative "whole-proteome" strategy is that it requires no *a priori* assumptions with respect to the specific proteins expressed in any body fluid. Rather, the approach made it possible to rigorously evaluate the entire complement of proteins in each proteome for those with potential utility as bodily fluid specific biomarkers. The reproducibility with which the proteomes of each body fluid were fractionated by 2D HPLC helped to ensure that proteomic profiles generated at different times and from different samples could be quantitatively compared to each other.

Accurately comparing different proteomes for biomarker discovery, however, still requires that subtle run-to-run variations in pH and solvent gradients and protein concentration be taken into consideration. Similarly, differences in protein expression levels in the same body fluid among different individuals must be taken into consideration. This required that data-mining algorithms be implemented to combine the proteome maps generated from individual samples into a single proteome map representative of each body fluid. These "consensus" proteome maps could then serve as a reliable basis for comparing the body fluids to each other.

ProteinMiner™ is a bioinformatics software application written specifically for this project. Using this software, we were then able to select specific pI/hydrophobicity coordinates within each "consensus" proteome map where the most promising candidate biomarkers for subsequent identification by mass spectrometry would likely be found. The software begins by applying a peak identification algorithm to the 2D pI/hydrophobicity maps (typically a 45 x 10,501 data point matrix) to extract a normalized representative dataset from each proteome map. This technique eliminated virtually all of the underlying "noise" associated with the original proteome map while increasing the resolution and the speed of additional downstream analyses. A second set of data mining algorithms (k-means clustering and hierarchical clustering) were then implemented to combine the normalized representative datasets for each proteome. Data mining for this purpose can be defined as grouping like objects together to extract only the significant features from a set of proteomes. This process yielded a set of six "clustered/consensus map" that were the basis of all comparisons between different body fluids. Using this bioinformatics application, a comprehensive list of the proteome map coordinates corresponding to candidate biomarkers for each body fluid of interest was generated. Subsequent retrieval of the second dimension fractions represented by these proteome map coordinates allowed the specific identity of candidate biomarkers to be rapidly determined by ESI-MS/MS analysis.

In toto, >1000 proteins were identified in the course of this comparative proteome mapping. This included candidate proteins identified by: (1) ms/ms analysis of peaks identified by our comparative mapping software as unique; (2) ms/ms analysis of pH fractions and; (3) ms/ms data on unfractionated body fluid samples. Through careful evaluation, it was possible to significantly narrow this initial list to a reasonably accurate listing of high quality candidate biomarkers.

Two criteria were employed in selecting individual proteins for inclusion in the final list of candidate biomarkers. These were peak uniqueness (i.e., body fluid specificity) and protein abundance as reflected by the ProteomeLab™ PF2D chromatogram. In general, a chromatographic peak of > 0.2 AU (absorbance units) can be expected to yield a quality ms/ms identification. The relative abundance of a given biomarker was also derived using the ProteomeLab™ PF2D peak integration feature for quantitation. Semenogelin is a highly abundant protein which accounted for an average of 44.16% of the total protein content in semen. By contrast, statherin accounted for a mean of only 0.78% of total saliva protein content.

The results, (Executive Summary Table 1), are extremely promising. For example, the first protein (statherin) identified by ProteinMiner™ as being highly-specific for saliva was selected independently of any information other than our mapped proteomes. This results was particularly encouraging because statherin has been independently identified as a possible saliva marker by gene expression database searches and by forensic researchers working on the development of mRNA markers for saliva[9, 17]. Similarly, the identification of semenogelin-1 and -2 as markers of seminal fluid[28] and periplakin as marker of vaginal secretions are consistent with what has been reported by biomedical researchers[18]. The apparent accuracy with which this comparative proteomics approach has been able to readily identify these markers bodes well for the specificity of many of the other candidate biomarkers that have been identified. This is all the more important as tissue specificity information for many of these proteins is lacking in the scientific literature.

Executive Summary Table 1: Candidate Protein Biomarkers for Body Fluid Identification

Body Fluid	Candidate Protein Biomarker	Accession Number
Semen	Semenogelin 1	P04279
	Semenogelin 2	Q02383
	Epididymal secretory protein E1	P61916
	Dual specificity testis-specific protein kinase 2	Q96S53
	Prostatic Acid Phosphitase	P15309
	G2/mitotic-specific cyclin-B3	Q8WWL7
Saliva	Statherin	P02808
	Salivary acidic proline-rich phosphoprotein	P02810
	Cystatin_SA	P09228
	Cystatin_D	P28325
	Submaxillary gland androgen-regulated protein	P02814
Vaginal Secretions	Extracellular matrix protein 1	Q16610
	Glycodelin	P09466
	Matrigel-induced gene C4 protein	O95274
	Secreted glypican-3	P51654
	Vimentin	P08670
	Stratifin	P31947
	Involucrin	P07476
	Periplakin	O60437

	Gelsolin	P06396
	Vinexin	O60504
	Mesothelin	Q13421
Urine	Uromodulin	P07911
	Osteopontin	P10451
Menstrual Blood	Pregnancy zone protein	P20742
	Matrilysin	P09237
	Calpastatin	P20810
	SH2B adapter protein 2	O14492
Peripheral Blood	Hemopexin	P02790
	Histidine-rich glycoprotein	P04196
	Apolipoprotein	P04114
	Plasminogen	P00747
	Transthyretin	P02766
	Antithrombin-III	P01008
	Ceruloplasmin	P00450
	Afamin	P43652
	Serum amyloid P-component	P02743

It is important to emphasize, however, that these protein biomarkers were identified by mapping the protein profiles of just five individuals per bodily fluid and thus only be considered candidate biomarkers. Future research will involve a second round of selection (from the list of candidate biomarkers shown in Table 1) and validation for biomarkers that are best suited to forensic applications. This will necessarily place greater emphasis on absolute abundance based on the view that more abundant candidates should have more utility with degraded samples as well as more complex mixed samples. It would not be unreasonable for some of the current candidate biomarkers to be eliminated from further consideration in the course of rigorous validation studies. While the use of even a relatively small sample group can help to reduce the potentially misleading impact of interindividual differences in protein expression through the creation of "consensus proteome maps", the ultimate applicability of a given biomarker for use with the general population necessitates a more comprehensive and thorough validation of each candidate marker for stain specificity with a larger population set. There are good reasons for this. For example, the possibility cannot be ignored that some candidate biomarkers might be secreted into non-target fluids in the same way that A, B, and Rh factors in blood are found in the saliva or semen of individuals termed secretors. Confounding factors such as this might be missed when looking at proteome data from only five individuals. Only when these larger-scale studies are completed, can these markers move from being candidates to serving as the foundation for a commercial multiplex assay system capable of characterizing both single source and mixed-source stains with high specificity.

In summary, all research objectives under award 2006-DN-BX-K001 have been successfully completed. A thorough validation of specificity of these candidate biomarkers in a larger population group, and using forensic casework type samples therefore, represents the next step

toward the development of a practitioner-ready high-specificity test for biological stain identification. Ongoing forensic validation studies in the author's laboratory have already begun to yield useful data on the specificity of several of these biomarkers.

Implications for Policy, Practice and Future Research

The identification of biological stains represents a significant challenge for the forensic serologist. While commercial kits for the identification of blood, semen and saliva have proven useful, the comparative proteomic research reported here indicate that it may be possible to achieve far greater specificity in biological stain identification.

The availability of protein markers for several significant body fluids could enable a single multiplexed approach to body fluid identification. After rigorous forensic validation of the candidate protein biomarkers presented in this report, the most obvious commercial application may be the development highly-specific immunochromatographic assays. The utility and cost effectiveness of these assays, as exemplified by ABA card and Seratec® kits, is well established in the forensic community. It is even conceivable that a hand-held assay card could be designed that would be capable of analyzing and identifying mixtures of different body fluids without having to perform multiple assays. High-sensitivity antibody-based assays also offer the potential for direct body fluid identification without the need for an amplification step. This can be important from an analyst's perspective because it could save time and minimizes sample handling while reducing the consumption of valuable evidence.

As researchers move forward, it is important for all investigators to remain cognizant of the standards for admitting scientific evidence in the federal courts. Experiments must be planned with both Frye's "general acceptance" test, the Daubert standard and federal rules of evidence in mind. In this way, future studies, coupled with publication in peer-reviewed journals would help to place the findings of this research on sound legal footing.

Literature Cited in the Executive Summary

1. *Biology Methods Manual*. 1978: Metropolitan Police Forensic Science Laboratory.
2. *Protocol Manual*. 1989: FBI Laboratory Serology Unit.
3. 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): 1057-60.
4. 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): 597-602.
5. *OneStep ABAcard® HemaTrace® for the Forensic Identification of Human Blood.*, Abacus Diagnostics (insert)
6. Sensabaugh, G.F., *Isolation and characterization of a semen-specific protein from human seminal plasma: a potential new marker for semen identification.* J Forensic Sci, 1978. 23(1): 106-15.
7. *OneStep ABAcard® p30 Test for the Forensic Identification of Semen*, Abacus Diagnostics (insert)
8. Juusola, J. and J. Ballantyne, *Multiplex mRNA profiling for the identification of body fluids.* Forensic Sci Int, 2005. 152(1): 1-12.
9. 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): 85-96.

10. 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): 217-226.

11. Griffin, T.J., et al., *Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae.* Mol Cell Proteomics, 2002. 1(4): 323-33.

12. Bauer, M., et al., *Quantification of mRNA degradation as possible indicator of postmortem interval--a pilot study.* Leg Med (Tokyo), 2003. 5(4): 220-7.

13. Inoue, H., A. Kimura, and T. Tuji, *Degradation profile of mRNA in a dead rat body: basic semi-quantification study.* Forensic Sci Int, 2002. 130(2-3): 127-32.

14. Dobberstein, R.C., Huppertz, J., von Wurmb-Schwark, N., Ritz-Timme, S. *Degradation of biomolecules in artificially and naturally aged teeth: implications for age estimation based on aspartic acid racemization and DNA analysis.* Forensic Sci Int. 2008. **179**(2-3): 181-91.

15. Laux, D.L., Tambasco, A. J., Benzinger, E. A., *Forensic Detection of Semen II. Comparison of the Abacus Diagnostics OneStep ABAcard p30 Test and the Seratec PSA Semiquant Kit for the Determination of the Presence of Semen in Forensic Cases*.

16. Wall, D.B., Kachman, M.T., Gong, S., Hinderer, R., Parus, S., Misek, D.E., Hanash, S.M. and D.M. Lubman (2000) Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry. *Anal. Chem.* **72**:1099-111.

17. 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.

18. Dasari, S., et al., *Comprehensive proteomic analysis of human cervical-vaginal fluid.* J Proteome Res, 2007. **6**(4): p. 1258-68.

FINAL TECHNICAL REPORT (MAIN BODY)

Introduction and Statement of the Problem

Blood and semen proteins which once held promise as discriminatory instruments for individualizing biological stains have been supplanted by DNA markers which can be amplified from tiny amounts of biological material. DNA markers, however, do not provide the forensic analyst with a means of knowing the type of tissue or biological fluid from which the DNA was extracted. Thus, the analysis of a swab from a suspect may reveal the presence of a DNA profile consistent with that of an alleged victim but the DNA profile cannot indicate if the DNA came from saliva, vaginal secretions, urine or a host of other possible biological sources. The ability to associate a DNA extract with a specific tissue source, however, can be critical to a successful criminal investigation[1].

Consider the case of an alleged sexual assault where a DNA profile consistent with the victim is found on the fingers of a suspect. The victim states that the suspect used his fingers to penetrate her vagina. The suspect counters that the alleged victim had been licking food off of his fingers and that no sexual contact occurred. Both claims could account for the presence of the victim's DNA on the suspect's fingers. The ability to unambiguously differentiate between saliva and vaginal secretions in this case could help to either confirm or refute these opposing claims. In another example, consider a case where DNA consistent with the victim of an alleged sexual assault is found on a hand towel recovered from the suspect's van where the alleged assault took place. The victim claims that the attacker wore a condom and that she had used the towel to wipe blood from her vaginal area after the assault. The suspect claims that the victim was a hitch hiker to whom he had offered a ride and that the blood on the towel came from a nose bleed that the victim developed in his van. Both claims could account for the presence of the alleged victim's DNA on the towel. The ability to reliably detect the presence of a mixture of both blood and vaginal secretions in this case could help to either confirm or refute these opposing claims. A number of other scenarios could easily be imagined where the ability to characterize mixtures of body fluids and to differentiate between menstrual and venous blood, or urine and saliva would have important probative value. This would benefit investigation and would enable forensic analysts to make more definitive statements about the potential tissue source of a DNA profile.

The research funded under DNA Research and Development Award 2006-DN-BX-K001 employed advanced protein separation technology to generate a rigorous and comprehensive profile of virtually every protein in six body fluids of value to the forensic community. By quantitatively comparing these "proteomic profiles", it had been hypothesized that it would be possible to identify highly-specific candidate protein markers of individual body fluids. These markers could then be forensically validated for their target specificity and ultimately serve as the foundation for a low-cost, high-speed assay that would make it possible to analyze an unknown biological stain for multiple biological fluids and to characterize even complex mixtures of body fluids.

At the onset of this project, it was recognized that tests for the detection of blood, semen, saliva, urine and fecal matter were available[2-5] but that some of these can be laborious and time consuming (*e.g.* tests for creatinine and urobilinogen as indicators of urine and fecal matter, respectively). Some tests also consume significant amounts of a valuable sample. For example, the test for amylase as a marker of saliva may consume half of an evidentiary swab and still fail

to provide the specificity or sensitivity needed by forensic practitioners. In addition, forensic serologists must be proficient at a variety of diverse methodologies some of which employ reagents that pose health risks. For example, long-term exposure (even at low levels) to mercuric chloride which is used to test for fecal matter, may lead to a build-up of mercury in body organs and irreversible neurological damage.

The development of commercial forensic kits has greatly facilitated routine forensic testing for blood and seminal fluid. Based on the detection of antigen-antibody interactions, these one-step immunoassay tests have provided forensic practitioners with high specificity and excellent sensitivity. For the detection of blood, the Abacus Diagnostics' ABAcard and kits from Seratec® use the protein hemoglobin as a diagnostic marker[5, 6]. Similarly, seminal fluid detection is based on use of the p30 (Prostate-Specific Antigen) protein as a diagnostic marker[4, 7, 8]. Saliva presents a more complex challenge for the forensic analyst. The detection of saliva is generally based on assays for the presence of α-amylase (*i.e.*, salivary amylase)[9]. However, α-amylase activity is also present in a variety of non-salivary body fluids including human blood serum, urine and cervical mucus[10-12], although normally at much lower levels than in saliva. As a result, tests for saliva identification are necessarily presumptive assays. Accordingly, it would be a misrepresentation to tell a jury that because a vaginal swab tested positive for amylase, the presence of saliva on that vaginal swab has been confirmed. Certainly, α-amylase activity can be confirmed but saliva cannot be specifically confirmed. Being well aware of this, forensic analysts would be limited to testifying in this scenario that a positive amylase result is "consistent with saliva" and by extension "perhaps consistent with oral-genital contact". For a range of other body fluids, forensically-validated commercial kits based on body fluid specific antigens are lacking entirely and this leaves the forensic analyst without the ability to make any substantive statement about the potential tissue source of a DNA profile. Part of the reason for this is that unlike hemoglobin and p30 which are abundant and relatively specific antigenic markers, much less is known about the antigens that make up the protein profiles of other forensically-relevant body fluids.

The antigenic profile of a given body fluid is fundamentally a function of the specific genes that are transcribed into mRNA and then translated into proteins. Among the hundreds to thousands of proteins present in any given body fluid, many are common to several body fluids while others are highly specific markers of a single body fluid. By rigorously mapping and comparing the full proteomic profiles of different body fluids, it has been possible to provide the forensic community with a comprehensive database of virtually every protein with potential forensic utility as a unique marker for any given body fluid. When combined with existing technology such as that used for ABAcard and Seratec® kits, high-specificity protein biomarkers would allow the development of low-cost multiplex assay systems for body fluid identification.

This same logic underlies studies on the use of comparative transcriptomics and reverse transcription PCR (RT-PCR) to identify body fluids on the basis of unique mRNA expression profiles[13-15]. For example, matrix metalloproteinase mRNA transcripts expressed in the endometrium have been proposed as a marker for differentiating between menstrual and venous blood[16]. Similarly, several candidate mRNAs have been evaluated as markers of saliva[14]. Other researchers have demonstrated that mRNA degrades more slowly than previously thought in some dried stain[17] and that false positive results from the amplification of processed pseudogenes can be prevented by treating samples with DNase prior to amplification. For all

these reasons, and because of its compatibility with existing DNA technology, mRNA profiling has attracted significant research interest in the past few years.

Rather than being competing technologies, the use of protein biomarkers and mRNA for body fluid identification is complementary in nature[18]. Both of these approaches have the potential to provide a means of differentiating between such body fluids as menstrual and venous blood, but protein profiling may also facilitate the identification of body fluids such as urine and saliva where mRNA-based strategies have had variable success. The use of protein markers with high-sensitivity antibody-based assays also offers the potential for direct body fluid identification without the need for an amplification step. This can be important from an analyst's perspective because it saves time and avoids the need to treat some samples with DNase prior to amplification – a step which may raise concern in an environment where every effort is made to protect valuable and often minimal forensic samples from exposure to endogenous and/or exogenous sources of DNase. In addition, while both mRNA and proteins have been shown to persist in dried forensic stains and post mortem tissues[19, 20], proteins have generally been found to be more stable than mRNA with degraded forensic samples[21, 22]. Thus low-cost, immunochromatographic assays may continue to yield reliable results in situations where mRNA profiling is found to be a less amenable approach. One additional and significant advantage of a proteomic approach to body fluid identification is the tremendous diversity of potential protein targets made possible due to post translational modification of proteins in different tissues. As a result, a single protein may be differentially modified by one's metabolism in two different body fluids. Such differential modification of proteins can result in the detectible presence of more than one form of the same protein. An example of this is blood serum where 420 proteins resolved on 2D gels were identified by Mass spectrometry and found to correspond to only 150 unique protein sequences. The ability of antibodies to differentiate among such post-translationally modified proteins has valuable potential for differentiating among body fluids even if they have numerous proteins in common.

Statement of Hypotheses and Core Research Objectives

The central goal of the current research project was to accurately and reliably isolate protein biomarkers that are highly specific to individual of biological stains of forensic utility (*i.e.*, saliva, semen, peripheral blood, menstrual blood, vaginal secretions, and urine). This can complement the use of DNA profiling by making it possible to more accurately and confidently associate a DNA sample with a specific type of biological stain. The lack of such biomarkers can present forensic serologists with a significant challenge in many criminal cases.

Core Hypotheses – The successful achievement of this goal rested upon four major hypotheses. Specifically it was hypothesized that:

1) sufficient differences exist in the proteomes of individual body fluids so as to allow for the identification of individual body fluids with a high degree of specificity – ideally to the exclusion of all other body fluids.
2) multidimensional HPLC fractionation of body fluid proteomes would enable the rapid and reproducible generation of comprehensive mapping of individual body fluid proteomes.

3) bioinformatics applications would enable the rapid identification of high-specificity candidate biomarkers of individual biological fluids through the rigorous comparisons of protein elution coordinates among all mapped proteomes.

4) Sufficient similarities exist across human populations that proteins specific to a given body fluid would be expressed in most if not all humans; thereby ensuring the broad applicability of stain identification assays based on the use of protein biomarkers.

Major Research Objectives – This research project sought to apply a comparative proteomics approach to six forensically significant body fluids to improve the tools for stain identification through the completion of three major research objectives. These are to:

1) Collect samples of six forensically relevant body fluids (*i.e.*, peripheral and menstrual blood, semen, saliva, vaginal secretions and urine) representing a minimum of 5 individuals each in accordance with established National Institutes of Health protocols.

2) Perform high-resolution fractionation and quantitative mapping of the complete proteomic profile for each body fluid by 2-Dimensional High-Performance Liquid Chromatography (2-D HPLC).

3) Identify unique protein markers for each of the body fluids by rigorously conducting quantitative pair-wise comparisons of the proteomic profiles obtained for each body fluid and from each individual donor.

The successful completion of these objectives would aid forensic analysts by identifying antigenic markers which are specific to each of the forensically relevant body fluids. This would facilitate the development of a multiplex high-throughput assay capable of accurately characterizing the biological stains from which a given DNA profile has been/could be determined.

Methods

Human Subjects – The University of Denver (DU) Institution review Board for Research Involving Human Subjects (IRB) reviews all research involving human subjects, regardless of funding source, to ascertain that the rights and welfare of subjects are being protected. The IRB is responsible for assuring that recruitment advertising is not misleading or coercive to the research subject. All projects using human subjects are reviewed no less than annually.

All research conducted under DNA Research and Development Award 2006-DN-BX-K001 was IRB reviewed, approved and conducted in full compliance with U.S. Federal Policy for the Protection of Human Subjects (Basic DHHS Policy for Protection of Human Research Subjects; 56 FR 28003). A total of 30 adult (>18 y.o.) human volunteers (12 males; 18 females) were recruited for this study from within the DU student population. The purpose and significance of the research and the methods that would be used to collect body fluid samples was thoroughly explained to each volunteer. All participants then signed a statement of informed consent to participate in the research. Recruitment notices were posted in campus science buildings to attract interested volunteers. The student traffic in these buildings consists primarily of science-oriented graduate and undergraduate students. As no health care associated information was collected, HIPPA authorization was not required.

Body Fluid Collection and Protein Extraction – Samples of six forensically-relevant body fluids (*i.e.*, peripheral and menstrual blood, semen, saliva, vaginal secretions and urine) were collected for proteome mapping. The choice of these specific body fluids was based on discussions with forensic serologists at the Colorado Bureau of Investigation. The procedures employed for sample collection were in accordance with the NIH guidelines.

Salvia: The donor was directed to gently brush their teeth and thoroughly rinse their mouth with sterile water to remove residual food particles. After 5 minutes to allow secretion of saliva, the donor was instructed to place a Sarstedt Salivette™ saliva collection sponge into their mouth and to gently chew and roll the sponge around in their mouth for 3-4 minutes. The sponge was then placed into a sterile plastic conical tube. This allowed for the collection of large quantities of relatively pure saliva while reducing protein contamination from food items. Salivette™ sponges were centrifuged for 2 min at 1500 x g at 4°C to recover saliva which was transferred to 15 ml conical vial and centrifuged again at 13,000 x g for 20 minutes at 4°C. Supernatant-containing proteins were filtered through a .45 µm filter to remove remaining debris prior to concentration.

Seminal Fluid: Donors were directed to refrain from sexual activity for a minimum of 24 hours and then to obtain a 3-6ml sample of seminal fluid by masturbation in the privacy of their home. The subject was requested 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 at the donor's earliest convenience (within 1 hour). Semen was then incubated at room temperature for at least 30 minutes to allow it to liquefy. After transfer to a 15 ml conical vial and dilution with 1/3 volume PBS, the sample was centrifuge at 13,000 x g for 20 minutes at 4°C to pellet spermatozoa. The protein-rich supernatant was then passed through .45 µm filter to ensure cellular removal.

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. Blood serum was removed to a 15 ml conical vial and then passed through a .45 µm filter to remove cellular material prior to immunodepletion and protein concentration.

Urine: Donors were directed 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 concentrated to ensure a sufficient quantity of protein for proteome mapping. After transfer to 50 ml conical vials, the urine was centrifuged at 13,000 x g for 20 minutes at 4°C and passed through a .45 µm filter to ensure cellular removal prior to concentration.

Vaginal Secretions: Following clinically accepted procedures, vaginal secretions were self collected by study participants in the privacy of their home. Donors were directed 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 at their earliest convenience (within 1 hour). Subjects were financially compensated for their participation.

Tampons were saturated with PBS and allowed to sit at room temperature for 30 min with occasional vortexing to elute proteins. The tampon was then placed in a 50cc syringe to force out the fluids and eluted proteins. The liquid was transferred to 50 ml conical vials and centrifuge at

13,000 x g for 20 minutes at 4°C. The resulting supernatant was passed through a .45 µm filter to ensure cellular removal prior to concentration.

Menstrual Blood: Following clinically accepted procedures, menstrual blood was self collected by study participants in the privacy of their home. The collection protocol employed an FDA-approved over-the-counter latex-free, hypoallergenic cup (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 1 hour. The cup was then gently removed; the contents were poured into a sterile 50ml conical tube and refrigerated until delivered to lab (within 1 hour). Subjects were financially compensated for their participation.

Blood serum was removed to a 15 ml conical vial and then passed through a .45 µm filter to remove cellular material prior to hemoglobin removal, immunodepletion and protein concentration.

Protein Concentration, Partitioning and Quantification – Corning Spin-X UF concentrators (3000 NMWL) (Corning, Lowell, MA) was used to concentrate low protein content body fluids such as saliva and urine while at the same time removing unwanted salts and other low molecular weight components.

Serum obtained from menstrual blood samples was typically contaminated with erythrocyte cellular components due to the lysing of fragile red blood cells that are abundant in the endometrial lining during menses. As hemoglobin comprises 32-36% of all the proteins found in red blood cells the serum from menstrual blood samples contained large quantities of hemoglobin which served to mask the detection of less abundant menstrual blood specific proteins. For this reason, hemoglobin was removed from collected serum prior to proteome fractionation through use of HemogloBind™ (Biotech Support Group, Monmouth Junction, NJ). This hemoglobin capture reagent is a solid-phase, non-ionic adsorbent product that binds specifically to hemoglobin allowing for the removal of 80-90% of hemoglobin from serum or red cell lysates. HemogloBind™ does not cross react with most common serum components, making it suitable for the proteomic applications of this research project.

Blood plasma presented as an extremely complex mixture of blood proteins and as well as proteins from tissue secretion or leakage into the circulatory system. While this abundance of different proteins bodes well for the primary project objective of identifying protein markers that can be used to differentiate between venous and menstrual blood, these fluids were more challenging to process. This is because the dynamic range of protein concentration in blood spanned more than ten orders of magnitude and because peripheral and menstrual blood are characterized by the presence of several high-abundance serum proteins common to both fluids. The presence of these high-abundance proteins make it difficult to accurately map those proteins that were less abundant but which were more likely be the most specific markers of each body fluid. To circumvent this problem, commercially available IgY-12 Proteome Partitioning columns were employed. These antibody-based columns made it possible to remove twelve highly abundant proteins 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.

The Thermo Scientific Pierce Micro BCA Protein Assay (Thermo Fisher Scientific, Rockford, IL) was used to determine final protein concentration of each extracted sample. All samples were stored in a locked -70°C freezer until analyzed.

Proteome Mapping – The proteome of each body fluid sample was fractionated and mapped by 2-Dimensional HPLC. This employed a commercial ProteomeLab™ PF2D System (Beckman Coulter, Fullerton, CA) specifically designed for comparative proteomics research.

A high-performance chromatofocusing column was used for the first dimension to fractionate the complex protein mixtures present in body fluid lysates based on the isoelectric point of the constituent proteins. An in-line pH meter controlled the output of the eluent to a 96-well plate in 0.1 pH increments from pH 8.5 to pH 4.0 which is the range across which nearly all proteins can be eluted. The ProteomeLab™ PF2D system can accommodate up to 5 mg of total protein with a maximum injection volume of 5 ml. The first dimension chromatofocusing (HPCF) column is initially equilibrated with 30 column volumes of start buffer at pH 8.5 at a flow rate of 0.2 ml/min for 130 minutes. Following equilibration the sample was injected into the HPCF module followed by 20 min of start buffer at 0.2 ml/min. At 20 minutes, eluent buffer (pH 4) was run at 0.2 ml/min for 115 min to create a pH gradient with fractions collected at 0.1 pH intervals and stored in chilled autosampler. At 115 min 1 M NaCl was run as a high ionic strength salt wash.

Following completion of HPCF fractionation each fraction collected from the first dimension were automatically injected into a high-performance non-porous silica reverse-phase (HPRP) column where proteins are further separated on the basis of hydrophobicity. This second dimension HPRP column was initially flushed with 5 column volumes of 0.08% TFA in acetonitrile followed by 10 column volumes of 0.1% TFA in H2O 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% of 0.08% TFA in acetonitrile gradient was performed over 30 minutes creating a 3.33% change in solvent/minute. At 0.5 min intervals, fractions were collected with a Gilson FC-204 fraction collector in a series of twenty or more 96-well plates. The collected fractions (approximately 400), containing intact proteins, were stored frozen in a locked -70°C freezer until required for further characterization (*e.g.*, by mass spectrometry).

The reproducibility of the system helped to ensures that different proteome maps could be readily and quantitatively compared to identify proteins whose expression was unique or quantitatively altered in a given sample. The ProteoVue™ software application was then used to generate preliminary high-resolution proteome data files and pI/ hydrophobicity expression maps that were color coded to facilitate data interpretation. These proteome maps were similar in format to proteome mapping by 2D gel electrophoresis.

Comparative Proteome Analysis – Due to the technical limitations associated with the commercial DeltaVue™ software package for interproteome map comparison, an in-house bioinformatics application (ProteinMiner™) was developed to manage and analyze the average of 500,000+ data points that comprised each proteome map. This custom application provided a robust means of comparing different proteomes while taking into account subtle run-to-run and interindividual differences between sample donors.

ProteinMiner™, which is based on a combination of C++, Perl, and Matlab, was used to address three aspects of data analysis. These were: 1) porting datasets from the ProteomeLab™ PF2D System to consistent and useful formats capable of making tab/comma delimited files; 2) data visualization and graphical manipulation to facilitate detailed visual analysis aimed at detecting possible forensic protein targets and; 3) “number crunching” to combine and average data from individual samples to create a single “consensus map” for a given body fluid and then to compare consensus maps across body fluid proteomes. The specific functionalities of

ProteinMiner™ software application are addressed in greater detail in the "Results and Discussion" section of this report.

Protein Identification – Proteins that were of interest as potential biomarkers were identified by mass spectrometry. Eluted fractions from the ProteomeLab™ PF2D System were transferred to 1.5 ml low retention microcentrifuge tubes and lyophilized in a vacuum evaporator. Dried protein samples were reconstituted in 40ul of 100 mM Tris-HCl pH 8.5, 1.2ul of 100 mM TCEP reducing agent and then shaken for 20 minutes at room temperature. Then 0.88ul of 500 mM IAA was added and the sample was shaken in the dark for an additional 15 minutes to alkylate the proteins. The proteins were digested with trypsin for 14-16 hours at 37°C. Samples were sonicated and digested with a second equal volume of trypsin for an additional 8-10 hours at 37°C. Digested samples were then purified on a C-18 spin column, dried and resuspended suspend in 3% acetonitrile and 0.1% formic acid.

Mass spectrometry was performed on an Agilent 6300 series ion trap mass spectrometer coupled to a 1200 series HPLC-Chip/MS system (Protein ID "short chip" specifications 43mm 300 A C18 chip) using 0.1 to 1.5ul of digested sample per injection. Columns were equilibrated in 0.1% Formic acid in water. At 2 min a 0-45% of 0.1% formic acid in acetonitrile gradient was performed over 22 minutes followed by a 3 minute column re-equilibration. Data analysis was performed using Spectrum Mill software suite by Agilent Technologies. The Swiss-Prot database was used to match MS/MS spectrum generated on mass spectrometer. Typically proteins identified with 2+ peptides and scores >16 were considered confident matches.

Results and Discussion

Proteome mapping – A minimum of five individuals was recruited to donate samples of each body fluid being analyzed. This redundancy was intended to help to discriminate between proteins that are characteristic of a specific body fluid and thus suitable for use in development of body fluid ID assays versus those that might reflect inter-individual variability in protein expression - and thus not be suitable as biomarkers. The analysis of samples of each body fluid from a minimum of five individuals reflected an effort to obtain the maximum amount of comparative proteomic information within a reasonable budget and timeframe. Although the value of mapping the proteomes from a larger number of study participants as a means of assessing proteomic variation within and among different human populations was recognized, the primary objective of this research project was to find highly-specific candidate biomarkers of body fluids that could subsequently be further validated for use across multiple populations. A comparative analysis of five proteomic profiles for each body fluid should yield sufficient data to achieve this objective. A rigorous, full-scale forensic validation of the body fluid markers identified thorough the proposed research, however, would necessarily involve a larger study population.

Fractionating and recovering proteins from complex mixtures for downstream identification and analysis is fundamental to the field of proteomics[23]. Traditionally, these tasks have been handled primarily by 2D gel electrophoresis (2DGE) and manual excision/purification of proteins of potential interest. While these methods have a long history, they have several critical shortcomings[24]. 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 requires that numerous gels be run to obtain data sets that can be reliably compared with each other.

Fluorescence-based difference gel electrophoresis has improved comparative protein quantization over traditional 2DGE, but this method is more labor intensive and does not appreciably expand the variety of proteins that can be analyzed.

More recently, researchers have used multi-dimensional chromatographic platforms to circumvent the limitations of 2DGE[25]. This approach to proteomic profiling offers several key technical advantages for the purposes of the proposed research. First, the use of liquid phase separation results avoids the solubility problems associated with gels. Second, a more complete profile of each body fluid can be generated because virtually all proteins present are fractionated and recovered. Third, less abundant proteins in complex mixtures can be screened as potential markers of specific body fluids because there was a higher efficiency of recovery (>95%) and more total protein could be injected (50µg-30mg) without the band distortion that occurs with 2DGE. The ProteomeLab™ PF2D System (Beckman, Fullerton, CA) was employed for the current research project. This is a commercial, high-resolution 2-D HPLC platform with optimized chemistries and software specifically designed for proteomics. This system made it possible to precisely fractionate and analyze complex body fluid proteomes while avoiding the limitations of 2DGE.

Individual comprehensive proteome maps were generated for each donated body fluid sample that was fractionated. In each of these maps, protein "pI" data from the chromatofocusing column (as a first dimension) is combined with ultra-high resolution hydrophobicity data from non-porous silica reverse phase column (as a second dimension) to yield a .dat file containing a detailed 2D pI/hydrophobicity map in that can be visualized by the ProteoVue™ software suite and then saved as a .vue file. These files were then ported as tab/comma delimited files to an in-house bioinformatics application for consensus map building and difference comparisons. Examples of the 2D pI/hydrophobicity maps for peripheral blood, urine, semen and saliva are shown in figure 1. The intensity and color of the bands represent the abundance of the protein detected. Red, orange and yellow bands represent more abundant proteins and while green and blue represent less abundant proteins. Additionally, a high resolution UV-based hydrophobicity elution chromatogram for each second dimension fraction (x-axis) is shown to the left of each proteome map. At the scaling shown in figure 1, only the most abundant proteins in each proteome are visually represented. By rescaling the maps proteins of moderate and lower abundance become readily identifiable. In this way it was possible to identify hundreds of individual proteins in each proteome.

Because of the excellent reproducibility with which different protein samples can be fractionated even visual comparisons of the resulting 2D pI/hydrophobicity maps from different body fluids can reveal the presence of proteins unique to each target body fluid. This is illustrated in figure 2 where a proteome map for a menstrual blood sample is shown next to a proteome map for a peripheral blood sample. There are a number of abundant proteins (circled) present in menstrual blood which appear to be absent from peripheral blood. The proteins represented by these bands could reasonably be considered to be candidates for identification and further investigation as potentially useful biomarkers of menstrual blood if their presence in menstrual blood and their corresponding absence from peripheral blood (and all other body fluids) were confirmed in the proteome maps of the remaining sample donors.

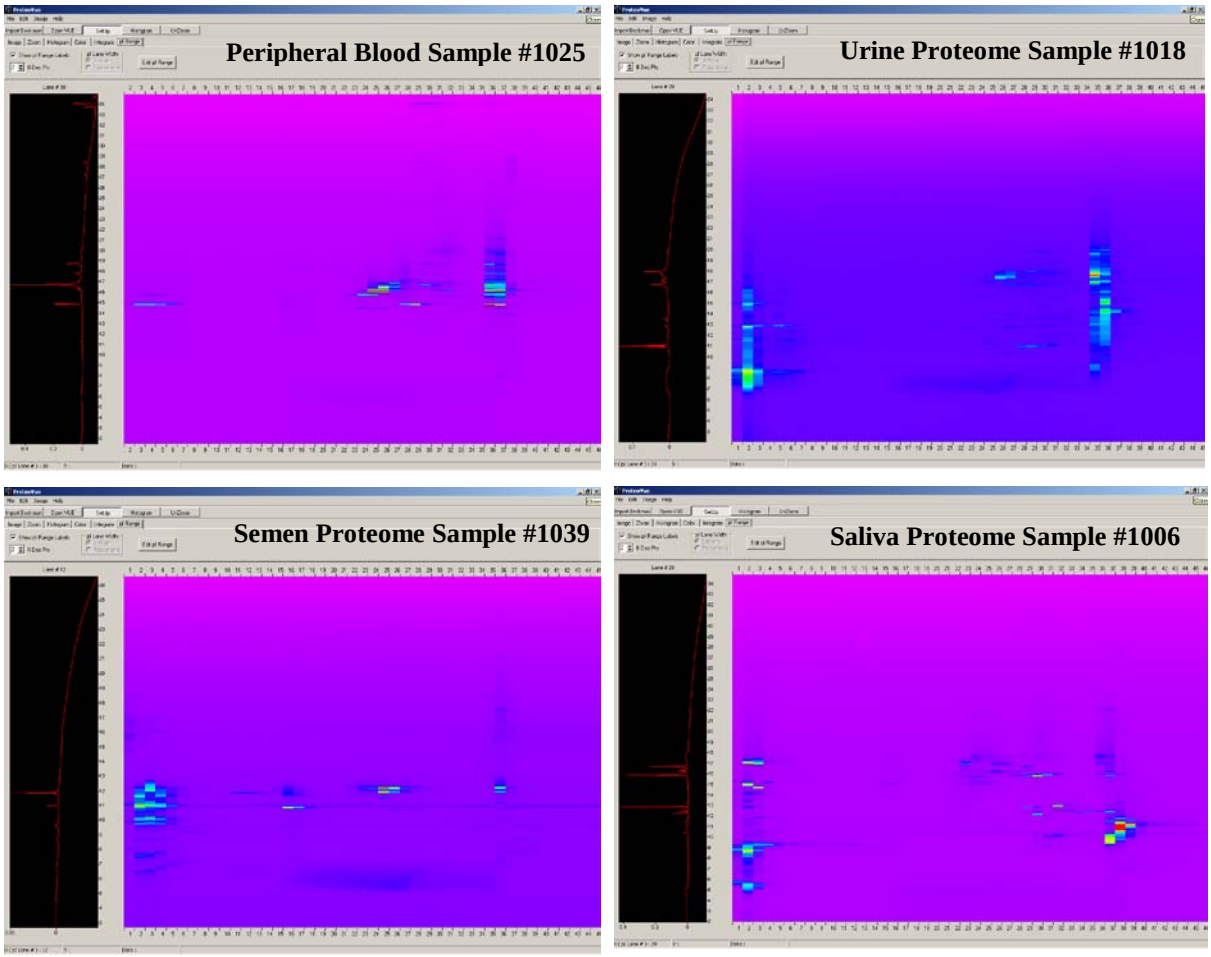

Figure 1: Examples of the 2D pI/hydrophobicity maps (*i.e.*, proteome maps) obtained from 4 different body fluid samples (peripheral blood, urine, semen and saliva). Differences in bands are indicative of potential protein markers of interest.

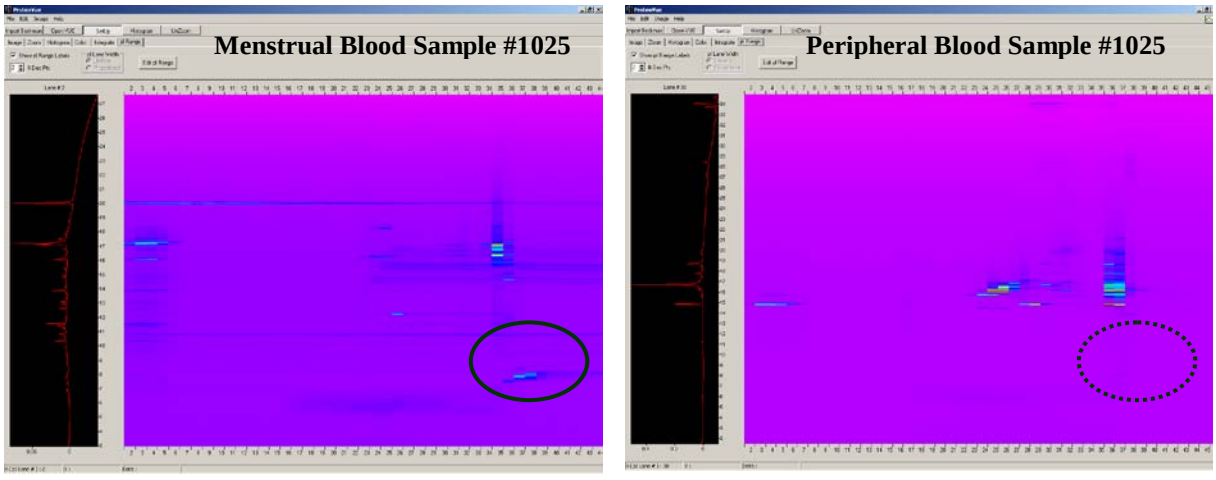

Figure 2: Comparison of 2D pI/hydrophobicity maps (*i.e.*, proteome maps) obtained from peripheral and menstrual blood from the same human subject. The bands encircled by the oval in the menstrual blood map (left side) represent proteins that are present in menstrual blood but which do not appear to be present in peripheral blood (dashed oval).

Critical to minimizing assay-to-assay variability in multi-dimensional HPLC proteome fractionation included the need to standardize the amount of total protein from each extract that was loaded onto the first dimension HPLC column at 3mg. This necessitated an efficient means of concentrating body fluids containing low concentrations of protein (*e.g.*, saliva and urine). It was also necessary to minimize potential protein input from non-human sources such as food particles and to ensure that abundant non-specific proteins such as albumin, immunoglobulins and hemoglobin in peripheral and menstrual blood were not allowed to mask the detection of less abundant proteins with potential utility as biomarkers. Attention to these details as outlined in the methods section of this report helped to ensure that differences in protein markers between individual body fluids reflected actual differences rather than artifacts arising from vicissitudes associated with the preparation of the protein extract.

Proteome Map Difference Comparison – Rigorous quantitative pair-wise comparisons of the 2D pI/hydrophobicity proteome maps obtained for each body fluid was critical to identifying the most promising proteins to be investigated as candidate high-specificity biomarkers of individual body fluids. The reproducibility with which the proteomes of each body fluid were fractionated by 2D HPLC helped to ensure that proteomic profiles generated at different times and from different samples could be quantitatively compared to each other. The commercial DeltaVue™ software suite which was integrated into the ProteomeLab™ PF2D System generate color-coded "differential display" maps that highlighted differences between the proteomic profiles of any two samples.

An important concern that arises with any comparative approach involving proteins is minimizing the number candidate biomarkers that are incorrectly identified as being uniquely expressed in a given type of sample. These "false positive" candidate markers may reflect interindividual differences in expression or the expression of low abundance proteins. In the later case, the "false positive" markers arise not because they are uniquely expressed in one sample versus another but rather because they are expressed in one sample at a detectable level and in another sample at a level that is nearer the threshold of detection. As a strategy to minimize the number of such "false positive" candidate markers multiple proteome maps were generated representing the samples provided by the five different volunteer donors for each body fluid. It had been expected that multiple quantitative pair-wise difference comparisons between the proteome maps of the individual donating the same body fluid would minimize the occurrence of false positives resulting from interindividual differences in protein expression. Similarly, it was expected that multiple quantitative pair-wise difference comparisons between different body fluids (*e.g.*, five proteome profiles of vaginal secretions compared to five proteome profiles of saliva = 25 pair-wise difference comparisons) would help to identify only the most promising candidate biomarkers. This was based on the fact that a candidate protein marker would need to be identified as a uniquely expressed protein in all 25 pair-wise difference comparisons for each of the six body fluids being examined.

These interproteome difference assays to identify potential candidate biomarkers were initially handled by the ProteomeLab™ PF2D System's DeltaVue™ software package. Following several pair-wise comparisons, however, a number of unexpected data analysis limitations were encountered. For example, when using the DeltaVue™ application, much of the finer resolution of the proteome map was lost as illustrated in figure 3. Furthermore, an inability to normalize for subtle retention time differences and pH variances between samples results in a

decrease in the confidence with which prospective biomarkers are identified (figure 3). As efficient biomarker discovery may hinge on identifying subtle differences between similar proteomes, a decrease in resolution impairs one's ability to locate unique protein markers. Recognizing the limitations in the current software, a custom software solution was developed to: 1) allow the difference comparison of multiple proteome maps at once without a loss of resolution; 2) facilitate normalization of pH fraction boundaries and reverse-phase retention times to account for subtle retention and pH variances between proteome maps; 3) take differences in protein expression levels into consideration and; 4) create a single "consensus proteome" for each body fluid which encompasses interindividual and interassay variations.

Figure 3: DeltaVue™ difference profile comparing urine (left) with saliva (right). Loss of resolution of lower abundance proteins and the inabilty to normalize 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.

Dataset Optimization: The dataset from the ProteomeLab™ PF2D System for each comprehensive proteome map consisted of a 45 x 10,501 data point matrix. A large portion of the >450,000 data points, however, represent uninformative background noise created by the buffers rather than actual proteins. Carrying these unnecessary data through analysis algorithms to identify protein differences between proteome maps results in seriously compromised computational performance. There is simply too much data being processed for most desktop computers to be able to handle. To circumvent this problem, a protein peak (*i.e.*, proteome map coordinates) identification algorithm was implemented to create a normalized representative dataset for each proteome (Figure 4). 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 can be eliminated while increasing the resolution and the speed of additional downstream analyses.

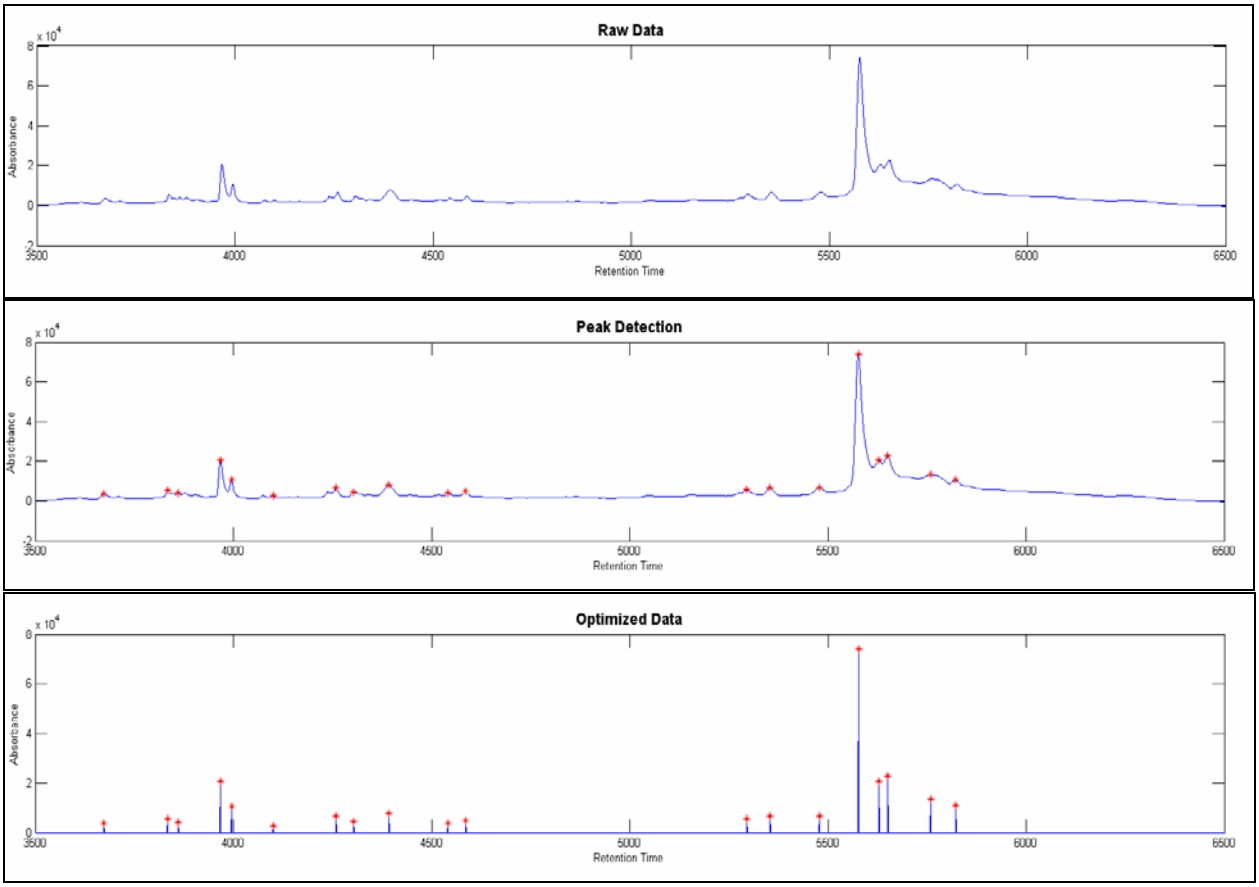

Figure 4 – Application of a protein peak identification algorithm to raw hydrophobicity data (top) generated from a single first-dimension ProteomeLab™ PF2D fraction. Protein peaks are called based on the point at which the slope changes from positive to negative or the point where a derivative changes signs (middle). Once identified, called protein peaks are used to create a normalized representative dataset from which > 99.9% of the underlying noise has been eliminated (bottom).

Dataset Merging and Consensus Map Creation: It was also necessary to creating a reliable means of comparing different proteomes while taking subtle differences between sample runs and individuals. 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 easily compare one body fluid to another. The specific algorithm that was implemented is known as the k-means clustering - an algorithm that organizes a data set into k subsets. The algorithm involves a four step procedure (figure 5). 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 consensus of a set of individual protein points from multiple proteome maps. The overall product of this procedure was a disjoint set of points split into k partitions. Meaning that all of the proteins from each fluid were grouped together in three dimensional space with a single central point (figure 6).

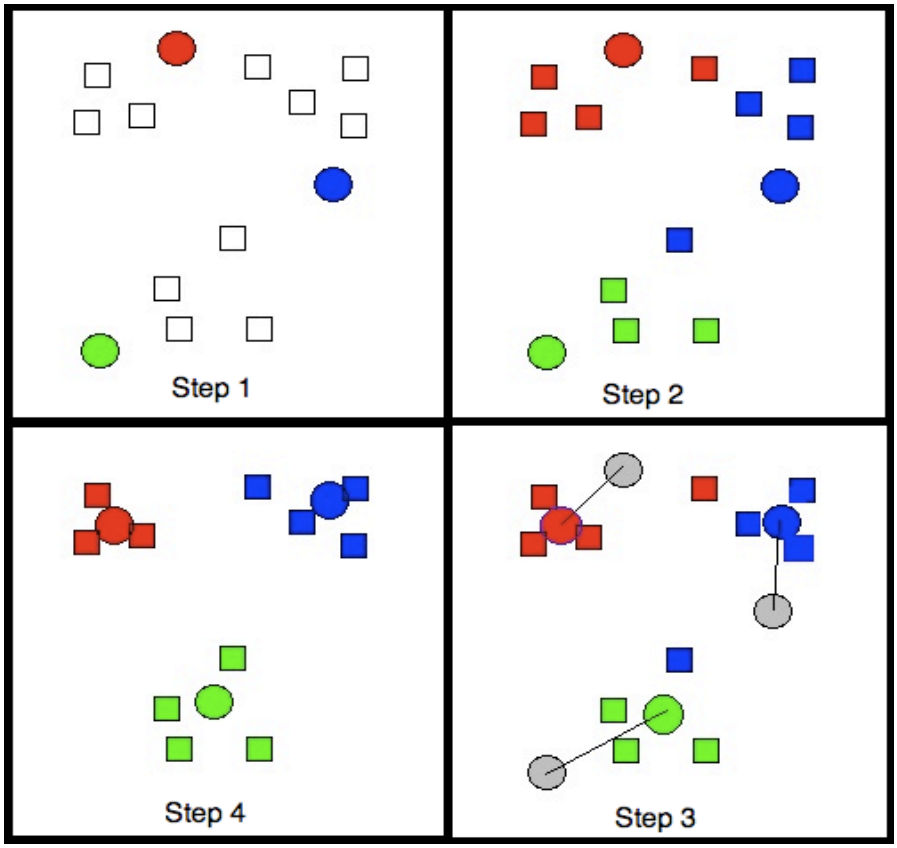

Figure 5 - Pictorial representation of k-means clustering algorithm. Step 1: circles representing centroids are randomly placed onto the map. Step 2: squares representing data points are assigned to the nearest centroid. Step 3: Centroid positioning is recalculated to minimize the variance between data points. Step 4: Steps 2 and 3 are repeated and data points are reassigned until a stable solution is achieved.

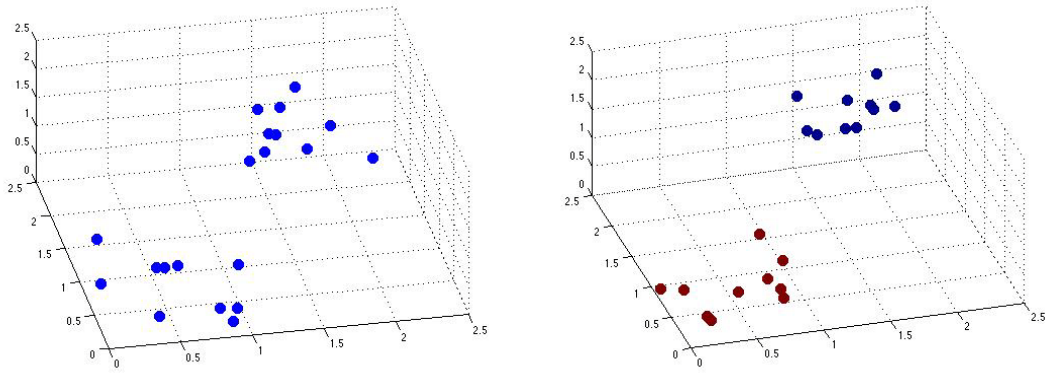

Figure 6 – Unclustered (left) versus clustered (right) groups of peaks. Each point representing a peak from an individual map. In this example, 20 points have been split into k = 2 partitions as indicted by the brown vs. blue coloring of the clustered data points.

Applying the K-means algorithm to a series of proteome maps where there may collectively be 10,000 or more data points presented a more complicated scenario. This was due to the difficulty of determining how many partitions k should be used and where they should be initially located. To account for subtle inter-individual variability in these cases a hierarchical clustering method was implemented to extract only the significant features from a set of proteomes. Termed *Protein Miner*™, this software application made it possible to scan through a series of proteome maps of the same body fluid from different individuals to create a consensus map for that fluid (figure 7**)**. The consensus maps contain significant protein peaks (proteome map coordinates) common across all individuals while eliminating peaks that were likely to reflect interindividual variations. The overall product of this procedure was a set of consensus peaks derived from a set of proteomes. Using these approaches a list of the proteome map coordinates of unique candidate biomarkers for subsequent protein identification by ESI-MS/MS was generated as illustrated in figure 8.

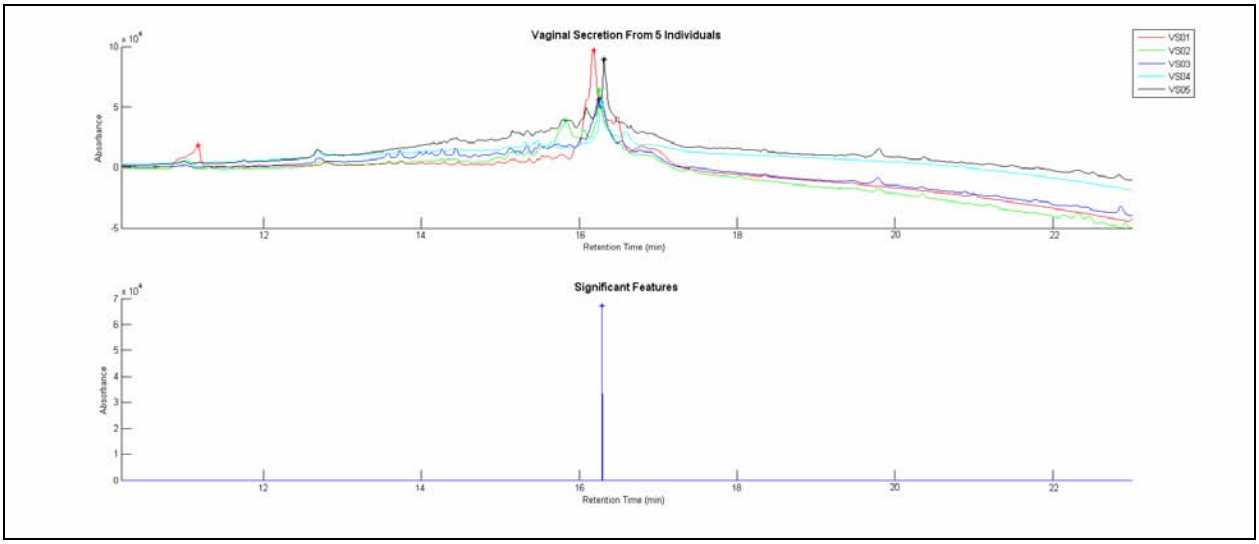

Figure 7: Detail from the application of the dataset clustering algorithm applied to vaginal secretion fraction 35. The top chromatographic trace shows raw data while the bottom trace shows the analyzed consensus peak.

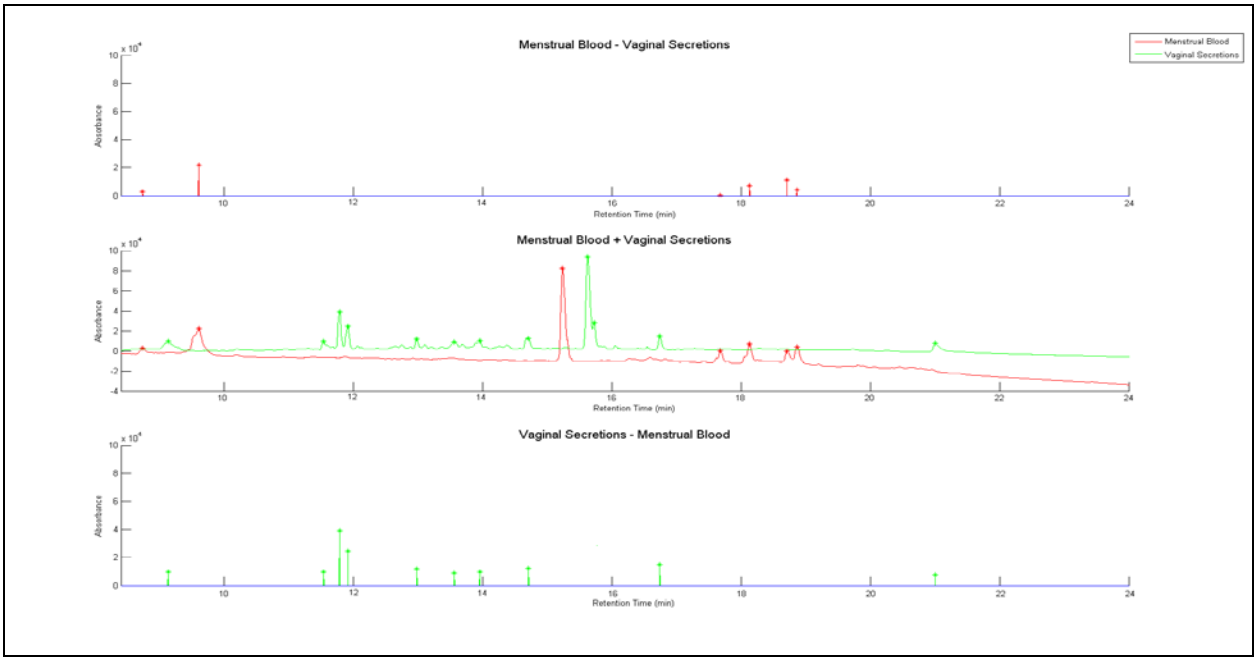

Figure 8: Identifying proteome map coordinates where highly-specific candidate protein biomarkers of individual body fluid biomarkers may be found. (Center Panel) By analyzing comparable ProteomeLab™ PF2D System based datasets for menstrual blood (red) and vaginal secretions (green), it is possible to identifying those proteome map coordinates where proteins that appear to be unique to menstrual blood (top panel) and vaginal secretions (bottom panel) may be found.

Identification of High-Specificity Candidate Biomarkers of Individual Body Fluids – The translation of identified proteome map coordinates output by *Protein Miner*™, into actual candidate protein biomarkers with potential utility for forensic practitioners was a relatively straight forward process. The basic workflow of this approach began with the retrieval of those ProteomeLab™ PF2D second dimension fractions identified as containing potentially unique proteins for a given body fluid. The proteins contained in these fractions were denatured and trypsin digested in preparation for peptide fingerprinting and identification by mass spectrometry on an Agilent electrospray ion trap mass spectrometer coupled with an HPCL-Chip system. Database searches were performed using Agilent's spectrum mill search engine. Figures 9, 10 and 11 show examples of the results of an identification assay run on Spectrum Mill. This software is able to analyze each fingerprint and match/rebuild a known protein. The result is a table of protein matches complete with names, accession numbers, and sequence information. For confident protein identification two or more distinct peptides needed to be present with scores exceeding sixteen.

Group (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Mean Peptide Spectral Intensity	Database Accession #	Protein Name
1	3	3	40.93	45	2.13e+007	P02808	Statherin OS=Homo sapiens GN=STATH PE=1 SV=2

#	Filename	z	Score	Fwd-Rev Score	SPI (%)	Spectrum Intensity	Sequence	RT (min)	MH+ Matched (Da)
1	SLSA30C10.0550.0550.3	3	15.09		82.5	7.57e+006	(R)FGYGYGPYQPVPEQPLYPQPYQPQYQQY(T)	5.27	3395.568
2	SLSA30C10.0519.0524.0	3	13.86		85.6	5.06e+007	(R)FGYGYGPYQPVPEQPLYPQPYQPQ(Y)	4.97	2813.325
3	SLSA30C10.0409.0409.2	2	11.98		99.4	5.63e+006	(Y)QPVPEQPLYPQPY(Q)	3.92	1555.779

Figure 9: Identification by Spectrum Mill software of statherin, a candidate high-specificity protein biomarker of saliva predicted by Protein Miner™.

Group (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Total Protein Spectral Intensity	Protein MW (Da)	Protein pI	Species	Database Accession #	Protein Name
1	28	13	206.17	29	1.98e+009	52131.1	9.30	HUMAN	P04279	Semenogelin-1 precursor - Homo sapiens (Human)
1	6	4	65.19	6	3.62e+008	65444.7	9.08	HUMAN	Q02383	Semenogelin-2 precursor - Homo sapiens (Human)
Totals:	34	17								

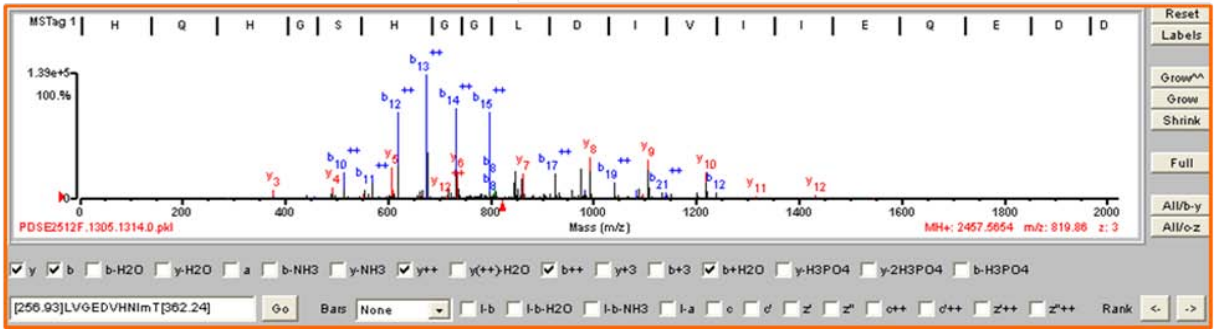

Figure 10: (Top) Identification by Spectrum Mill software of semenogelin-1 and -2, candidate high-specificity biomarkers for semen predicted by Protein Miner™. MS/MS scores of >16 with two or more distinct peptides are considered confident. (Bottom) Mass spectrometry scan of ProteomeLab™ PF2D System second dimension fractions identified as containing potentially unique proteins in human semen.

KMLSEMEN1TEST # spectra mean intensity	Database Accession #	%AA Coverage	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	Group #	Protein Name
5 2.04e+007	P02768	51	25	438.85	1.1	2 X Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2
5 2.21e+007	P02788	37	24	434.16	2.1	Lactotransferrin OS=Homo sapiens GN=LTF PE=1 SV=6
35 8.18e+007	P04279	53	26	432.97	3.1	4 X Semenogelin-1 OS=Homo sapiens GN=SEMG1 PE=1 SV=2
1 9.84e+007	P07288	72	10	195.70	4.1	Prostate-specific antigen OS=Homo sapiens GN=KLK3 PE=1 SV=2
1 1.47e+007	P10909	25	9	155.84	5.1	Clusterin OS=Homo sapiens GN=CLU PE=1 SV=1
2 2.18e+007	P15309	29	8	135.62	6.1	Prostatic acid phosphatase OS=Homo sapiens GN=ACPP PE=1 SV=3
0 0.00e+000	P02751	6	8	127.84	7.1	Fibronectin OS=Homo sapiens GN=FN1 PE=1 SV=3
1 2.29e+007	P12273	53	7	113.85	8.1	Prolactin-inducible protein OS=Homo sapiens GN=PIP PE=1 SV=1
1 1.63e+007	P61916	32	3	55.58	9.1	2 X Epididymal secretory protein E1 OS=Homo sapiens GN=NPC2 PE=1 SV=1

Figure 11: Identification by Spectrum Mill software of multiple candidate protein biomarkers of semen (*e.g.*, semenogelin, prostate specific antigen, prostatic acid phosphitase and Epididymal secretory protein E1). All proteins were identified by analysis of ProteomeLab™ PF2D System second dimension fractions identified as containing proteins that were potentially unique to semen.

The overriding core objective of the research funded under DNA Research and Development Award 2006-DN-BX-K001 was to compile a comprehensive panel of highly-specific candidate protein biomarkers for each of six specific body fluids for subsequent forensic validation. Having successfully generated comprehensive multi-dimensional proteome profiles comprising virtually every protein in peripheral and menstrual blood, vaginal secretions, semen, urine, and saliva, a defined subset of proteins that appear to be unique to each of these body fluids has been identified. These are presented in Table 1.

The results obtained to date are extremely promising. For example, the first protein (statherin) identified by the *Protein Miner*™ software as being highly-specific for saliva was selected independently of any information other than the k-means clustered proteomes. We are particularly encouraged by these results because statherin has been independently identified as a possible saliva marker by gene expression database searches and by forensic researchers working on the development of mRNA markers for saliva[14, 26]. Similarly, our identification of semenogelin-1 and -2 as markers of seminal fluid[27] and periplakin as marker of vaginal secretions are consistent with what has been reported by biomedical researchers[28]. The apparent accuracy with which our comparative proteomic approach has been able to identify these markers bodes well for the likely specificity of numerous other candidate biomarkers that we have identified - but for which information on tissue specificity in the literature is lacking.

Priority for inclusion in the list of candidate biomarkers was based primarily on peak uniqueness and secondarily on protein abundance. The inclusion of a specific protein in the table of candidate markers also employed two quantitative criteria: the relative abundance of the biomarker-containing peak from the ProteomeLab™ PF2D chromatogram and; the overall quality of the subsequent ms/ms data. With regard to the ProteomeLab™ PF2D, the limiting factor was the threshold value (*i.e.*, peak height) needed for quality ms/ms data. Based on findings from the current study and the manufacturer's standard operating procedure for the ProteomeLab™ PF2D, a peak of >0.2 AU was needed for quality ms/ms identification. This being said, a very promising saliva candidate (statherin), had a peak height of only 0.17 AU.

An indication of the relative abundance of each candidate biomarker was derived from the ProteomeLab™ PF2D peak integration feature for quantitation. For example, semenogelin was a highly abundant protein whereas statherin was on the lower abundance end of the biomarker panel. In fact, statherin, approached the detection limits of the ProteomeLab™ PF2D. Semenogelin, by contrast, encompassed 3 peaks across 3 pH fractions with heights of 0.797 AU, 0.919 AU, and 1.40 AU and area percents of 11.53%, 14.35%, and 18.28%, respectively. Semenogelin accounted for an average of 44.16% of the total protein content in semen while statherin with a chromatographic peak height of 0.17 AU accounted for only 0.78% of the total saliva protein content.

In all, there were >1000 proteins identified in the course of comparative proteome mapping. This included candidate proteins identified by: (1) ms/ms analysis of peaks identified by our comparative mapping software as unique; (2) ms/ms analysis of pH fractions and; (3) ms/ms data on unfractionated body fluid samples. All of these approaches identified proteins might be considered "putative" candidates. However, data analysis by an experienced mass spectroscopist was able to eliminate many candidates as being redundant and/or non-unique. Further analysis of each possible candidate based on information in swissprot/uniprot/ncbi and the profession literature made it possible to arrive at a reasonably accurate listing of high quality candidate biomarkers.

Table 1: Candidate Protein Biomarkers with Potential Utility for Body Fluid Identification

Fluid	Protein	Accession	Function
Semen	Semenogelin 1	P04279	They participate in the formation of a gel matrix entrapping the accessory gland secretions and spermatozoa.
	Semenogelin 2	Q02383	
	Epididymal secretory protein E1	P61916	May be involved in the regulation of the lipid composition of sperm membranes during maturation in the epididymis.
	Dual specificity testis-specific protein kinase 2	Q96S53	Phosphorylates cofilin at 'Ser-3'. May play a role in spermatogenesis.
	Prostatic Acid Phosphitase	P15309	Acid phosphitase produced in the prostate.
	G2/mitotic-specific cyclin-B3	Q8WWL7	Involved in cell cycle control.
Saliva	Statherin	P02808	Stabilizes saliva supersaturated with calcium salts by inhibiting the precipitation of calcium phosphate salts
	Salivary acidic proline-rich phosphoprotein	P02810	Inhibitor of calcium phosphate crystal growth. Provides a protective / reparative environment for dental enamel.
	Cystatin_SA	P09228	Thiol protease inhibitor.
	Cystatin_D	P28325	Inhibitor possibly involved in protection against oral cavity proteinases.
	Submaxillary gland androgen-regulated protein	P02814	Phosphoprotein
Vaginal Secretions	Extracellular matrix protein 1	Q16610	Negative regulator of bone mineralization. Stimulates endothelial cell proliferation and angiogenesis.
	Glycodelin	P09466	Main protein synthesized and secreted in the mid-luteal phase endometrium and during the first trimester of pregnancy.
	Matrigel-induced gene C4 protein	O95274	Supports cell migration. May be involved in urothelial cell-matrix interactions
	Secreted glypican-3	P51654	May be involved in the suppression/modulation of mesodermal tissue growth.
	Vimentin	P08670	Vimentins are class-III intermediate filaments in some non-epithelial cells
	Stratifin	P31947	Adapter protein implicated in the regulation of some signaling pathways.
	Involucrin	P07476	Part of the insoluble cornified cell envelope of some stratified squamous epithelia.
	Periplakin	O60437	May link the cornified envelope to desmosomes and intermediate filaments.
	Gelsolin	P06396	Calcium-regulated, actin-modulating protein that binds to the ends of actin monomers or filaments, preventing monomer exchange.

	Vinexin	O60504	Promotes up-regulation of actin stress fiber formation.
	Mesothelin	Q13421	Membrane-anchored forms may play a role in cellular adhesion.
Urine	Uromodulin	P07911	Unknown. May play a role in regulating the circulating activity of cytokines.
	Osteopontin	P10451	Possibly important to cell-matrix interaction.
Menstrual Blood	Pregnancy zone protein	P20742	Inhibitor of four classes of proteinases by a unique 'trapping' mechanism.
	Matrilysin	P09237	Degrades casein, gelatins of types I, III, IV, and V, and fibronectin.
	Calpastatin	P20810	Specific inhibition of calcium-dependent cysteine protease. Plays a key role in postmortem muscle degradation. May also be involved in degradation of living tissue.
	SH2B adapter protein 2	O14492	Adapter protein for several members of the tyrosine kinase receptor family. Involved in multiple signaling pathways.
Peripheral Blood	Hemopexin	P02790	Binds heme and transports it to the liver for breakdown and iron recovery.
	Histidine-rich glycoprotein	P04196	Physiological function not yet known. It binds heme, dyes and divalent metal ions.
	Apolipoprotein	P04114	Apolipoprotein B is a major protein constituent of chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100).
	Plasminogen	P00747	Plasmin dissolves the fibrin of blood clots and acts as a proteolytic factor in a variety of other processes including embryonic development, tissue remodeling, tumor invasion, and inflammation. It activates urokinase-type plasminogen activator, collagenases and several complement zymogens. It cleaves fibrin, fibronectin, thrombospondin, laminin and von Willebrand factor.
	Transthyretin	P02766	Thyroid hormone-binding protein. Probably transports thyroxine to the brain.
	Antithrombin-III	P01008	Most important serine protease inhibitor in plasma. Regulates blood coagulation.
	Ceruloplasmin	P00450	Ceruloplasmin is a copper-binding glycoprotein.
	Afamin	P43652	Possible role in the transport of a yet unknown ligand.
	Serum amyloid P-component	P02743	Can interact with DNA and histones and may scavenge nuclear material released from damaged circulating cells. May also function as a calcium-dependent lectin.

Implications for Policy and Practice

We have established excellent working relationships with forensic practitioners in the US and abroad which have been invaluable in productively guiding our R&D efforts. For the research reported here we have worked in collaboration with forensic analysts from the Colorado Bureau of Investigation and forensic research scientist from the Forensic Biology Group of New Zealand's Crown Institute for Environmental Science and Research. The advice of these collaborators played an important role in shaping many of the preliminary experiments. Our collaborators have repeatedly stressed that the identification of biological stains can still be a significant challenge for the forensic serologist. Commercial kits for the identification of blood, semen and saliva use proteins as diagnostic markers of these forensically important substances. While these protein markers have proven useful, they were selected at a time when the field of comparative proteomics was in its infancy. The comparative proteomic research reported here, however, has made it possible to obtain a far more complete proteomic map based on the hundreds to thousands of proteins present in many human body fluids. These proteins have been a rich source of information with enormous forensic utility.

After further and very rigorous forensic validation of the candidate protein biomarkers presented in this report, the most obvious commercial application may be the development highly-specific immunochromatographic assays. The utility and cost effectiveness of these assays, as exemplified by ABA card and Seratec® kits, is well established in the forensic community. The identification of dozens of additional protein markers could enable a single multiplexed approach to body fluid identification. It is even conceivable that a hand-held assay card could be designed that would be capable of analyzing and identifying multiple body fluids or mixtures of different body fluids, without having to perform multiple assays. The use of protein markers with high-sensitivity antibody-based assays also offers the potential for direct body fluid identification without the need for an amplification step. This can be important from an analyst's perspective because it saves time and minimizes sample handling. This could also significantly reduce the consumption of valuable evidence. Furthermore, by using protein markers for all body fluids, it would be possible to completely eliminate loss of valuable evidentiary material by processing swabs and other evidence to simultaneously extract both nucleic acids and proteins. There are a number of kits that now available from commercial suppliers (*e.g.*, Qiagen and Sigma) allow such separations.

Another potential area of impact would be the use of unique protein biomarkers of individual body fluids in the development of more sensitive or advanced next generation detection technologies. Where the ABA card and Seratec® kits represent ideals in terms of cost and speed, the same protein markers identified under this proposal would be equally suited to assay systems based on high-sensitivity ELISA tests or even future antibody/aptamer chip based assay systems.

It is important to emphasize, however, that these protein biomarkers were identified by mapping the protein profiles of just five individuals per bodily fluid and thus may only be considered candidate biomarkers. While the use of even a relatively small sample group can help to reduce the potentially misleading impact of interindividual differences in protein expression through the creation of "consensus proteome maps", the ultimate applicability of a given biomarker for use with the general population necessitates a more comprehensive and thorough validation of each candidate marker for stain specificity with a larger population set. There are good reasons for this. For example, the possibility cannot be ignored that some candidate

biomarkers might be secreted into non-target fluids in the same way that A, B, and Rh factors in blood are found in the saliva or semen of individuals termed secretors. Confounding factors such as this might be missed when looking at proteome data from only five individuals. Only when larger-scale studies are completed, can these markers move from being candidates to serving as the foundation for a commercial multiplex assay system capable of characterizing both single source and mixed-source stains with high specificity.

Implications for Further Research

All specific aims under award 2006-DN-BX-K001 have been successfully completed resulting in the identification of multiple candidate biomarkers of six body fluids of forensic significance. A thorough forensic validation of specificity of these candidate biomarkers in a larger population group, and using forensic casework type samples represents the next step toward the development of a practitioner-ready high-specificity test for biological stain identification. There are a number of approaches that could be used to accomplish these tasks. The traditional pipeline for biomarker development and commercialization begins with two objectives, discovery and validation. Significant progress has already been made and reported under award 2006-DN-BX-K001. In developing an accurate and efficient means of validating the specificity of numerous candidate biomarkers across multiple body fluids multiple approaches should be considered beginning with the use of antibodies and mass spectrometry.

Though antibodies have a long history of robust reliability, their use as a means of validating the specificity of our candidate biomarkers presents some significant shortcomings. First is the need to obtain relatively large quantities of purified protein for the immunization process. Although the ProteomeLab™ PF2D can be used to fractionate large quantities of protein extract from crude body fluids, each fraction is still likely to contain multiple proteins. Such mixtures can complicate the production of antibodies. Even if this were not a concern, the binding specificity/cross-reactivity of the resulting antibodies would still need to be individually characterized. This is a time and labor intensive process which would likely have to be repeated before manufacturing a commercial assay system for biological stains. In short, the use of antibodies for larger scale biomarker validation work represents an exceptionally expensive strategy that could require several years to complete.

The use of mass spectrometry based approaches to biomarker validation would circumvent the limitations of an antibody-based strategy. More advanced types of mass spectrometers (*e.g.*, Quadrupole Time-of-Flight (Q-TOF) mass spectrometers) allow specific ions of interest to be selectively isolated and identified from among the thousands present in any given body fluid. Such a "targeted Q-TOF approach" it becomes possible to rapidly assay virtually any body fluid for the presence or absence of biomarkers of interest. If a candidate biomarker is not present in a given bodily fluid, then no protein would be detected by the Q-TOF assay. The important bottom line for future research aimed at validating the specificity of candidate protein biomarkers, therefore, is that Q-TOF-based assays allows unprocessed biological stains to be directly scanned for biomarkers of interest. Furthermore, even though protein abundance will obviously vary among individuals, this should not adversely impact the specificity or potential utility of these markers. For example, even though statherin encompasses <1% of the total saliva protein content, it can be readily detected at the sub-picogram detection limits that are possible using a targeted mass spectrometer. Ongoing research in the author's

laboratory focusing on exactly this approach has begun to yield useful data on the specificity of several of the biomarkers that were presented in Table 1.

In future assay development studies, a second round of selection (from the list of candidate biomarkers shown in Table 1) for biomarkers that are best suited to forensic applications will likely place greater emphasis on absolute biomarkers abundance. This is based on the view that more abundant candidates should have more utility with degraded samples as well as more complex mixed samples. That being said, researchers should remain cognizant of the fact that just because a protein is more abundant does not necessarily mean it is necessarily a better candidate. This is because not all proteins will degrade at the same rate. As a result, a low abundant protein may actually be a better (*i.e.*, more persistent) biomarker. Future research would properly include studies of biomarker degradation rates. A second question that arises is how abundance will affect the processing of casework samples. For example a peripheral blood sample with all high abundance markers mixed with lower abundant vaginal secretion markers. Will the vaginal markers be masked by the blood proteins? These are important issues but ones that can only be resolved after the final detection assay is developed and employed as part of a larger-scale study. Efforts to do this are currently underway.

As researchers move forward, it is important for all investigators to remain cognizant of the standards for admitting scientific evidence in the federal courts. Experiments must be planned with both Frye's "general acceptance" test, the Daubert standard and federal rules of evidence in mind. Future studies coupled with publication in peer-reviewed journals, therefore, would help to place the findings of this research on sound legal footing.

Cited References

1. Virklera, K. and Lednev, I.K. *Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene*. Forensic Sci Int, 2009. **188**(1-3): p. 1-17.
2. *Biology Methods Manual*. 1978: Metropolitan Police Forensic Science Laboratory.
3. *Protocol Manual*. 1989: FBI Laboratory Serology Unit.
4. 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.
5. 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.
6. *OneStep ABAcard® HemaTrace® for the Forensic Identification of Human Blood.*, Abacus Diagnostics, Inc. p. Product insert.
7. Sensabaugh, G.F., *Isolation and characterization of a semen-specific protein from human seminal plasma: a potential new marker for semen identification.* J Forensic Sci, 1978. **23**(1): p. 106-15.
8. *OneStep ABAcard® p30 Test for the Forensic Identification of Semen*, Abacus Diagnostics, Inc. p. Product insert.
9. Keating, S.M., *Oral Sex--a review of it's prevalence and proof.* Journal of the Forensic Science Society, 1988. **28**: p. 341-355.
10. Balsells, D., et al., *Reference values for alpha-amylase in human serum and urine using 2-chloro-4-nitrophenyl-alpha-D-maltotrioside as substrate.* Clin Chim Acta, 1998. **274**(2): p. 213-7.

11. Quarino, L., et al., *An ELISA method for the identification of salivary amylase.* J Forensic Sci, 2005. **50**(4): p. 873-6.
12. Singh, V.N., *Human uterine amylase in relation to infertility.* Horm Metab Res, 1995. **27**(1): p. 35-6.
13. Juusola, J. and J. Ballantyne, *Multiplex mRNA profiling for the identification of body fluids.* Forensic Sci Int, 2005. **152**(1): p. 1-12.
14. 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.
15. 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.: 217-226.
16. Bauer, M. and D. Patzelt, *Evaluation of mRNA markers for the identification of menstrual blood.* J Forensic Sci, 2002. **47**(6): p. 1278-82.
17. 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.
18. Griffin, T.J., et al., *Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae.* Mol Cell Proteomics, 2002. **1**(4): p. 323-33.
19. Bauer, M., et al., *Quantification of mRNA degradation as possible indicator of postmortem interval--a pilot study.* Leg Med (Tokyo), 2003. **5**(4): p. 220-7.
20. Inoue, H., A. Kimura, and T. Tuji, *Degradation profile of mRNA in a dead rat body: basic semi-quantification study.* Forensic Sci Int, 2002. **130**(2-3): p. 127-32.
21. Dobberstein, R.C., Huppertz, J., von Wurmb-Schwark, N., Ritz-Timme, S. *Degradation of biomolecules in artificially and naturally aged teeth: implications for age estimation based on aspartic acid racemization and DNA analysis.* Forensic Sci Int. 2008. **179**(2-3): 181-91.
22. Laux, D.L., Tambasco, A. J., Benzinger, E. A. *Forensic Detection of Semen II. Comparison of the Abacus Diagnostics OneStep ABAcard p30 Test and the Seratec PSA Semiquant Kit for the Determination of the Presence of Semen in Forensic Cases*. [cited; Available from: http://mafs.net/pdf/laux2.pdf.
23. Service, R.F., *Proteomics. High-speed biologists search for gold in proteins.* Science, 2001. **294**(5549): p. 2074-7.
24. 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.
25. 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.
26. 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.
27. 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.
28. Dasari, S., et al., *Comprehensive proteomic analysis of human cervical-vaginal fluid.* J Proteome Res, 2007. **6**(4): p. 1258-68.

Dissemination of Research Findings

A total of seven progress reports on this research program have been provided to the National Institute of Justice. Research findings were also disseminated through poster presentations and invited research seminars listed below. With the completion of the biomarker discovery objectives under award 2006-DN-BX-K001, preparations of a manuscript is underway for submission to the Journal of Forensic Science, the Journal of Proteomics and equivalent publications that serve the forensic and human proteomic communities. An invitation to publish a manuscript in the Journal Current Bioinformatics has also been extended for work relating to the development and use of ProteinMiner™ software application.

Invited Research Talks and Poster Presentations

July 2007 National Institute of Justice, "Isolation of Highly Specific Protein Markers for the Identification of Biological Stains: Adapting Comparative Proteomics to Forensics", Invited Talk, The NIJ Conference 2007 Forensic DNA: Tools, Technology, and Policy.

July 2008: "Comparative Proteomics of Human Body Fluids for Forensic Applications", Poster Presentation, The National Institute of Justice Conference 2008 - Forensic DNA: Tools, Technology, and Policy. Washington, D.C.

September 2008: DNA Forensics Technology Working Group, "Isolation of Highly-Specific Protein Markers for the Identification of Biological Stains: Adapting Comparative Proteomics to Forensics", Invited Talk, National Institute of Justice. Washington, DC

July 2009: "Isolation of Highly Specific Protein Markers for the Identification of Biological Stains: Adapting Comparative Proteomics to Forensics", Poster Presentation: The National Institute of Justice Conference 2009, Washington, D.C.

July 2009: "Isolation of Highly Specific Protein Markers for the Identification of Biological Stains: Adapting Comparative Proteomics to Forensics", Invited Talk, 2nd Annual Green Mountain DNA Conference. Department of Public Safety, Vermont Forensic Laboratory. Burlington, VT

September 2009 “Isolation of Highly Specific Protein Markers for the Identification of Biological Stains: Adapting Comparative Proteomics to Forensics”, Invited Talk, National Authority for Scientific Research, Tripoli, Libya

September 2009 “Isolation of Highly Specific Protein Markers for the Identification of Biological Stains: Adapting Comparative Proteomics to Forensics”, Invited Talk, Biotechnology Research Center, Tripoli, Libya

November 2009: “Forensic Analysis of Human DNA and Proteins in Criminal Investigations”, Invited Talk, National Associate of Biology Teachers Conference, Denver, CO

March 2010: “Isolation of Highly Specific Protein Markers for the Identification of Biological Stains: Adapting Comparative Proteomics to Forensics”, Poster Presentation, 6th Annual US Human Proteomics (HUPO) 2010 Conference, Denver CO

Panel of Candidate Protein Biomarkers: To promote scientific collaboration and to facilitate the research efforts of other scientists working in this area, the identities of all candidate protein biomarkers identified in the biomarker discovery phase of this research project are made freely available upon request.

Protein Miner™ Software Availability: This bioinformatics application for porting and analyzing ProteomeLab™ PF2D .dat files is made freely available to any interested researchers interested in extending their proteome mapping capabilities.