Development of a Proteomic Assay for Menstrual Blood, Vaginal Fluid and Species Identification

Final Draft Technical Report NIJ Grant 2010-DN-BX-K192

Principal Investigator:
Donald Siegel, Ph.D.
Principal Scientist
Office of Chief Medical Examiner
421 East 26th Street
New York, NY 10016
Tel: 212-323-1434
Fax: 212-323-1560
Email: dsiegel@ocme.nyc.gov
Web: www.nyc.gov/ocme

TABLE OF CONTENTS

ABBREVIATIONS

A2ML1 = alpha-2-macroglobulin-like 1 precursor

Alb = albumin

AP = acid phosphatase

BCA = bicinchoninic acid

BV = bacterial vaginosis

CBG = corticosteroid-binding globulin

CE= capillary electrophoresis

CHCA = *alpha*-Cyano-4-hydroxycinnamic acid

CPLC = combinatorial ligand peptide chromatography

CRYS = Beta-crystallin S, vaginal fluid marker protein

CST = Community State Types

EDTA = ethylenediaminetetraacetic acid

ELISA = Enzyme-linked immunosorbent assays

Hb = hemoglobin

HPLC = high performance liquid chromatography

MiGE = microfluidic gel electrophoresis

MRM = multiple reaction monitoring

MS = mass spectrometry

PAGE = polyacrylamide gel electrophoresis

PAEP = progestagen-associated endometrial protein

PLCD1 = phospholipase C delta-1

PSA = prostate-specific antigen

PZP = pregnancy zone protein

RT = reverse transcription

RT-qPCR = reverse transcription – quantitative polymerase chain reaction

SDS = sodium dodecyl sulfate

SPRR3 = small proline-rich protein 3

SUP = supernatants

SWATH = *Sequential Window Acquisition of all THeoretical fragment-ion spectra*

TCEP = TCEP (tris(2-carboxyethyl)phosphine)

TMPRSS11B = Transmembrane protease, serine 11B

TMPRSS11D = Transmembrane protease, serine 11D

WH = whole homogenates

ABSTRACT--

Body fluid identification plays a vital role in forensic investigations contributing to case evidence as well as in directing the course of further investigations - particularly with respect to sexual assaults. Yet the methods used for body fluid detection have, for the most part, failed to keep pace with scientific and technological advances. As a consequence, current testing is a diverse mixture of techniques that vary dramatically in sensitivity, reproducibility, speed and above all specificity – i.e. their ability to confirm that test results are real. Additionally, there are many body fluids for which no tests are currently available. For example, forensic cases arise where the ability to distinguish menstrual blood from venous blood could prove critical to the outcome of criminal investigations or trials. Similarly, a confirmatory test capable of identifying vaginal fluid, particularly in sexual assault cases, would be valuable to the criminal justice system. Currently, however, no routine forensic tests are performed for either menstrual blood or vaginal fluid.

Part of the difficulty of developing confirmatory assays for menstrual blood and vaginal fluid is that unlike blood, saliva and semen, which have known functional marker proteins that are specific to or enriched in them (e.g. hemoglobin, amylase and semenogelins respectively), menstrual blood is composed mostly of blood, while vaginal tract epithelial cells share many of the functions of other mucus membranes open to the environment and consequently have similar proteomes. Identifying unique or enriched proteins in their respective proteomes (e.g. from the uterine lining in menstrual blood, or epithelial cells or mucosal secretions from the vaginal tract) is the purpose of this application.

Using menstrual blood samples collected from 50 volunteers during each day of menses and vaginal fluid samples collected twice over the course of their menstrual cycle, we extracted and analyzed the proteomes of both body fluids to identify candidate markers for each. These markers were present in all subjects regardless of ethnicity, age or use of hormonal contraception, and consequently can be used to establish robust and confirmatory assays for both body fluids.

Further, we have demonstrated that as part of routine blood sample analysis, species information is inherent in the results – including the ability to distinguish primates. While not directly related to menstrual blood assay, the ability to routinely distinguish species by blood analysis has other forensic applications.

EXECUTIVE SUMMARY –

• SYNOPSIS OF THE PROBLEM: Body fluid identification plays a vital role in forensic investigations contributing to both case evidence and in directing the course of further investigations - particularly with respect to sexual assaults. Yet the methods used for body fluid detection have, for the most part, failed to keep pace with scientific and technological advances. As a consequence, current testing is a diverse mixture of techniques that vary dramatically in sensitivity, reproducibility, speed, cost and above all specificity – i.e. their ability to confirm that test results are real. Even quantitative immunoassays like ELISA suffer from the inherent limitations of antibody specificity, affinity, narrow working ranges and manufacturing variability, as well as relatively high costs and long assay times. Additionally, there are many body fluids for which no tests are currently routinely available, e.g. menstrual blood and vaginal fluid.

Unlike DNA testing, which relies on a single, dominant technology for testing all samples regardless of their source, there is no single uniform methodology to simultaneously evaluate an unknown forensic sample for all possible body fluids. Consequently, each body fluid test must be performed separately, consuming time *and* evidence, as well as requiring multiple instruments and laboratory personnel trained in the nuances of each method. In short, a single, reproducible, confirmatory test that can identify all body fluids simultaneously is sorely lacking.

What would constitute an ideal body fluid assay? A single method able to simultaneously detect all body fluids accurately, reproducibly, rapidly, with high sensitivity, at low cost and above all - it must be confirmatory. Movement in this direction is the subject of much current research, particularly in the areas of mRNA, micro-RNA and DNA methylation. However, we believe a mass spectrometry (MS) proteomics assay offers several advantages over these methods: i) proteins are less susceptible to environmental degradation than RNA and DNA, ii) partial degradation of proteins does not preclude marker identification (nucleic acid assays require unbroken sequences from primer to primer), iii) multiple specific protein markers are present in each body fluid and simultaneously detected – increasing confidence that test results are correct, iv) mass spectrometry is unbiased with regard to sample type, i.e. no prior knowledge of a sample is necessary prior to assay (e.g. no specific body fluid or species primers), MS simply identifies those proteins that are present, v) species information is inherent in the sample and reported with results and vi) the assay is amenable to high throughput, and costs are low.

Finally, as mentioned, there are currently no routine assays available for the identification of menstrual blood and vaginal fluid. Part of this problem stems from the fact that, unlike blood, saliva and semen – which possess abundant and unique, or enriched, marker proteins inherent to their biological function - few such markers are known to be present in menstrual blood and vaginal fluid. Menstrual blood is composed primarily of blood and the external layer of the endometrium expelled during menses. Common blood specific markers demonstrate the presence of blood, but contribute nothing to the identification of menstrual blood. And while cells of the endometrial surface contribute to the specific function of egg implantation in the uterus, the endometrium also shares many functional characteristics of other mucus membranes, and consequently shares many of the same proteins. Similarly, the vaginal tract functions in many of the same ways as other mucosal membranes exposed to the environment (e.g. mouth, nasal passages, etc.), i.e. maintaining moisture, electrolytes, pH and defense against invading pathogens. Consequently the vaginal tract and its secretions also share many of the proteins that participate in these mucosal functions (e.g. secretion of mucins, immunoglobulins, defensins, etc.). It is because of the shared *functional* nature of the endometrium and vaginal tract with other mucus membranes (and because any specific differences in protein expression are likely to be relatively small in comparison to overall function of these organs) that unique or enriched markers have been difficult to identify.

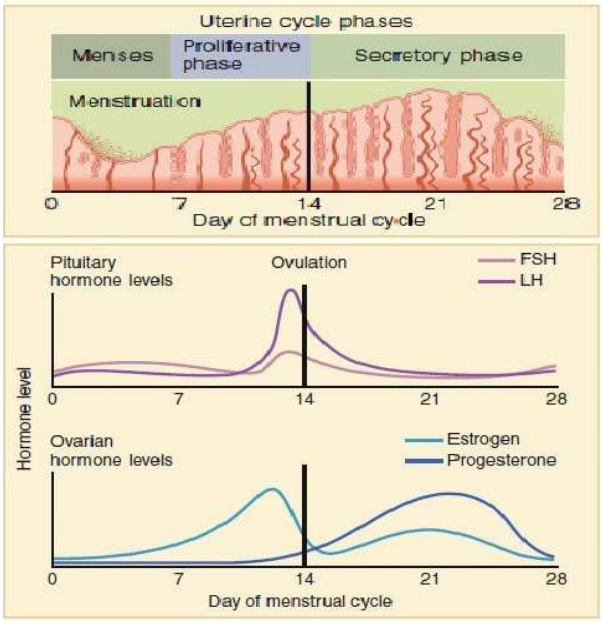
• PURPOSE: The overarching goal of this work has been to exploit the power of mass spectrometry as an accurate and sensitive analytical tool to identify and evaluate menstrual blood and vaginal fluid candidate markers. Such markers must meet specific criteria: i) they must be specific to, or highly enriched in, each of body fluid so that there is no overlap with markers from body fluids of other mucus membranes, ii) multiple markers must be identified to ensure confidence that no other body fluid or mixture of body fluids could give the same results, iii) menstrual blood markers must be present in all women regardless of age, ethnicity, use of hormonal contraceptives, and present in every day of a women's period, and iv) vaginal fluid markers must also be present in all women regardless of age (including post-menopausal), ethnicity, and use of hormonal contraceptives. The reason for testing all these criteria is to ensure that selected markers are present in all women and that the use of hormonal contraceptives, which function to prevent pregnancy through multiple mechanisms including suppression of ovulation, as well as changes in cervical mucus do not result in changes in protein expression that could be missed in women using them.

• RESEARCH DESIGN: As stated, the major objectives of this project were to evaluate menstrual blood and vaginal fluid samples from a large number of women of varying ages, ethnicities and use of oral contraception. Menstrual blood samples were taken during each day of a woman's period and vaginal fluid twice during the length of her cycles. As the goal of this work was to unbiasedly evaluate the proteomes of menstrual blood and vaginal fluid (i.e. not select candidate markers from the literature) we chose not to extract proteins from tampons, as there is always the possibility of non-specific bio-molecule binding to tampon material which could confound results. Consequently, both body fluids were collected in disposable menstrual blood cups as described below. To ensure that no menstrual blood candidate markers were also in circulating blood, venous blood was taken during each woman's period and used as a control.

I) *MENSTRUAL & VENOUS BLOOD COLLECTION*: Menstrual blood samples were obtained from 45 volunteers on each day of their periods. Samples were collected anonymously in sterile, individually packaged, disposable menstrual cups. Women were asked to keep cups in place for a minimum of four but no longer than twelve hours. When cups were removed they were inserted with contents into sterile 50 ml conical centrifuge tubes, placed in small black plastic bags and stored at -20^{o}C. When all samples had been collected, they were brought to the laboratory and stored at -80^{o}C until analyzed. Venous blood was taken from volunteers on the first day of their periods, unless this occurred on a weekend or holiday, in which case it was taken on the first weekday opportunity thereafter. Venous blood was obtained by finger lance and collected in sterile tubes containing EDTA as an anticoagulant. Samples ranged between ~50 and 500 µl. A total of 193 menstrual blood and 45 blood samples were obtained.

II) *VAGINAL FLUID COLLECTION*: Vaginal fluid was collected from these same 45 women (using the same menstrual cup procedure) on the 13th and 19th days following the first day of their periods – equivalent to approximately one week following the end of their last period and one week prior to their next period (assuming a 28 day menstrual cycle) thus capturing the proliferative and secretory stages of the uterus, respectively (see figure below). In addition, five postmenopausal women were asked to use menstrual cups to collect two vaginal fluid samples at one week intervals. Venous blood was taken from these women at the time of the first vaginal fluid collection.

Figure i 28 day menstrual cycle showing endometrial changes with hormone



Source: OpenStax College - Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/, Jun 19, 2013

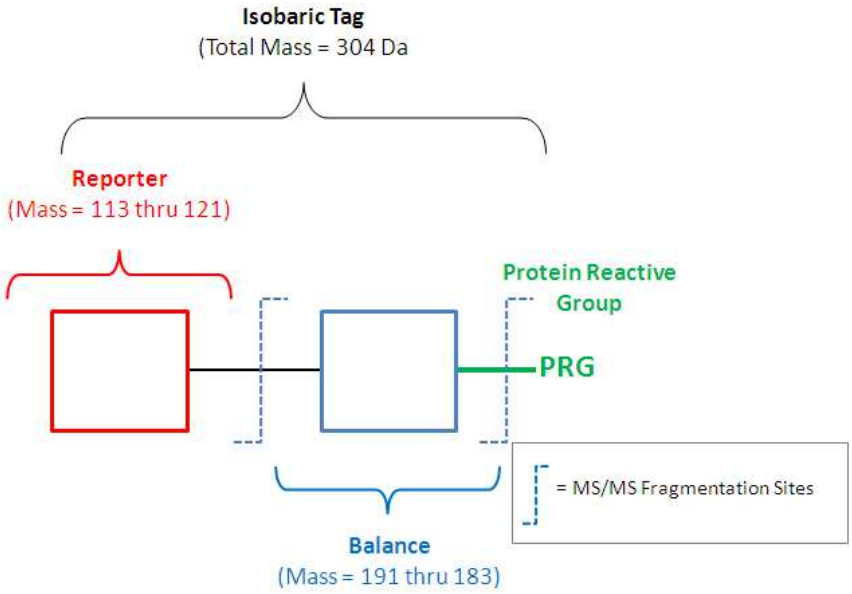
III) *SAMPLE PREPARATION AND PROCESSING*: Menstrual/venous blood and vaginal fluid samples were homogenized and debris removed by centrifugation. Protein quantitation was performed at each step and the quality of samples analyzed by gel electrophoresis. Because of the large number of abundant common proteins shared between menstrual and venous blood (e.g. hemoglobin, albumin, transferrin, immunoglobulins, etc.) these proteins were partially removed by combinatorial peptide ligand chromatography (CPLC) - a method that reduces protein dynamic range (i.e. the difference between the most abundant and least abundant proteins in a sample). This method allowed us to find those proteins specifically expressed in menstrual blood.

Prior to mass spectrometry analysis both individual menstrual/venous blood and vaginal fluid samples were labeled with isobaric barcodes which allowed us to multiplex samples and consequently perform relative quantitation of the protein expression between samples. Importantly, we included a common standard menstrual blood sample in all multiplexed runs and were thereby able to compare all menstrual candidate proteins between all women as well as between the different days of their periods.

◦ *How Isobaric Barcoding Works*: The isobaric barcode used (iTRAQ from SCIEX, Framingham, MA) is schematically represented in Figure 15. It is composed of three functional groups: i) the *Protein Reactive Group* which allows the tag to covalently react with primary amines from N-terminal or lysine on each peptide, ii) the *Reporter* which is the actual barcode composed of eight different masses (113–119 Da plus 121 Da), and iii) the *Balance* which, in combination with the reporter always results in an isobaric mass of 304 Da for all iTRAQ labels (e.g. Reporter 113 + Balance 191, Reporter 114 + Balance 190, etc.). The importance of keeping the mass of the *Reporter* plus *Balance* constant (isobaric) is to ensure that the same marker peptide from different samples (e.g. different days of menstruation from one individual) migrate together in the mass spectrometer which separates and identifies peptides based on their mass.

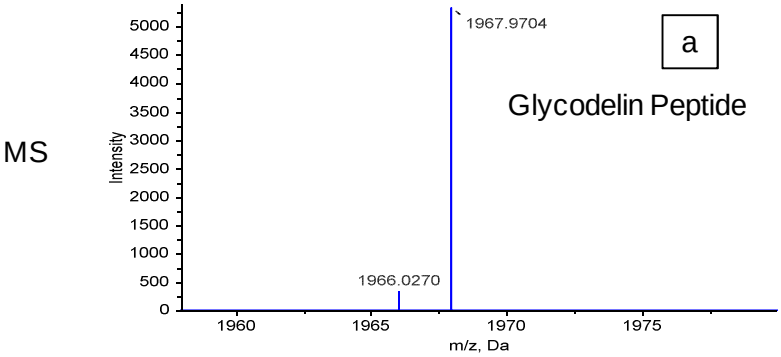
Subsequent fragmentation of the co-migrating peptides in the mass spec's collision chamber releases the *Reporters* (barcodes) which are then detected. A comparison of each barcode's peak height (intensity) shows the relative quantity of the marker peptide present in the different samples. The relative intensities of these ions are proportional to the amount of this peptide in each volunteer's sample. These relative peptide quantities are reported by Protein Pilot software (AB Sciex) as ratios relative to a labeled standard in the run. Protein ratios are obtained using a weighted average of the ratios of all the peptides contributing to the protein identification.

Figure ii Schematic Representation of iTRAQ Isobaric Tag



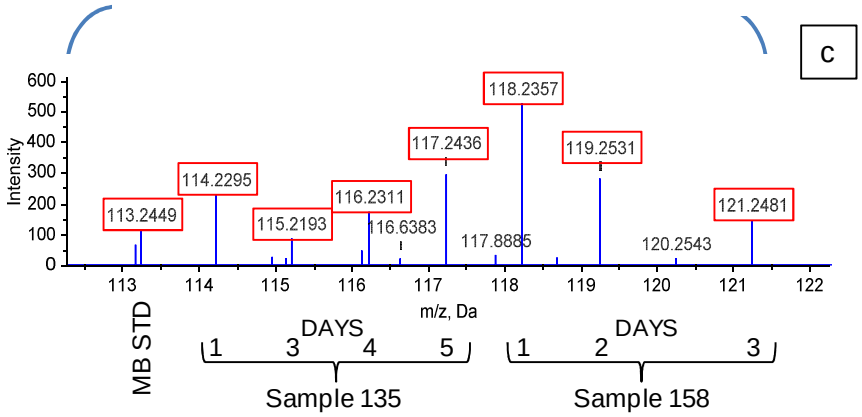
An example of iTRAQ multiplexing and quantitation of barcoded samples is shown in Figure iii. Here seven menstrual blood samples from two volunteers were individually labeled with iTRAQ isobaric tags containing reporters 114 - 119 plus 121 Da. After labeling, samples were combined, peptides separated by nano-HPLC and then analyzed by MALDI TOF/TOF. Figure iiia is the mass spec signal of a single peptide with a mass of 1967.97 Da. MS/MS of this peptide (Figure iiib) reveals it is from the menstrual blood marker protein glycodelin with the amino acid sequence *VLVEDDEIMQGFIR*. The bracket below the MS/MS peak at ~118 Da is expanded in Figure iiic and shows each of the reporter ions 113 - 119 plus 121.

Figure iii Multiplexing Isobaric Barcode Labeled Samples for Mass Spec Analysis



MS/MS

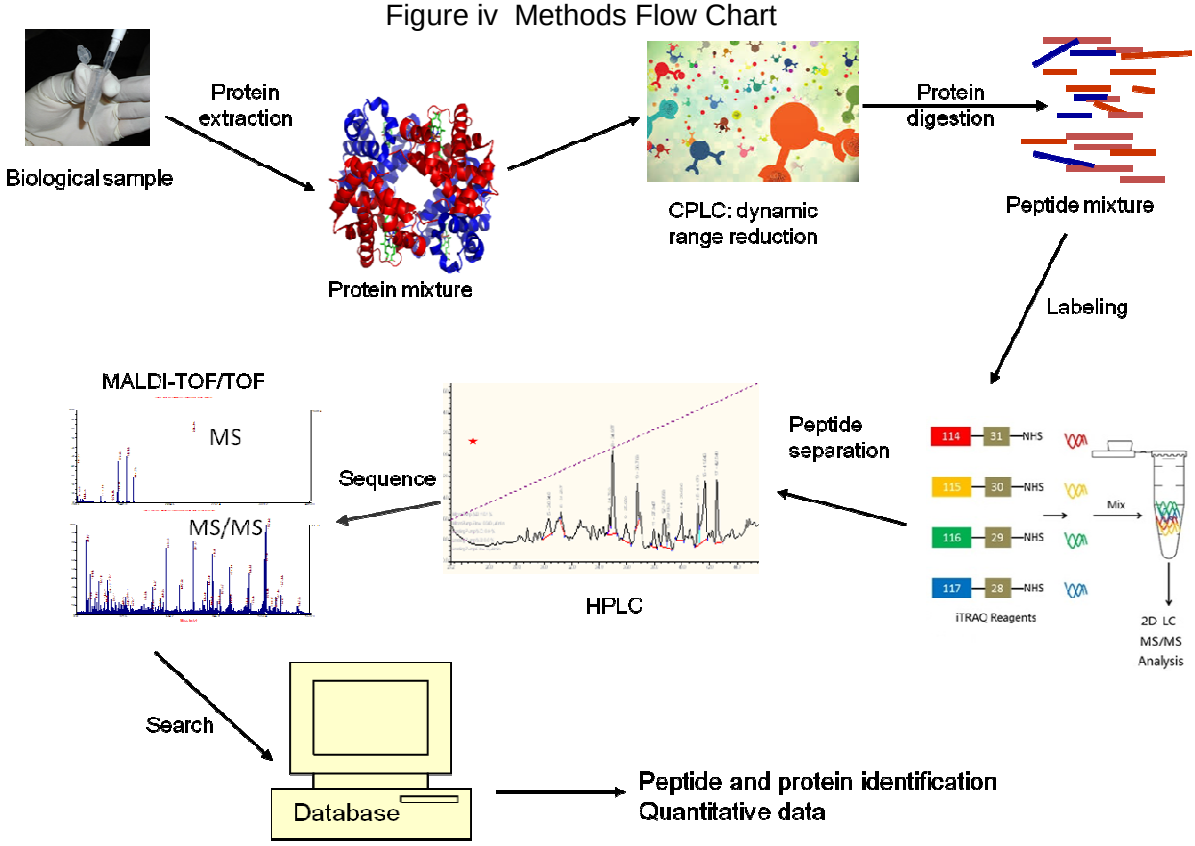
Peptide Sequence
VLEGVWIFYELPNY



MS and MS/MS spectra of peptide VLEGVWIFYELPNY from menstrual blood marker protein glycodelin labeled with isobaric iTRAQ tagging reagent. **a**, MS spectrum of the peptide; **b**, MS/MS spectrum which is used to determine the peptide's sequence; **c**, expanded MS/MS spectrum of iTRAQ barcodes showing the relative amounts of peptides from each sample. Barcode Key: 113 = menstrual blood standard used in all experiments (volunteer 280, day 2); 114-117 = volunteer 135, days 1, 3-5; 118, 119 & 121 = volunteer 158, days 1-3. (Volunteer 135 day 2 had insufficient protein for MS analysis.)

The intensities of the reporter ions indicate the relative concentrations of the glycodelin menstrual blood marker in each woman's sample. Thus, for example, in volunteer 135 peptide intensities ranged from ~ 100 to 350 during the four days of her period. In volunteer 158, intensities ranged from ~200 to 550. In all experiments reporter ion 113 was used to label a single menstrual blood sample (volunteer 282, day 2) that was spiked into all mass spec runs in order to allow quantitative comparisons of marker proteins across runs.

Figure iv Methods Flow Chart



Following TOF/TOF mass spectrometry, proteins were identified by searching spectra against protein databases. Unique or highly enriched menstrual blood candidate marker proteins were identified by comparing menstrual blood and venous blood results using multivariate analysis software (Qlucore). Marker proteins found only in menstrual blood were then compared against 13 other body fluid proteomes in the Sys-BodyFluid database (saliva, seminal fluid, serum/plasma, vaginal fluid, amniotic fluid, bronchoalveolar lavage, cerebrospinal fluid, milk, nipple aspirate, synovial fluid tear and urine) in order to determine those markers that could be used for menstrual blood identification. Vaginal fluid specific or enriched candidate markers were determined by comparing all identified proteins against the 13 body fluid proteomes in the Sys-BodyFluid database (Li et al. 2009).

• FINDINGS: Multiple specific and/or enriched markers were found in both menstrual blood and vaginal fluid that we believe can be used for establishing proteomic assays for both body fluids.

i) *MENSTRUAL BLOOD*: Menstrual blood markers may be divided into three broad categories: i) markers found in all women, but also detected in other body fluids ii) markers found only in menstrual blood but not found in all women, and iii) markers enriched in menstrual blood but detected in other body fluids and not found in all women. Each category is discussed below.

i) ***Markers Found in All Women, but Also Detected in Other Body Fluids –-*** Five proteins were consistently and robustly identified in all menstrual blood samples and can easily distinguish menstrual blood from venous blood. These proteins are: 1) progestagen-associated endometrial protein (PAEP aka glycodelin); 2) glycogen phosphorylase, liver form; 3) small proline-rich protein 3; 4) ADP/ATP translocase 2 and 5) keratin, type II cytoskeletal 7.

Two of these proteins, progestagen-associated endometrial protein and ADP/ATP translocase 2, have been reported in comprehensive proteomic analyses of venous blood at low levels. However, we have not detected these proteins in the 45 venous blood samples used in this study. Progestagen-associated endometrial protein is an excellent marker for distinguishing menstrual blood from venous blood as it is a well documented endometrial protein and known to increase in the secretory phase of the menstrual cycle. However, we, and others (Pilch & Mann 2006; Chiu et al. 2007) have also detected PAEP in seminal fluid. Nevertheless, the presence of PAEP in a blood stain in the absence of the specific and abundant seminal fluid markers semenogelin 1 and 2 would suffice to exclude the possibility of a mixture of semen and menstrual blood.

Table i Unique or Enriched Menstrual Blood Marker Proteins

No.	Protein ID	Protein Name	No. Peptides	No. Women	Saliva	Seminal Fluid	Serum/ Plasma	Vaginal Fluid	Amniotic fluid	Broncho-alveolar Lavage	Cerebro spinal Fluid	Milk	Nipple Aspirate	Synovial Fluid	Tear	Urine
1	P09466	Progestagen-associated endometrial protein (PAEP aka Glycodelin)	2	ALL			*									
2	P06737	Glycogen phosphorylase, liver form (PYGL)	5	ALL												
3	Q9UBC9	Small proline-rich protein 3 (SPRR3)	2	ALL												
4	P05141	ADP/ATP translocase 2 (ADT2)	4	ALL			*									
5	P08729	Keratin, type II cytoskeletal 7 (K2C7)	4	ALL												
6	P14314	Glucosidase 2 subunit beta (GLU2B)	1	37												
7	Q9UIV8	Serpin B13 (SPB13)	8	35												
8	Q00325	Phosphate carrier protein, mitochondrial (MPCP)	2	35												
9	P22749	Granulysin (GNLY)	1	33												
10	P43307	Translocon-associated protein subunit alpha (SSRA)	1	32												
11	P39656	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit (OST48)	3	32												
12	P12236	ADP/ATP translocase 3 (SLC25A6)	4	31												
13	Q9UBG3	Cornulin (CRNN)	2	29												
14	P30536	Translocator protein (TSPOA)	1	28												
15	P00403	Cytochrome c oxidase subunit 2 (COX2)	1	27												
16	P21397	Amine oxidase [flavin-containing] A (AOFA)	1	24												
17	P20962	Parathymosin (PTMS)	1	24												
18	P51571	Translocon-associated protein subunit delta (SSRD)	1	22												
19	P49755	Transmembrane emp24 domain-containing protein 10 (TMEDA)	2	22												
20	A8K2U0	Alpha-2-macroglobulin-like protein 1 (A2ML1)	5	21												
21	Q9NRM1	Enamelin (ENAM)	1	20			*									
22	P24158	Myeloblastin (PRTN3)	2	20												

* Found in plasma/serum in very low amounts.
= Unique to menstrual blood. Not found in Sys-BodyFluid database.
= Identified by NYC OCME.
= Found in Sys-BodyFluid database
= Identified by NYC OCME & found in Sys-BodyFluid database

Except for PAEP, the four remaining proteins have been identified in the Sys-BodyFluid database to be present at varying levels in saliva, although we routinely find only small proline-rich protein 3, likely due to its high abundance. However, here again, the absence of other saliva specific and abundant markers (e.g. alpha-amylase and cystatin-S) would be sufficient to determine that the source of these proteins are not saliva. Finally, it should be noted that small proline-rich protein 3 is also present in, and we consider it a marker for, vaginal fluid. However, a mixture of *venous* blood and vaginal fluid could not be mistaken for menstrual blood as such a mixture would be missing nine known menstrual blood markers (see Table i).

ii) ***Markers Found Only in Menstrual Blood but Not Found in All Women –-*** When checked against the *Sys-BodyFluid* database, five menstrual blood proteins were identified that had not been reported in any of the 12 body fluids in the database. These were: 1) granulysin found in 33 women (85%); 2) dolichyl-diphosphooligosaccharide-protein glycosyltransferase found in 32 women (82%); 3) translocator protein found in 28 women (72%); 4) amine oxidase A found in 24 women (62%); and 5) parathymosin also found in 24 women (62%). Of these only dolichyl-diphospho-oligosaccharide-protein glycosyltransferase was identified by more than one peptide. While not found in all women, when identified these markers give the greatest confidence that menstrual blood is present. However, it should be noted that while not possible on a MALDI mass spectrometer, using multiple reaction monitoring (MRM) or SWATH (a SCIEX proprietary data-independent acquisition method - *Sequential Window Acquisition of all THeoretical fragment-ion spectra*) additional peptides are likely to be detected in positive samples and, because of the sensitivity of these methods, additional samples are likely to be identified as positive.

iii) ***Markers Enriched in Menstrual Blood but Detected in Other Body Fluids and Not Found in All Women –-*** In addition to the ten markers described above, 12 additional proteins were identified in menstrual blood that were not detected in venous blood. These 12 proteins were found in 51% - 95% of all women. Nearly all were identified in the Sys-BodyFluid database in saliva. However, as described above, in the absence of alpha-amylase and cystatin-S the contribution of these markers could not be confused for saliva. We have also identified four of these 12 proteins in vaginal fluid, two of which (small proline-rich protein 3 and alpha-2-macroglobulin-like protein 1) are consistent markers for vaginal fluid. It is not possible to say whether these vaginal fluid markers are normal constituents of menstrual blood or whether they become mixed with menstrual blood as it passes through vaginal canal. However, the presence of these markers in a mixture of vaginal fluid and venous blood could not be confused for menstrual blood since the five menstrual blood specific markers would be absent.

It is important to point out that the markers identified in this study are proteins inherent in uterine function and consistent with the dynamic processes of the menstrual cycle – in particular the secretory phase (the period just prior to menstruation) and the beginning of menstruation which is currently regarded as an inflammatory response. In the secretory phase, the endometrium has reached its full thickness and cells are preparing for blastocyst implantation through increased metabolic activity, protein synthesis and secretion. Increased levels of PAEP as well as ADP/ATP translocase 2 & 3, phosphate carrier protein and cytochrome c oxidase

subunit 2 are reflective of this stage of the menstrual cycle. Lack of implantation leads to increases in proteolytic enzymes, proteins involved in apoptosis as well as proteins from the increased numbers of immune cells that are an integral part of menstruation, and, of course, proteins from the breakdown of the uterine wall. These proteins include serpin B13, granulysin, glucosidase 2 subunit beta, parathymosin and keratin, type II cytoskeletal 7.

Finally, for all menstrual blood markers identified, both those found in all women and those found only in some women, when present, there was no difference between their expression on different days of a woman's period, or on a woman's age, ethnicity or use of hormonal contraception – meaning these markers are commonly expressed during a woman's entire menstrual cycle. Perhaps the most significant finding of this work, however, was the discovery that multivariate unsupervised hierarchical clustering analysis of menstrual and venous blood naturally separated them into their respective types. These data suggest a new method for serology analysis not restricted to menstrual blood but to all body fluids and body fluid mixtures.

II) *VAGINAL FLUID*: Four marker proteins were consistently found in both Day 13 and Day 19 vaginal fluid samples from women of child bearing age, post menopausal women and one volunteer using the hormonal contraceptive Depo-Provera who did not have a period. These markers are: 1) small proline-rich protein 3 (SPRR3,100% D13 / 100% D19), 2); alpha-2-macroglobulin-like 1 precursor (A2ML1,100% D13 / 100% D19); 3) transmembrane protease, serine 11B (TMPRSS11B, 100% D13 / 100% D19) and 4) transmembrane protease, serine 11D (TMPRSS11D, 92% D13 / 96% D19). Of these TMPRSS11B has not been reported in any of the 13 body fluids in the Sys-BodyFluid database and consequently is the strongest candidate for vaginal fluid identification. SPRR3, A2ML1 and TMPRSS11D have been reported in saliva, although we have only found SPRR3 in saliva. Here again, however, a vaginal fluid/saliva mixture could be identified by the presence of TMPR11B, which is not present in saliva, and by amylase, cystatin SA and histatin which are found in saliva but not found in vaginal fluid. The absence of the abundant saliva markers amylase, cystatin SA and histatin would demonstrate that no saliva is present in a sample of vaginal fluid.

Interestingly, three proteins: Beta-crystallin S (CRYS); phospholipase C delta-1 (PLCD1) and pregnancy zone protein (PZP) appear to show some differential expression between days 13 and 19. CRYS (83% D13 / 33% D19) and PLCD1 (50% D13 / 0% D19) are found more frequently on day 13 than on day 19, while PZP (8% D13 / 50% D19) is found more frequently on day 19 than day 13. That the data are not completely "on and off" on days 13 and 19, is to be expected from a biological system (see for example hormonal levels in the menstrual cycle figure above). If these data hold up in future studies, quantitative levels of expression of these proteins could be used to determine where a women is in her menstrual cycle from a vaginal fluid stain.

Table ii Vaginal Fluid Marker Protein Expression on Days 13 & 19 of the Menstrual Cycle

PROTEIN MARKERS	D-13 EXPRESSION (%)	D-19 EXPRESSION (%)	BODY FLUID EXPRESSION	COMMENTS
SPRR3	48/48 (100)	48/48 (100)	found in saliva, reported in amniotic fluid and urine	Sys-BodyFluid reported in vaginal fluid
A2ML1	48/48 (100)	48/48 (100)	reported in saliva, and urine	Sys-BodyFluid reported in vaginal fluid
T**MPRSS11 B**	48/48 (100)	47/48 (98)	**not reported in other body fluids**	Sys-BodyFluid reported in vaginal fluid

TMPRSS11D	44/48 (92)	46/48 (96)	reported in saliva	Sys-BodyFluid reported in vaginal fluid
CRYGS	40/48 (83)	16/48 (33)	**not reported in other body fluid**	structural components of the vertebrate eye lens
PLCD1	24/48 (50)	0/48 (0)	reported in plasma/serum	not found by us in serum/plasma
PZP	4/48 (8)	24/48 (50)	reported in bronchoalveolar lavage, cerebral spinal fluid, plasma/ serum, saliva, synovial fluid	not found by us in serum/plasma or saliva, we did not analyze other body fluids mentioned

SPRR3 = small proline-rich protein 3
A2ML1 = alpha-2-macroglobulin-like 1 precursor
TMPRSS11B = Transmembrane protease, serine 11B
TMPRSS11D = Transmembrane protease, serine 11D
CRYS = Beta-crystallin S
PLCD1 = phospholipase C delta-1
PZP = pregnancy zone protein

Additionally, as previously mentioned, mass spectrometry is an unbiased methodology and identifies all peptides found in a sample. Consequently, when vaginal fluid sample data were blasted against the entire proteome database (all species), we identified several of the microorganisms know to be present in the vaginal microbiome and currently proposed to be specific forensic markers for vaginal fluid. Thus, while our data serve to confirm current efforts to identify vaginal fluid by its microbiome, we believe that the human vaginal fluid markers protein we identified give significantly greater confidence for forensic testing than microbial markers.

III) *SPECIES IDENTIFICATION***:** Tandem mass spectrometry effectively sequences all peptides it detects. As with DNA, species identity is inherent in protein sequence information. Consequently, without any additional assay requirements (sample preparation or instrument programing) species are determined. Using blood from 11 different animals (cat, chicken, cow, deer, dog, ferret, pig, rabbit, sheep, turkey and human) we were easily able to identify species by differences in multiple common blood proteins (albumin, β-enolase, carbonic anhydrase 3, creatine kinase M-type, hemoglobin-α, hemoglobin-β, myoglobin and phosphoglucomutase-1) using HPLC-MS/MS with as little as 0.2 nl of blood. Additionally, because the abundance of these proteins in blood, identifications can be made by MS/MS alone (i.e. without HPLC purification) saving significant amounts of time. Finally, comparing human and gorilla blood, we demonstrate the ability to distinguish primates by protein sequence differences in the common blood proteins hemoglobin-α, hemoglobin-β, carbonic anhydrase 1 and catalase.

• CONCLUSIONS INCLUDING IMPLICATIONS FOR POLICY AND PRACTICE:
The major goals of this grant were to use mass spectrometry to identify menstrual blood and vaginal fluid protein markers that would be useful for the forensic identification of these body fluids, and determine if the identified markers were present in all women regardless of age, ethnicity, use of hormonal contraception and, for menstrual blood, present on each day of women's periods, while for vaginal fluid, present on the two most distinct day of the menstrual cycle – the peak of the proliferative phase (Day 13) and peak of the secretory phase (Day 19) –

when hormonal and endometrial cell changes are greatest. An additional goal was to demonstrate the ability of MS to easily identify species. These goals have been achieved.

The discovery of menstrual blood and vaginal fluid protein markers in over 300 samples from fifty women demonstrate their reliability to accurately identify these body fluids. However, additional basic and applied research needs to be done to take these discoveries from the research lab to routine forensic laboratory testing. Additional samples should be evaluated to confirm these data. Robust extraction and downstream processing methods need to be established along with SOPs. Limits of detection need to be determined from both "pure" samples and mock forensic samples applied to multiple substrates to determine if such substrates affect recovery. Finally, mixtures of body fluids need to be examined to determine the limits of detection in mixed samples.

Beyond this grant, the ultimate purpose of this work is to use the knowledge generated here to establish robust, confirmatory and routine forensic assays to aid in criminal investigations, and provide evidence of unquestioned reliability. Protein analysis by mass spectrometry effectively sequences proteins. These data then are similar in their specificity, quality and reliability to DNA analysis of STRs and, consequently, may be used to establish confirmatory serology tests of unparalleled reliability. Nor should such testing be limited to menstrual blood and vaginal fluid. We have already demonstrated (Yang et al. 2013) that serology testing by mass spectrometry for blood, saliva and semen are robust, confirmatory, rapid and more cost effective than other methods currently in the market. Consequently, we believe that policy makers should press for confirmatory serology assays for all relevant forensic samples and continue funding the basic and applied research in order to bring these assays into routine use in public laboratories. As DNA analysis has helped balance the scales of justice by providing impartial evidence – both incriminatory and exculpatory - so too will the establishment of confirmatory serology testing.

INTRODUCTION – This section is organized following Final Technical Report Guidelines. It is composed of three sections: i) Statement of the Problem, ii) Literature Review and Citations, and iii) Rationale for Research.

• STATEMENT OF THE PROBLEM: This section addresses both the limitation of i) current serology assays, as well as ii) menstrual blood and vaginal fluid testing.

I) *SEROLOGY ASSAYS:* Body fluid identification plays a vital role in forensic investigations contributing to case evidence as well as in directing the course of further investigations - particularly with respect to sexual assaults. Yet the methods used for body fluid detection have, for the most part, failed to keep pace with scientific and technological advances. As a consequence, current testing is a diverse mixture of techniques that vary dramatically in sensitivity, reproducibility, speed, cost and above all specificity – i.e. their ability to confirm that test results are real. For example, neither the hemoglobin (Kastle-Meyer) nor amylase tests are confirmatory; both give false positives for a variety of naturally occurring compounds, both rely on the subjective evaluation of qualitative results, both use technology that is nearly 100 years old (Patzelt 2004) and both are in common use today. Even the more quantitative and confirmatory immunoassays like ELISA suffer from the inherent limitations of antibody specificity, affinity, narrow working ranges (i.e. antigen-antibody concentration ratios – the Hook effect), and manufacturing variability, as well as relatively high costs and long assay times. Additionally, there are many body fluids for which no tests are currently available, e.g. menstrual blood and vaginal fluid.

Unlike DNA testing, which relies on a single, dominant technology for testing all samples regardless of their source, there is no single uniform methodology to simultaneously evaluate an unknown forensic sample for all possible body fluids. Consequently, each body fluid test must be performed separately, consuming both time *and* evidence, as well as requiring multiple instruments and laboratory personnel trained in the nuances of each method – e.g. the subjective evaluation of the clearing of an opaque field around a suspected amylase sample in an agarose gel, or the relative appearance of a band on an immuno-stick. In short, a single reproducible, confirmatory test that can identify all body fluids simultaneously is sorely lacking.

What would constitute an ideal body fluid assay? A single method able to simultaneously detect all body fluids accurately, reproducibly, rapidly, with high sensitivity, at low cost and above all - it must be confirmatory. Movement in this direction is the subject of much current research, particularly in the areas of mRNA, micro-RNA and DNA methylation (which are reviewed below). However, we believe a mass spectrometry (MS) proteomics assay offers several advantages over these methods: i) proteins are less susceptible to environmental degradation than RNA and DNA, ii) partial degradation of proteins does not preclude marker identification (nucleic acid assays require unbroken sequences from primer to primer), iii) multiple specific protein markers are present in each body fluid and simultaneously detected – increasing confidence that test results are correct, iv) MS assay sensitivity is greater than that of mRNA assays (Yang et al. 2013), v) mass spectrometry is unbiased with regard to sample type, i.e. no a priori knowledge of a sample is necessary prior to assay (e.g. no specific body fluid or species primers), MS simply identifies those proteins that are present, vi) species information is

inherent in the sample and reported with results and vii) the assay is amenable to high throughput, and costs are low.

II) *MENSTRUAL BLOOD AND VAGINAL FLUID TESTING:* While not frequent, forensic cases do arise where the ability to distinguish menstrual blood from venous blood could prove critical to the outcome of criminal investigations or trials (*New York Times; Bangor Daily News)*. Similarly, a confirmatory test capable of identifying vaginal fluid, particularly in sexual assault cases, would be valuable to the criminal justice system. Currently, however, no routine forensic tests are performed for either menstrual blood or vaginal fluid – for reasons that will be described presently.

Classically, attempts to distinguish menstrual from venous blood have been based on measurements of fibrinolytic activity and relative amounts of fibrinogen and fibrin degradation products analyzed using an unknown blood stain (Whitehead & Divall 1973; Akutsu et al. 2012) This method is based on the fact that menstrual blood does not typically coagulate due to increased protease activity. While this method can show some specificity under ideal, controlled laboratory conditions, its reliability in routine casework is uncertain as non-menstrual blood samples (e.g. blood mixed with semen and/or vaginal fluid) also show activity – albeit at lower levels. However, as samples age and activity declines can an old sample of menstrual blood be distinguished from a fresh mixture of venous blood and semen or vaginal fluid? Recently use of a modified assay for fibrin–fibrinogen degradation products (latex agglutination test) has shown promise, but even if more reliable, it too would only be presumptive. Historically, vaginal fluid identification has been significantly more difficult to assay for, few studies have been able to show differences between vaginal track epithelial cell and those from other mucus membranes (Gaensslen Ph.D 2000). As a result there has been a renewed focus on the vaginal microbiome (Akutsu et al. 2012; Benschop et al. 2012; Fleming & Harbison 2010b) However, here too positive results would only be presumptive. Consequently, none of the classical methods for identifying menstrual blood and vaginal fluid are sufficiently accurate to use as confirmatory tests for these body fluids. (Newer molecular approaches are described below.)

Part of the difficulty of developing a confirmatory assay for menstrual blood and vaginal fluid is that unlike blood, saliva and semen, which have known functional marker proteins that are specific to or enriched in them (e.g. hemoglobin, amylase and semenogelins respectively), menstrual blood is composed mostly of blood, while vaginal tract epithelial cells share many of the functions of other mucus membranes open to the environment and consequently have similar proteomes. Identifying unique or enriched proteins in their respective proteomes (e.g. from the uterine lining in menstrual blood, or epithelial cells or mucosal secretions from the vaginal tract) is the purpose of this application.

- LITERATURE REVIEW: This section is divided into two parts. The first only briefly discuss standard serology methods currently in use as they are well know to the forensic community. The second reviews current research.

I) *CURRENT SEROLOGY METHODS IN USE:*

i) ***Chemical Assays –-*** Identification of blood is typically carried out by catalytic colorimetric assays (e.g. phenolphthalein, benzidine) or crystal formation tests (Teichmann, Takayama). While the colorimetric assays (presumptive) can be fairly sensitive, they lack specificity. False positives can result from the presence of chemical oxidants and catalysts (including rust), as well as from vegetable and animal matter (e.g. apple, potato, blackberry, saliva, mucus, and others (Saferstein 1982) Crystal tests (confirmatory) require, by today's standards, relatively large amounts of material (0.1 mg hemoglobin), and crystal formation can decline with age of the sample (Takayama), or when samples are exposed to substances that cause hematin to lose iron (Teichmann). A negative result does not confirm the absence of blood. Species cannot be determined by this method.

ii) ***Enzymatic Assays –-*** Enzymatic assays are commonly used to identify semen and saliva. The semen assay relies on acid phosphatase (AP), while the saliva assay relies on amylase. Both assays are presumptive, and suffer from limitations inherent on relying on enzyme activities that decline over time and can be accelerated by environmental conditions that denature or degrade them (e.g. heat, chemicals, drying, etc.(Ende 1961; Kohn 1986)). Further, neither alkaline phosphatase nor amylase activities are specific to their respective body fluids. AP is also found in vaginal fluid and amylase in blood. Consequently even if identified, they are not definitive evidence for semen or saliva. Similarly, the absence of AP or amylase activity does not preclude the presence of semen or saliva. Neither test can determine species.

iii) ***Microscopic Observations –-*** While microscopic identification of sperm cells is confirmatory, staining and assays do not always reveal their presence. This can be due to low or no sperm numbers in the ejaculate because of hereditary, pathological conditions or elective surgery. Thus, failure to identify sperm by microscopic examination is not conclusive for the absence of semen (Saferstein 1982)**.**

iv) ***Immunochemical Assays –-*** Enzyme-linked immunosorbent assays (ELISA) for specific body fluid markers (e.g. PSA in semen) address some of the potential problems outlined above since they do not require body fluid enzymes to be active for detection and are confirmatory. However, limitations do exist: antibodies must be highly specific with strong binding affinities and antigenic sites must be present in appropriate conformations (i.e. not denatured or degraded). Because antibodies are products of living organisms, there can be significant batch-to-batch variations in binding affinities which may require assays to be recalibrated after each new lot is validated. False positives are not unknown, and for reliable results material adjacent to a stain should be tested for "environmental epitopes" as non-specific binding of primary antibodies to extracted materials can occur. In addition, as with all of the above assays, ELISA testing is done on one body fluid at a time. That is, a forensic sample of unknown body fluid composition must be divided and tests carried out separately for each. Finally, while species identification can be performed by ELISA, it requires production of individual antibodies for each species tested with all aforementioned limitation for each antibody. Importantly, there is no current, routine test for vaginal fluid and menstrual blood.

II) *CURRENT RESEARCH*: This section reviews the three most extensively studied confirmatory methods currently under development: mRNA, miRNA and DNA methylation assays.

i) ***Messenger RNA Assays –-*** In view of the limitations of extant methods (discussed above) research has been conducted in search of better alternatives. One such alternative, body fluid mRNA expression profiling, has been studied by several laboratories (Juusola & Ballantyne 2005; Juusola & Ballantyne 2007; Nussbaumer et al. 2006; Bauer & Patzelt 2008; Haas et al. 2009; Fleming & Harbison 2010b; Richard et al. 2011) **.** The goal of this technique is to identify body fluid specific mRNA markers that can be used to unambiguously identify each body fluid. The technology employed is typically reverse transcription (RT) followed by quantitative (real time) PCR (RT-qPCR) or capillary electrophoresis (CE). Ballantyne and others have identified specific mRNA and developed multiplex RT-qPCR/CE assays to detect body fluid-specific mRNAs for blood, saliva, semen, vaginal secretions and menstrual blood from single or mixed stains (Juusola & Ballantyne 2005; Haas et al. 2009; Fleming & Harbison 2010a; Haas et al. 2008) There are, however, several potential drawbacks to this method that may limit its usefulness. Primary among these is the stability of mRNA and the ubiquitous nature of RNases which digest it (Ibberson et al. 2009; Hanson et al. 2009; Jung et al. 2010)**.** RNases are generally small, stable molecules that function at neutral pH, and can be difficult to inactivate (Gilman 2002; Brown et al. 2004)**.** RNases are not only present within cells, where they function in cellular metabolism, but are part of the body's defense mechanism and are secreted into such body fluids as tears, blood, saliva and perspiration (Blank & Dekker 1981; Anon n.d.)**.** They are also produced by microorganisms. Studies have shown that RNA stability varies between tissues (Nussbaumer et al. 2006) and is dependent on storage conditions (Tanner et al. 2002)**.** Consequently, digestion of mRNA by RNases from within cells and body fluids, as well as possible RNase contamination of evidence from collection through analysis by perspiration (fingertips) and/or microorganisms, has the potential to limit the usefulness of this method. Even partial degradation of a sample can be detrimental to the assay, as the entire region of amplification from the 5' primer to the 3' primer must be intact for the assay to be effective. In addition, polymorphisms at primer binding sites can reduce or eliminate amplification. Use of smaller amplicons might benefit this approach, however, unless designed to cross exons, precautions against contaminations by genomic DNA will be required.

Another potential problem is that not all body fluid "specific" markers are unique to a body fluid, although they may be predominantly expressed there (Richard et al. 2011; Shaw & Diamandis 2007). This is particularly a problem for PCR assays which can amplify extremely small amounts of nucleic acids, and therefore potentially detect signals from contaminating tissues/body fluids. Consequently, quantitative RT-PCR assays using several markers for each body fluid will need to be established in order to construct body fluid "mRNA fingerprints." Recent work by Hass et al. has moved in this direction (Haas et al. 2014).

Currently, mRNA body fluid assays sensitivity appears to be nearly one order of magnitude less than that attained by MALDI mass spectrometry for blood, saliva and semen (Yang et al. 2013).

ii) ***MicroRNA Assays –-*** In recent years several laboratories have explored the possibility of using micro-RNAs (**miRNA**) for body fluid identification (Hanson et al. 2009; Zubakov et al.

2010; Wang et al. 2012). miRNAs are a class of small (18-24 nucleotides in length), non-coding RNAs that primarily function in the regulation of protein expression by binding to the 3' untranslated region of mRNAs and blocking translation (Shomron et al. 2009). It is estimated that ~30% of human mRNAs (Lewis et al. 2005) are regulated by miRNAs and that a single miRNA can bind ~200 different mRNA transcripts (Lim et al. 2005). Conversely, a single transcript can bind more than one miRNA (Lim et al. 2005). (The complexity of this regulatory system can make unambiguous interpretation of data difficult, see below.) Over 700 miRNAs have been identified in humans, and it is predicted that the human genome contains about 1,000 different species (Berezikov et al. 2005).

Because miRNA expression patterns have been shown to vary between tissues they have become a focus of both clinical and forensics research, the former to distinguish diseased from non-diseased tissues (a relatively simple task), the latter to distinguish body fluids. However, because it is the pattern of multiple commonly expressed miRNAs, not the presence of a single uniquely expressed miRNA that distinguishes tissues and body fluids, data interpretation requires the evaluation of a complex fingerprint. The techniques typically used to study miRNAs are reverse transcription followed by quantitative PCR or microarrays; both methods require care and attention to detail.

To date two groups have examined the five most common encountered forensic body fluids (blood, saliva, semen, menstrual blood and vaginal fluid) and have identified multiple, but *different* miRNA markers (Hanson et al. 2009). Recently, Wang et al. reexamined three of the published markers using more rigorous analytical tools, but could confirm only one, miR-16 in blood.

It is not clear if these dissimilar findings are the result of technical differences (RT-qPCR vs microarray) or inherent biological obstacles, i.e. the actual sources of miRNAs. For example, Pigati et al. (2010) have demonstrated the release of different miRNAs from normal and diseased mammary epithelial cells into the blood, milk and ductal fluids of women. One of the miRNAs released from malignant cells was miR-451, which was previously identified as a blood and menstrual blood miRNA marker (Hanson et al. 2009). Release of different miRNAs into the circulatory system occurs in both lactating and non-lactating women, and has also been shown to occur in lung and prostate cancers, and gliomas (Pigati et al. 2010). Further, Wang et al. have demonstrated altered miRNA expression profiles in seminal plasma from patients with unobstructed azoospermia, asthenozoospermia and oligospermia when compared to normal controls. Finally, evidence from the literature demonstrates that epigenetic effects (in response to chemotherapeutic drugs (Saito et al. 2006; Scott et al. 2006) and trauma (Mor et al. 2011) can significantly influence miRNA expression profiles in cell culture.

These data demonstrate that both normal and diseased cells routinely release miRNAs into blood and other body fluids, and pose at least a theoretical problem of body fluid identification in the presence of disease, trauma and perhaps even changes that can occur with age, puberty, menstrual cycle, therapeutic treatments, as well as with gender and ethnicity. At the moment, most of these are hypothetical possibilities and may not preclude use of miRNAs as body markers. However, because miRNA markers are not unique to the function of specific body fluids (as are hemoglobin to blood, and semenogelin to semen), and because they are not always the most abundant miRNAs found in particular body fluids (Hanson et al. 2009), they

may prove difficult to validate. For these same reasons, the use of miRNAs may also pose significant obstacles for identifying body fluid mixtures.

Conflicting data have also been published regarding the stability of miRNAs. One study, entailed "Robust microRNA stability in degraded RNA preparations from human tissue and cell samples" (Jung et al. 2010), appears to contradict a second "RNA degradation compromises reliability of microRNA expression profiling" (Ibberson et al. 2009). However, it does appear that miRNAs are inherently more stable than mRNA, likely due to their small size which makes them less of a target for RNases.

iii) ***DNA Methylation Analysis for Forensic Tissue Identification –-*** Recent evidence from the literature suggests that DNA methylation patterns may be different in different tissues (Eckhardt et al. 2006) and that these differences may be useful for establishing a forensic DNA methylation assay for tissue identification (Frumkin et al. 2011). The advantages of such an assay would be that it uses amplification and detection technologies similar to those already available in DNA forensic labs, and that DNA is less labile than RNA (see above). While there is promise for this method, there are also many hurdles that need to be overcome.

The principle of the test - that it is possible to identify unique methylation patterns for each tissue and body fluid - will require significant testing in very large numbers of individuals. This is because it is already well established that methylation patterns differ with: age (Gronniger et al. 2010), gender (El-Maari et al. 2007), ethnicity (Adkins et al. 2011), health and disease (Saito et al. 2006; Andraos et al. 2011) as well as environmental conditions (Vercelli 2004) and diet (Choi & Friso 2010). Methylation patterns can also differ between the same tissues in different individuals, as well as within the same tissue of a single individual (Egger et al. 2004). Consequently, it will be necessary to identify "invariable" tissue/body fluid specific methylation sites so that a unique tissue/body fluid specific pattern can be routinely recognized across all the variables described above.

In addition, of the three technologies commonly in use for methylation detection: endonuclease-based, affinity-based and bisulfite-based, the latter is the current gold standard. However, methylcytosine conversion (i.e. C to U) of the bisulfite based test does not currently reach 100%, and DNA degradation during bisulfite treatment can be as great as 90%. Such extensive degradation can be problematic when DNA input is limited (Grunau et al. 2001; Ehrich et al. 2007). Consequently, for this technique to be used for routine forensic testing, conversion and recovery rates will need to be improved. Finally it seems unlikely this method will lend itself to species identification, as amplicons would likely require different primers and methylation patterns may be very different.

• RATIONALE FOR RESEARCH: The revolution that has swept proteomics over the past decade rivals what occurred in genomics in the previous decade. Much of this progress is due to advances in mass spectrometry instrumentation which is much faster, simpler, more accurate and more sensitive than previous generations of instruments (Lill et al. 2007). Similar advances have occurred in nano-liquid chromatography and equally as important is the power of bioinformatics to rapidly search and recover annotated protein information from ever increasing databases (Lill et al. 2007). While these advances have propelled proteomics to the forefront of basic research and pharmaceutical discovery (Brewis & Gadella 2010; Isaaq et al. 2011), they

are also being eyed for routine rapid testing in the clinic (Righetti et al. 2005; Sparbier et al. 2009), particularly as a diagnostic tool to identify diseases and monitor therapeutic treatments (Remily-Wood et al. 2011) from easily obtainable body fluids including: blood (Hachani et al. 2011), saliva (Bencharit et al. 2011), semen (Thacker et al. 2011) and vaginal fluid (Zegels et al. 2009).

These same MS tools are directly applicable to the forensic identification of body fluids and we have used them, with the support of NIJ, to establish a rapid, sensitive and confirmatory test for blood, semen and saliva (Yang et al. 2013). Below we present our results for its use for menstrual blood and vaginal fluid.

SECTION 1 – MALDI TOF/TOF ANALYSIS OF MENSTRUAL BLOOD – The aim was to assess menstrual blood candidate markers in a large number of women of varying ages, ethnicities and oral contraceptive use over the entire length of their periods, and vaginal fluid samples taken one week prior to and one week following their periods. Marker expression analysis was aimed at identifying menstrual blood and vaginal fluid proteins that are consistently found in all women at all times during menstruation and in all samples of vaginal fluid respectively. Detailed methods of sample collection, processing and mass spectrometry conditions are given below followed by analyses.

- BODY FLUID COLLECTION METHODS: Menstrual blood samples were obtained from 44 volunteers on each day of their periods (approved by the New York City Department of Health and Mental Hygiene's Internal Review Board). Samples were collected anonymously in sterile, individually packaged, disposable menstrual cups (Softcup, Evofem Inc. San Diego, CA). Women were asked to keep cups in place for a minimum of four but no longer than twelve hours. When cups were removed they were inserted with contents into sterile 50 ml conical centrifuge tubes (Corning, Tewksbury, MA), placed in small black plastic bags and stored at -20^{o}C. When all samples had been collected, they were brought to the laboratory and stored at -80^{o}C until analyzed. Venous blood was taken from volunteers on the first day of their periods, unless this occurred on a weekend or holiday, in which case it was taken on the first weekday opportunity thereafter. Venous blood was obtained by finger lance and collected in sterile tubes containing EDTA as anticoagulant. Samples ranged between ~50 and 500 µl. A total of 193 menstrual blood and 45 blood samples were obtained. Vaginal fluid was collected from these same 45 women (using the same menstrual cup procedure) on the 13th and 19th days following the first day of their periods – equivalent to approximately one week following the end of their last period and one week prior to their next period (assuming a 28 day menstrual cycle) thus capturing the proliferative and secretory stages of the uterus respectively (DeSouza et al. 2005). One woman of child bearing age using the long-acting, progestin-only contraceptive Depo-Provera experienced no bleeding during her menstrual cycle, a common effect of this contraceptive (Westhoff 2003). Two vaginal fluid samples and one venous blood sample were obtained from this volunteer one week apart. In addition, five postmenopausal women were asked to use menstrual cups to collect two vaginal fluid samples at one week intervals. Venous blood was taken from these women at the time of the first vaginal fluid collection.

• BODY FLUID COLLECTION RESULTS

I) *LENGTH OF PERIODS*: The number of days of menstruation ranged from between two and five for the 44 volunteers, with only two women (5%) with periods of two days and the vast majority (61%) with five day periods (Table 1).

TABLE 1 LENGTH OF PERIODS 44 WOMEN

DAYS OF PERIOD	1	2	3	4	5
NUMBER OF INDIVIDUALS	0	2 (5%)	6 (14%)	9 (20%)	27 (61%)
NUMBER OF MENSTRUAL BLOOD SAMPLES	0	4	18	36	135

TOTAL MENSTRUAL BLOOD SAMPLES = 193

II) *VOLUNTEER DEMOGRAPHICS*: The 50 volunteers ranged in age from 21 to 59 years. A demographic overview (Table 2) and more detailed breakdown (Table 3) are provided below. In short, volunteers represented four ethnic groups: 4 Asians (8%), 13 Blacks (26%), 4 Hispanics (8%), and 29 Whites (58%). Five women (10%) were post menopausal (3 Black, 2 White) and ranged in age from 49-59 years. In women of childbearing age, 18 (36%) reported using oral contraception (25-39 years of age) when samples were collected. One volunteer (aged 30, White) used the NuvaRing (estrogen/progesterone), and one volunteer (aged 22, Black) used Depo-Provera (progesterone only, by injection); 19 (38%) reported using no contraceptive method.

Table 2 Demographics Overview

Ethnicity	Number
Asian	4 (8%)
Black	13 (26%)
Hispanic	4 (8%)
White	29 (58%)

Age	Number
21-30	28 (56%)
31-40	14 (28%)
41-50	4 (8%)
51-59	4 (8%)

Contraception	Number
oral	16 (32%)
oral/condom	2 (4%)
condom	11 (22%)
other	2 (4%)
NA	19 (38%)

Post Menopausal	Age
5 (10%)	49-59

Table 3 Demographics Specifics

Ethnicity	Number	Age	Contraception	Post Menopausal
Asian	4 (8%)	26-39	1 oral, 3 NA*	None
Black	13 (26%)	22-55	1 oral, 4 condom, 1 other, 7 NA*	3 (age 49, 55)
Hispanic	4 (8%)	25-30	1 oral, 3 condom	None
White	29 (58%)	21-59	13 oral, 2 oral/condom, 4 condom, 1 other, 9 NA*	2 (age 58, 59)

Age	Number	Ethnicity	Contraception	Post Menopausal
21-30	28 (56%)	5 Black, 2 Asian, 4 Hispanic,17 White	11 oral, 5 condom, 2 oral/condom, 2 other, 8 NA*	No
31-40	14 (28%)	4 Black, 2 Asian, 8 White	5 oral, 6 condom, 3 NA*	No
41-50	4 (8%)	2 Black, 2 White	NA*	1
51-59	4 (8%)	2 Black, 2 White	NA*	4

Contraception	Number	Age	Ethnicity	Post Menopausal
oral	16 (32%)	25-39	1 Black, 1 Asian, 1 Hispanic, 13 White	No
oral/condom	2 (4%)	25-27	2 White	No
condom	11 (22%)	25-37	4 Black, 3 Hispanic, 4 White	No
other	2 (4%)	22-30	1 Black, 1 White	No
NA*	19 (38%)	21-59	7 Black, 3 Asian, 9 White	5

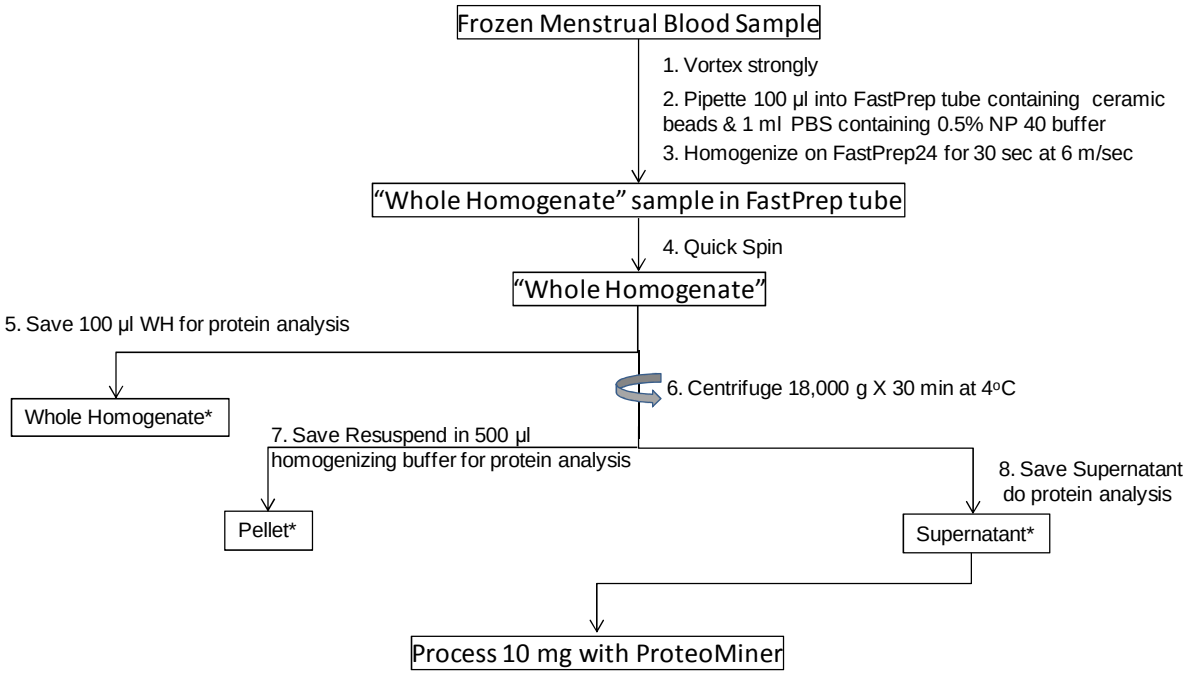
Post Menopausal	Age	Number	Ethnicity	Contraception
Yes	49-59	5 (10%)	3 Black; 2 White	NA
No	21-46	45 (90%)	10 Black, 4 Asian; 4 Hispanic, 27 White	16 oral, 2 oral/condom, 11 condom, 2 other, 14 NA*

*NA = Not Applicable, no contraceptive use.

• MENSTRUAL BLOOD SAMPLE PREPARATION METHODS: Menstrual cups were thawed on ice. Many had both liquid and membranous fragments. Samples were mixed with a rubber policeman and aspirated with a 1 ml pipette tip (cut to increase orifice size) equally into 2 ml screw cap tubes and weighed. It was not uncommon that samples from the first and particularly the last days of a woman's period had less material than days in between and often more vaginal mucus (white/tan secretion) than blood or membranous uterus lining (blood red). After transfer to tubes samples were stored at -80^{o}C until use. For analysis (see schematic Figure 1), tubes were thawed on ice, vigorously vortexed to help break apart membranous tissue

fragments and 100 µl quickly pipetted (again with pipette tips cut to increase orifice size so as not to exclude membranous fragments) into 2 ml screw cap homogenizing tubes containing 1.4 mm ceramic beads (MP Biomedical, Santa Ana, CA). One ml of ice-cold homogenizing buffer (PBS containing 0.5% NP-40 with Halt Protease Inhibitor Cocktail (Thermo Scientific, Waltham, MA)) was added and samples homogenized in a "bead beater" reciprocating at 6 m/sec for 30 sec (FastPrep24, MP Biomedical). Tubes were spun for 5 sec to remove homogenate from caps and samples transferred (without ceramic beads) to 1.5 ml microfuge tubes. Tubes were vigorously vortexed and 100 µl saved as the "whole homogenate". The remaining sample was spun at 18,000 x g for 30 min at 4°C. Supernatants were transferred to new 1.5 ml microfuge tubes and pellets resuspended in 500 µl 4% SDS. Venous blood samples were treated similarly.

Figure 1 Schematic of Sample Preparation Workflow



* Determine protein content and run 10 µg on Agilent chip

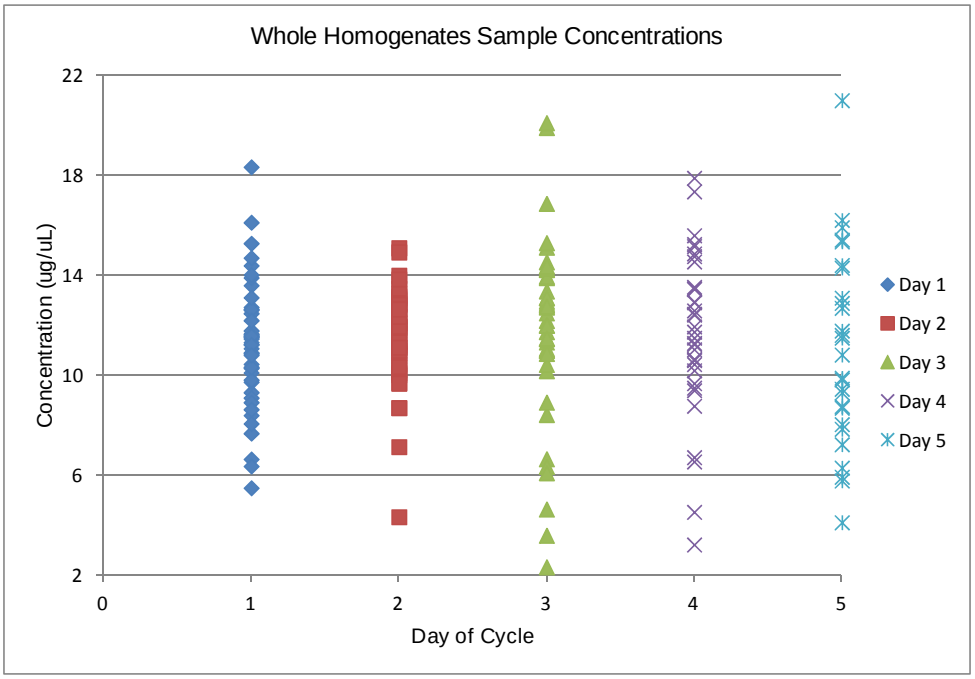
• PROTEIN QUANTITATION RESULTS

I) *VARYING LENGTHS OF MENSES*: Total protein for whole homogenates (WH), supernatants (SUP) and pellets were measured for all menstrual (193) and venous blood (45) samples using the bicinchoninic acid (BCA) assay (Pierce, Life Technologies, Grand Island, NY).

Average protein content of whole homogenates was ~11.5 µg/µl for each of the five days of the women's periods with perhaps less on the first and last days and the greatest variation found on day 3 (see Figure 2a & 2b). While the range of values on each day is relatively wide, the decline of protein concentration near the end of menses is more obvious in sample

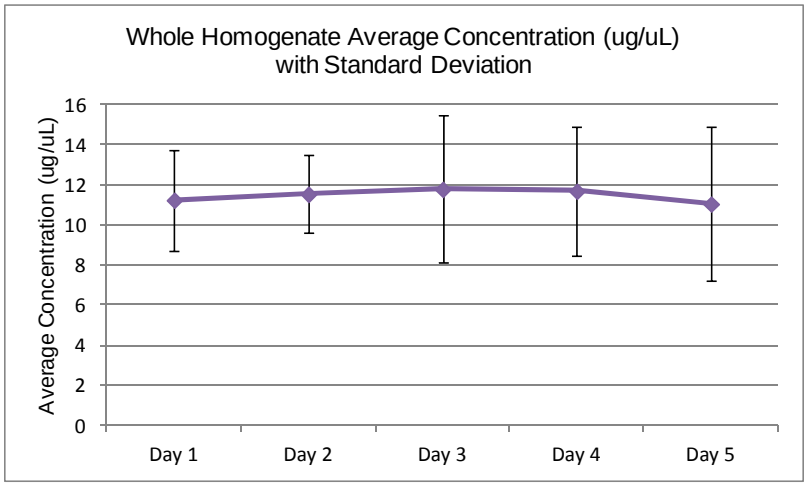
supernatants (Figures 3a & 3b) which are the fractions used for menstrual blood specific marker analysis. (Indeed, the bulk of samples with insufficient protein for MS analysis were from the last days of women's periods (Table 4)). Finally, because these data include women whose periods ended on days 2, 3 and 4 with those whose periods ended on day 5 (~61% of volunteers, Table 1), and because of the reduction of protein concentration at the end of menses, samples from women with 5 day periods only were analyzed to more accurately evaluate protein concentration over menses (see below).

Figure 2a Menstrual Blood Whole Homogenate Protein Concentration on Each Day of the Women's Periods



Day 1 N=44; Day 2 N=44; Day 3 N=41; Day 4 N=35; Day 5 N=27

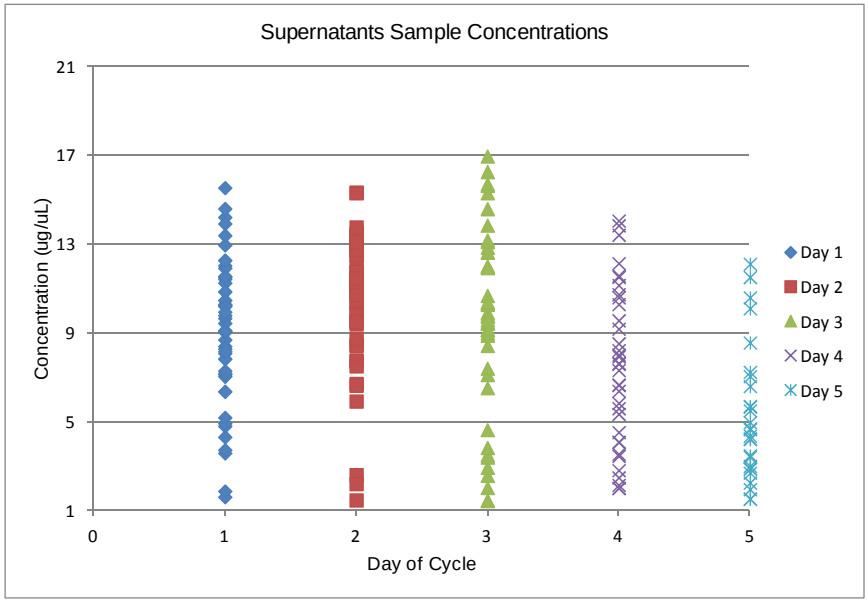
Figure 2b Menstrual Blood Whole Homogenate Protein Content on Each Day of the Women's Periods with Standard Deviations.



Day 1 N=44; Day 2 N=44; Day 3 N=41; Day 4 N=35; Day 5 N=27

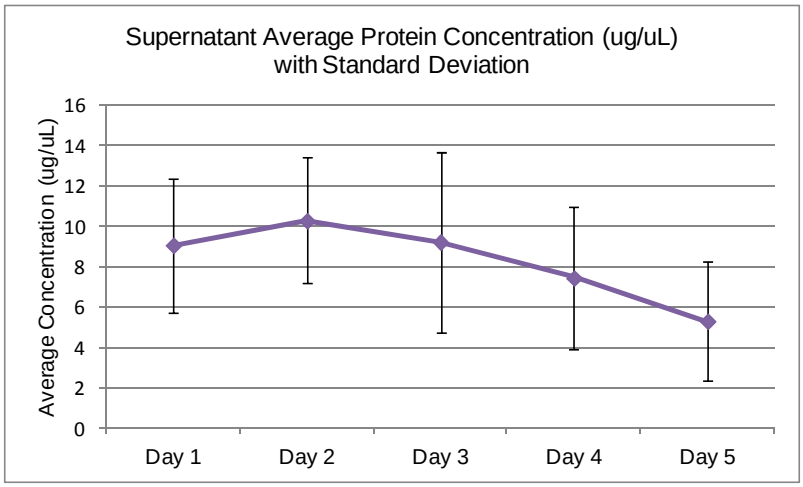
The average protein content of supernatants was 8.3 µg/µl over the five days of the women's periods. However, while protein concentrations in whole homogenates showed only a modest decline over time, supernatant concentrations showed a significantly greater reduction in protein concentration after day three (Figures 3a & 3b), suggesting that protein solubilization by the homogenizing buffer (0.5% NP-40 in PBS) was less effective in these samples. One possible reason for this is that vaginal mucus represented a greater proportion of samples collected on the last days of menstruation than earlier on, and that the mucus is more difficult to solubilize than blood and uterine membranes.

Figure 3a Menstrual Blood Supernatant Protein Content on Each Day of the Women's Periods



Day 1 N=44; Day 2 N=44; Day 3 N=41; Day 4 N=35; Day 5 N=27

Figure 3b Menstrual Blood Supernatant Protein Content at Each Day of the Women's Periods with Standard Deviations



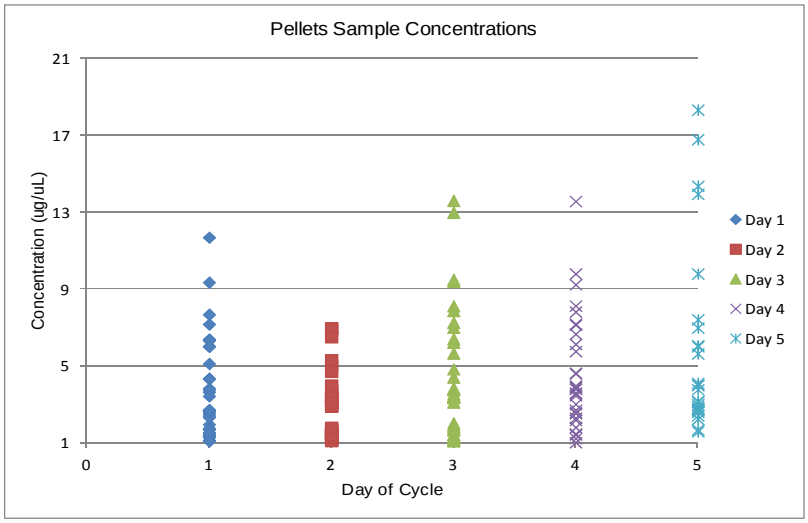
Day 1 N=44; Day 2 N=44; Day 3 N=41; Day 4 N=35; Day 5 N=27

TABLE 4 NUMBER OF SUPERNATANT SAMPLES WITH INSUFFICIENT PROTEIN FOR CPLC ANALYSIS BASED ON DAY OF MENSES

First Day	Middle Days	Last Day	Last 2 Days	Last 3 Days
1	5	14	4	3
1st of 5	2nd of 5 (X 1)	2nd of 2 (X 1)		
	3rd of 5 (X 3)	3rd of 3 (X 2)	3rd & 4th of 4 (X 1)	3rd, 4th & 5th of 5 (X 1)
	4th of 5 (X 1)	4th of 4 (X 3)	4th & 5th of 5 (X 1)	
		5th of 5 (X 8)		

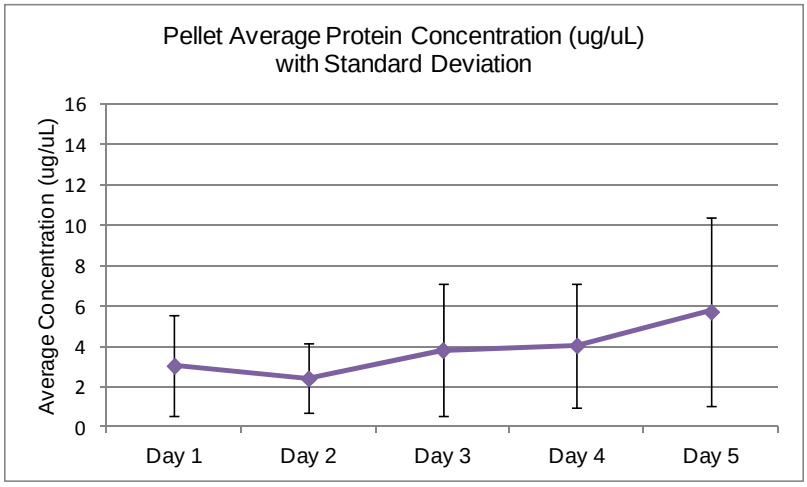
As expected, the average concentration of protein recovered from pellets (Figure 4a & b) was less than found in supernatants (Figure 3a &b above) and appeared to increase over the length of women's periods. These data are consistent with the decline in protein content seen in the supernatants, and again may be due to the increased contribution of vaginal mucus observed during the last days of menses and the difficulty of solubilizing it with the non-ionic detergent NP-40.

Figure 4a Menstrual Blood Pellet Protein Concentration During Each Day of Menses



Day 1 N=44; Day 2 N=44; Day 3 N=41; Day 4 N=35; Day 5 N=27

Figure 4b Menstrual Blood Pellet Protein Content During Each Day of Menses with Standard Deviations.



Day 1 N=44; Day 2 N=44; Day 3 N=41; Day 4 N=35; Day 5 N=27

II) *FIVE DAY MENSES ONLY*: As protein concentration varies over the duration of menses, combining data from all menstrual blood samples (i.e. those of women with 2 through 5 day cycles) in a single analysis (as above) could give a skewed measure of protein concentration over time. Consequently, women with five day cycles (the majority of samples (61%)) were analyzed separately (Figures 5-7). Results show that the overall trends (i.e. a slight decline in protein concentration of whole homogenates near the end of menses, a significant drop in supernatant protein concentrations at that time, and an increase in protein concentration in pellets at the end of menses) are similar to those identified above. However, as expected, there is a slight increase in average protein concentration at days 1, 2 and 3 in the 5 day only samples, as the lower protein concentrations of shorter periods have been removed.

Figure 5a Menstrual Blood Whole Homogenate Protein Concentration on Each Day of the Women's Periods - 5 Day Menses Samples Only

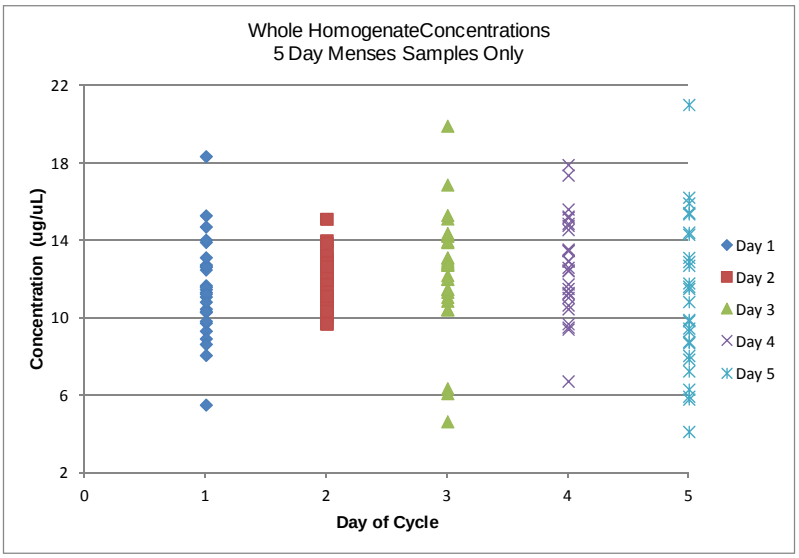


Figure 5b Menstrual Blood Whole Homogenate Protein Content during Each Day of Menses with Standard Deviations - 5 Day Menses Samples Only

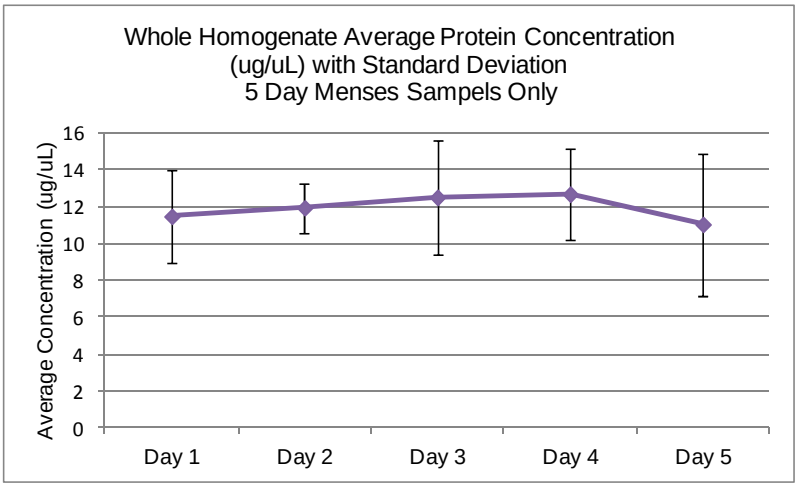


Figure 6a Menstrual Blood Supernatant Protein Concentrations on Each Day of the Women's Periods - 5 Day Menses Samples Only

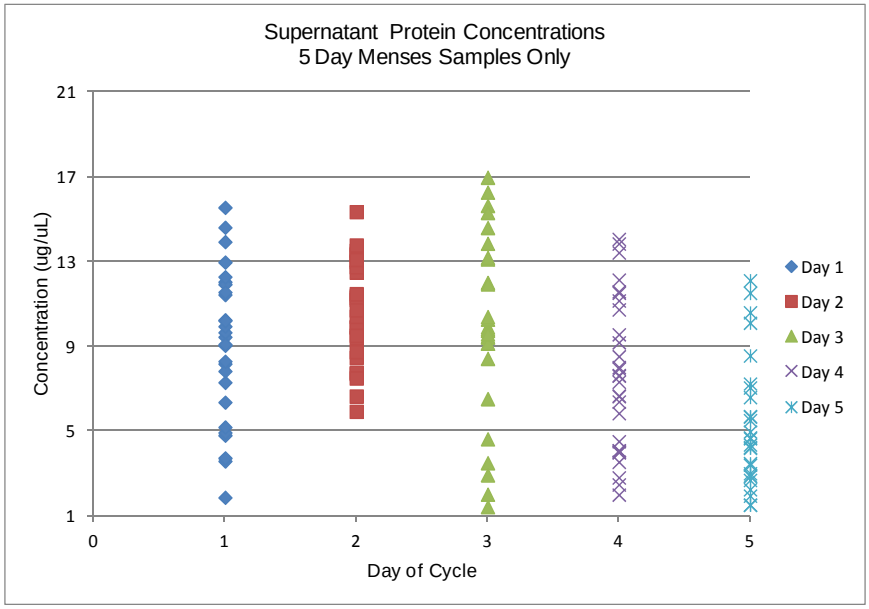


Figure 6b Menstrual Blood Supernatant Protein Content during Each Day of Menses with Standard Deviations – 5 Day Menses Samples Only

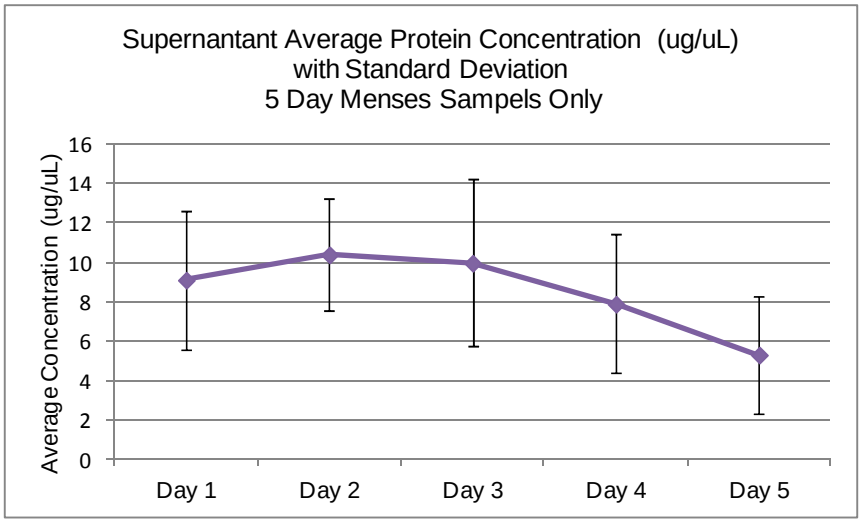


Figure 7a Menstrual Blood Pellet Protein Concentrations on Each Day of the Women's Periods - 5 Day Menses Samples Only

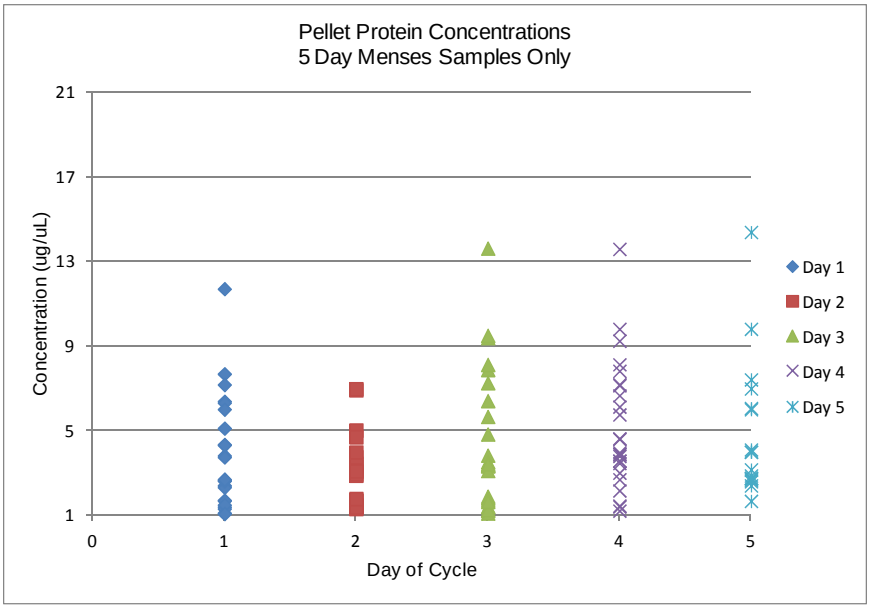
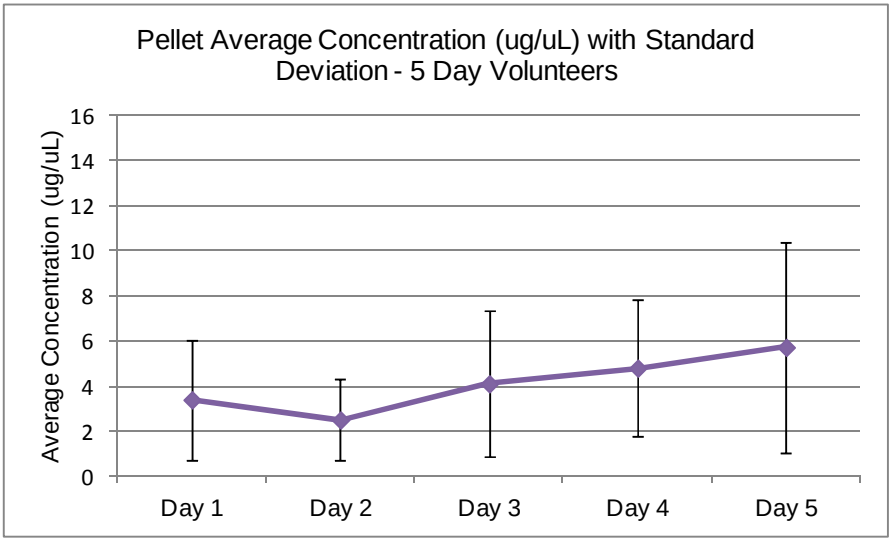


Figure 7b Menstrual Blood Pellet Protein Content during Each Day of Menses with Standard Deviations – 5 Day Menses Samples Only



• REDUCTION OF PROTEIN DYNAMIC RANGE: Supernatant fractions were used for all menstrual blood marker analyses. However, because menstrual blood specific marker proteins represent only a fraction of total supernatant protein their detection can be masked by the abundant blood proteins that contribute to menstrual blood (e.g. hemoglobin, albumin, immunoglobulins, etc.). Consequently, to improve menstrual blood marker protein detection, the dynamic range of menstrual blood proteins (i.e. the difference between the largest and smallest amounts of specific proteins in a sample) was reduced using combinatorial ligand peptide chromatography (CPLC).

CPLC employs random six-mer amino acid aptamers (synthesized with all 20 amino acids in all possible combinations) linked to a bead matrix packed in micro-spin columns. Each bead has thousands of the same aptamer bound to it and all beads are present in equal amounts. As different proteins have different affinities for different bead aptamers, each protein will bind most strongly to its cognate bead(s). This means that proteins of low abundance will accumulate as they pass through the column, while proteins of high abundance (after saturating their associated bead(s)) will pass through the column effectively reducing the overall protein dynamic range by concentrating low abundance proteins and shedding high abundance proteins. In order to ensure reproducibility and maximum dynamic range compression column binding capacity needs to be saturated.

While CPLC saturation varies with the type and complexity of the protein sample used, the manufacturer suggests a starting point of 10 mg protein for the pre-loaded CPLC spin columns (ProteoMiner, Bio-Rad Hercules, CA). Due to the differences in menstrual blood samples (day of the menstruation, volumes collected, blood/mucus content, etc.) protein content varied widely (see above). However, all samples with greater than 2 mg protein remaining after Agilent gel-chip analysis were used to determine CPLC column binding capacity. Of the 193 menstrual blood samples collected, 29 (15%) had less than 2 mg protein. The remaining 164 (85%) samples were processed through CPLC (Table 5). Applied protein (supernatants) ranged from ~2-12 mg (average 7.8 mg; standard deviation 2.7 mg).

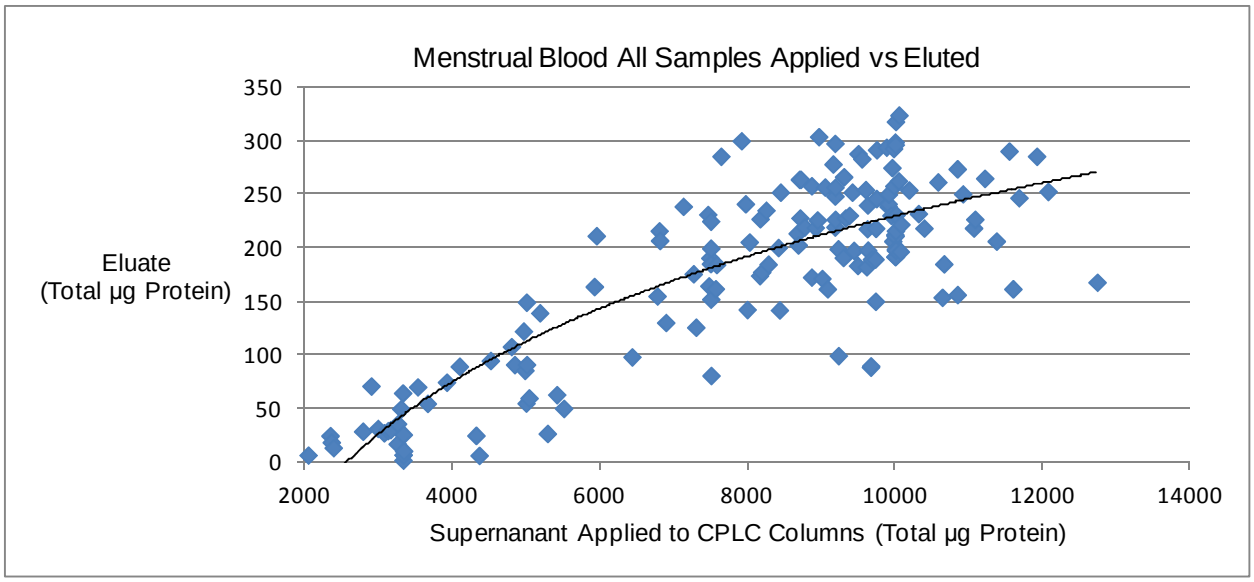
Saturation was determined by comparing the amounts of eluted protein against the amounts applied – i.e. once saturation occurs column binding should be maximized and eluate protein should plateau. Figure 8 shows a scatter plot of all menstrual blood samples.

Table 5 Menstrual & Venous Blood Samples

Sample	Collected	Extracted	Evaluated by SDS MiGE (Agilent)	Run on CPLC		
				Insufficient Protein	164 (85%)	
					<7,500 µg	≥7,500 µg
Menstrual Blood	193	193	193	29 (15%)	52* (32%)	112* (68%)
Venous Blood	45	45	45	-	13 (29%)	32 (71%)

***** Four ≤7,500 and one >7,500 columns failed; there was insufficient protein to repeat CPLC.

Figure 8 Determining CPLC Column Saturation for Menstrual Blood



A breakdown of the data (Figure 9) show a relatively linear relationship between applied and eluted samples up to approximately 7,500 µg applied protein (linear regression). Figure 10 shows most samples with more than 7,500 µg applied protein appear to cluster between 150 and 300 total µg protein recovered, suggesting that saturation occurs at about 7,500 µg applied protein.

Figure 9 CPLC Linear Binding Range for Menstrual Blood

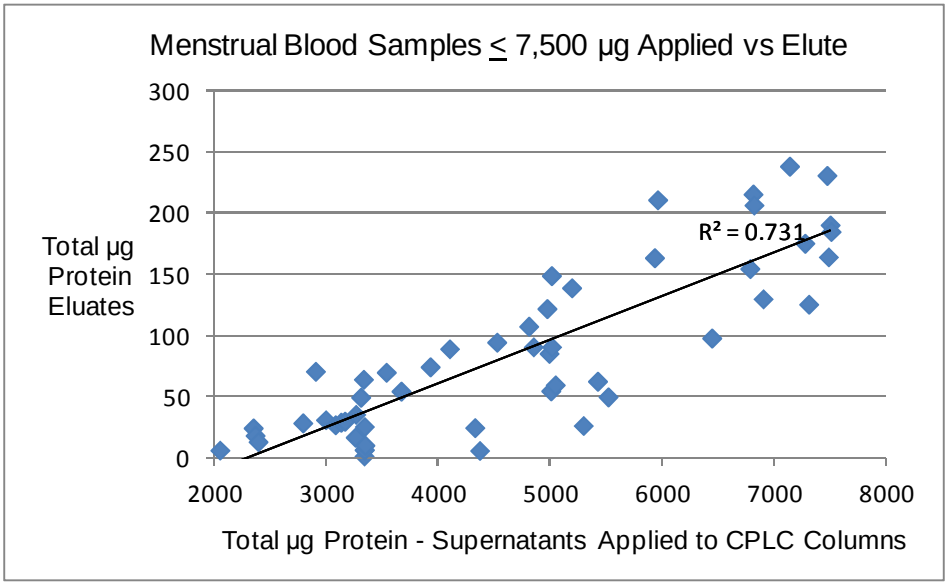
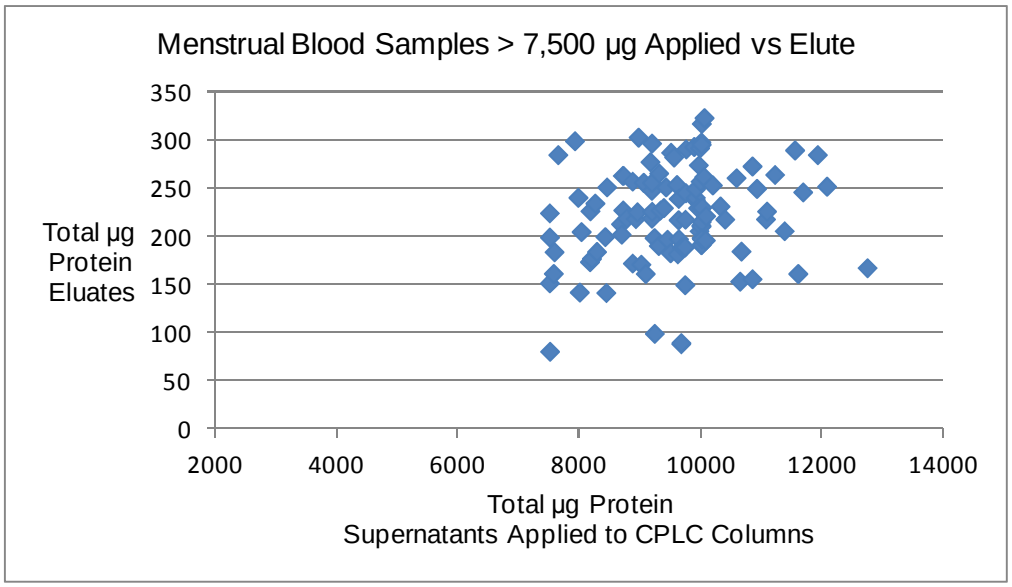


Figure 10 CPLC Protein Saturation for Menstrual Blood



Venous blood showed similar results (Table 5 above, Figures 11-13) with a linear relationship between applied protein to eluted protein up to approximately 7,500 µg and CPLC column saturation at protein amounts greater than 7,500 µg.

Figure 11 Determining CPLC Column Saturation for Venous Blood

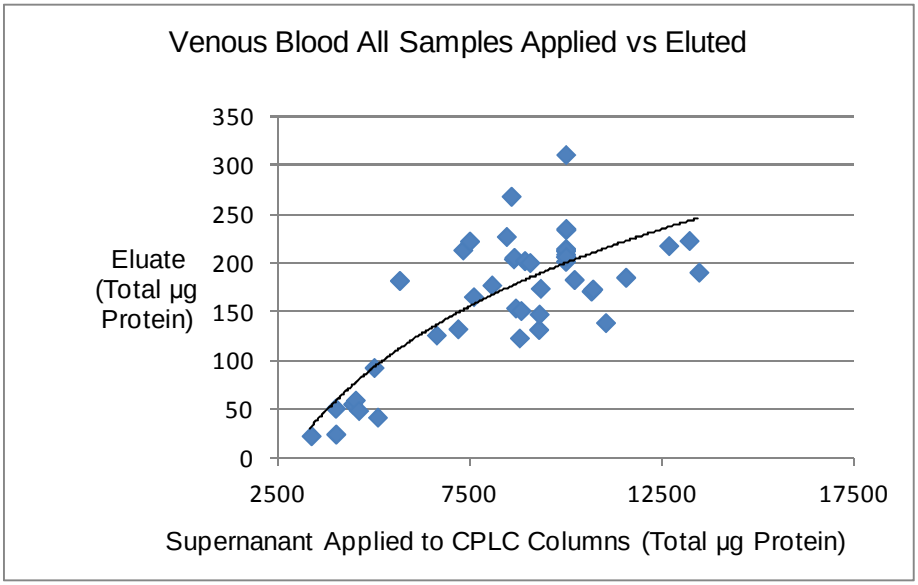


Figure 12 CPLC Linear Binding Range for Venous Blood

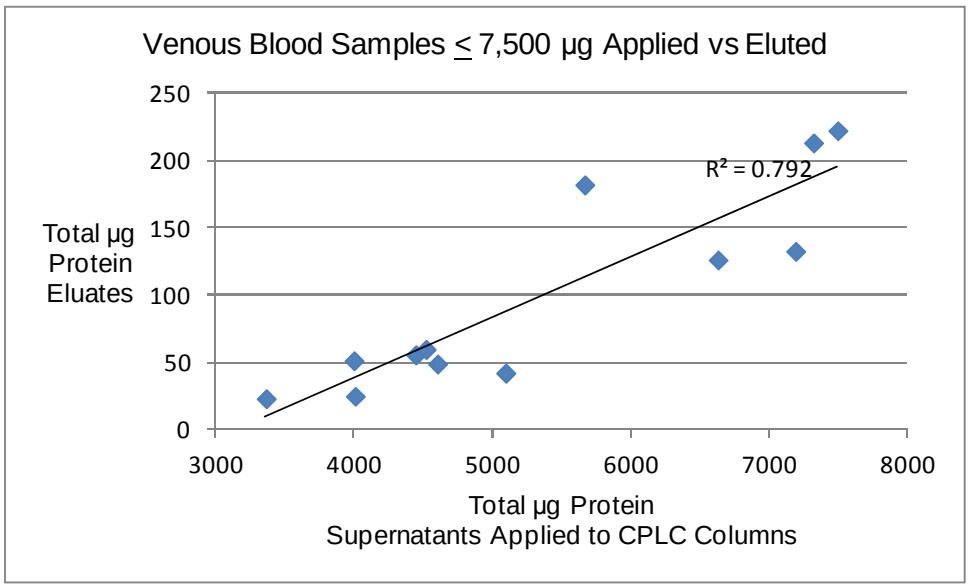
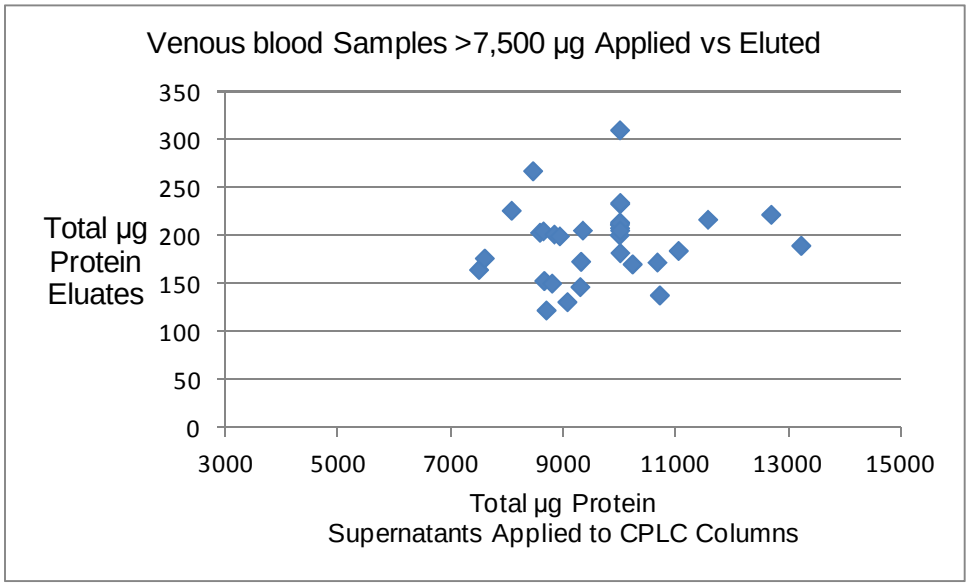


Figure 13 CPLC Protein Saturation for Venous Blood



Thus, assuming CPLC column saturation for menstrual blood and venous blood requires ~7,500 µg protein, 107 (55%) of 193 menstrual blood samples and 32 (71%) of 45 blood samples met saturation criteria (see Table 6). On average, approximately 2.3 percent of applied protein was recovered in CPLC elutes from samples where more than 7,500 µg protein were applied to the columns. Approximately 2.0 percent of applied protein was recovered from venous blood samples where ≥7,500 µg protein were applied (Table 6).

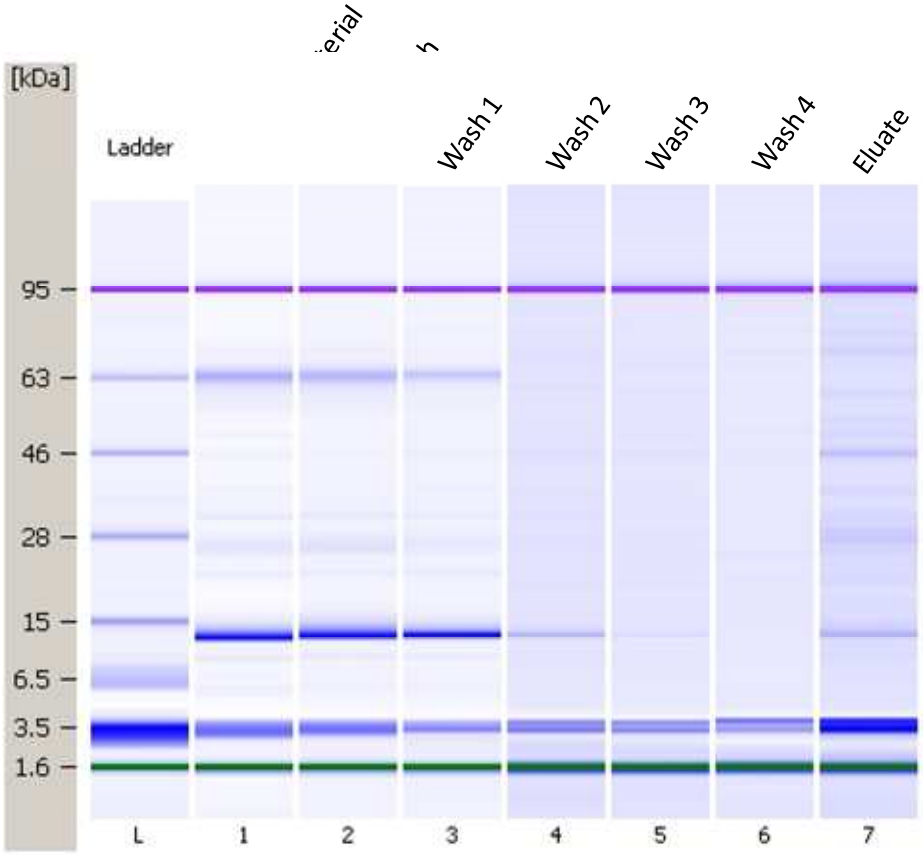
Table 6 Statistical Analyses of Menstrual and Venous Blood Samples Meeting CPLC Saturation Criteria (>7,500 µg Applied Protein)

Body Fluid/ Fraction	Number of Samples >7,500 µg Protein	Range µg Protein	Mean µg Protein	Std Deviation	Median µg Protein
Menstrual Blood Applied	107	7,502 - 12,744	9,500	1,100	9,600
Menstrual Blood Eluted	107	81 - 234	220	50	230
Venous Blood Applied	32	7,595 - 13,470	9,900	1,400	10,000
Venous Blood Eluted	32	165 – 191	200	40	200

• SDS MICROFLUIDIC GEL ELECTROPHORESIS (SDS MiGE): Qualitative analysis of all samples (menstrual and venous blood) was performed by microfluidic gel electrophoresis (Agilent Technologies, Wilmington, DE) in order to evaluate CPLC protein dynamic range reduction.

Figure 14 shows a typical example of protein dynamic range reduction of a menstrual blood sample following CPLC as analyzed by SDS MiGE. In this pseudo SDS PAGE image equal volumes of starting material (supernatant), pass-through, washes and eluate were loaded onto an Agilent Protein 80 microfluidic gel chip with resolution between 5-80 kDa. The chip shows that the bulk of applied supernatant proteins do not bind to the CPLC column as evidenced by the similar band intensities of the loading material, pass through and even the first wash. This confirms that the columns were overloaded and likely saturated as is necessary for protein dynamic range reduction. The most prominent bands in these samples (Lanes 1-3) at ~15 and 63 kDa are hemoglobin (Hb) and albumin respectively. As may be seen, little protein is released from the CPLC column following the forth wash. The eluent shows a dramatic reduction in dynamic range. The albumin band has nearly disappeared and Hb is significantly reduced. Importantly, many protein bands that were not detectable in the starting material are now easily seen.

Figure 14 Protein Dynamic Range Reduction of Menstrual Blood by Combinatorial Ligand Peptide Chromatography



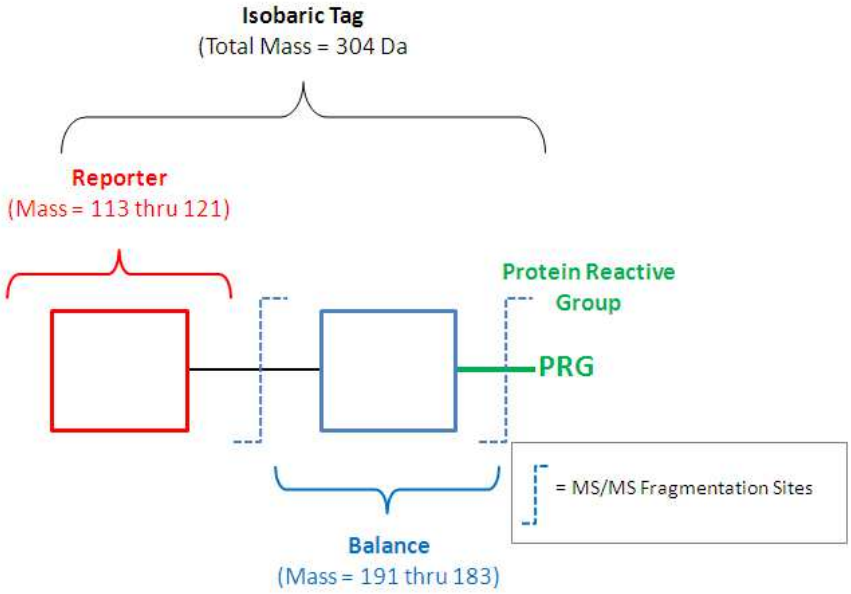
Pseudo SDS-PAGE image of menstrual blood sample 118, Day 4 evaluated by SDS microfluidic gel electrophoresis before and after protein dynamic range reduction. 9.2 mg of supernatant protein (Lane 1) were applied to the CPLC column; 99.2 µg protein (~1.1%) were eluted (Lane 7).

• PEPTIDE ISOBARIC LABELING FOR MULTIPLEXING AND QUANTITATION: Following CPLC all menstrual and venous blood elutes were reduced, alkylated and digested overnight with trypsin. Prior to trypsin digestion, disulfide bonds are reduced with TCEP and alkylated with iodoacetamide to prevent cysteine cross linking again and open proteins for more thorough digestion. Following digestion, 20 µg of tryptic peptides from each sample were labeled with isobaric tags (iTRAQ 8plex reagent, SCIEX, CA) following the manufacturer's protocol. Isobaric labeling is used to barcode all peptides in a single sample with a unique mass tag that can be identified by a mass spectrometer. Isobaric barcoding serves two functions: i) it allows the mixing of differently barcoded samples for simultaneous analysis by mass spec (i.e. multiplexing) and therefore significantly reduces instrument use time, and ii) it allows for quantitative comparisons between the different barcoded samples. Thus, this technique permitted us to combine all menstrual blood sample days from each volunteer in a single mass spec analysis and therefore to compare the levels of specific menstrual blood markers during each day of that women's period. Further, by choosing a single volunteer's menstrual blood sample (volunteer 282, day 2) as a common standard and tagging it with a unique barcode (113

Da), we were able to spike all samples with this standard sample and consequently compare all mass spec runs to one another.

i) *How Isobaric Barcoding Works*: The isobaric barcode used (iTRAQ from SCIEX, Framingham, MA) is schematically represented in Figure 15. It is composed of three functional groups: i) the *Protein Reactive Group* which allows the tag to covalently react with primary amines from N-terminal or lysine on each peptide, ii) the *Reporter* which is the actual barcode composed of eight different masses (113–119 Da plus 121 Da), and iii) the *Balance* which, in combination with the reporter always results in an isobaric mass of 304 Da for all iTRAQ labels (e.g. Reporter 113 + Balance 191, Reporter 114 + Balance 190, etc.). The importance of

Figure 15 Schematic Representation of iTRAQ Isobaric Tag



keeping the mass of the *Reporter* plus *Balance* constant (isobaric) is to ensure that the same marker peptide from different samples (e.g. different days of menstruation from one individual) migrate together in the mass spectrometer which separates and identifies peptides based on their mass. Subsequent fragmentation of the co-migrating peptides in the mass spec's collision chamber releases the *Reporters* (barcodes) which are then detected. A comparison of each barcode's peak height (intensity) shows the relative quantity of the marker peptide present in the different samples. The relative intensities of these ions are proportional to the amount of this peptide in each volunteer's sample. These relative peptide quantities are reported by Protein Pilot software (AB Sciex) as ratios relative to a labeled standard in the run. Protein ratios are obtained using a weighted average of the ratios of all the peptides contributing to the protein identification.

An example of iTRAQ multiplexing and quantitation of barcoded samples is shown in Figure 16. Here seven menstrual blood samples from two volunteers were individually labeled with iTRAQ isobaric tags containing reporters 114 - 119 plus 121 Da. After labeling, samples were combined, peptides separated by nano-HPLC and then analyzed by MALDI TOF/TOF. Figure 16a is the mass spec signal of a single peptide with a mass of 1967.97 Da. MS/MS of

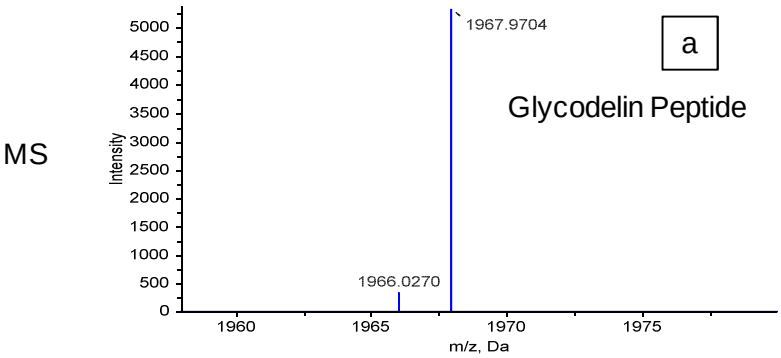
this peptide (Figure 16b) reveals it is from the menstrual blood marker protein glycodelin with the amino acid sequence *VLVEDDEIMQGFIR*. The bracket below the MS/MS peak at ~118 Da is expanded in Figure 16c and shows each of the reporter ions 113 - 119 plus 121.

The intensities of the reporter ions indicate the relative concentrations of the glycodelin menstrual blood marker in each woman's sample. Thus, for example, in volunteer 135 peptide intensities ranged from ~ 100 to 350 during the four days of her period. In volunteer 158, intensities ranged from ~200 to 550. In all experiments reporter ion 113 was used to label a single menstrual blood sample (volunteer 282, day 2) that was spiked into all mass spec runs in order to allow quantitative comparisons of marker proteins across runs.

• NANO-HPLC*:* After isobaric labeling, 2 µg of up to seven differently labeled samples plus the universal control (volunteer 282 day 2, isobaric label 113) were combined and peptides separated by reverse phase C-18 nano-HPLC using a 5% - 45% linear gradient of acetonitrile over 3 hours. HPLC runs were performed on Ultimate 3000 system (Dionex, CA). Tryptic peptides were desalted using an inline C18 trap column (300 µm x 5 mm, 5 µm, 100 Å, Dionex). Separation was performed on a PepMap100 RP C18 column (75 µm× 250 mm; 3 µm, 100 Å, Dionex) using a 5–45% linear ACN gradient for 160 min at flow rate of 300 nl/min (solvent A 2% acetonitrile, 0.1% TFA; and solvent B 98% acetonitrile, 0.1% TFA). A Dionex Probot was used to spot approximately 1 µl (1:1 volumes eluate:matrix) on a 384 SCIEX MALDI plate (matrix: α-cyano-4-hydroxycinnamic acid at a concentration of 5 mg/ml in 60% ACN, 40% water, 0.1% TFA).

• MASS SPEC ANALYSIS: Peptides were analyzed on a Sciex 4800 MALDI TOF/TOF instrument. MS data was acquired at a laser repetition rate of 200 Hz with 600 laser shots/spectrum (50 laser shots/sub-spectrum). MS/MS data was acquired at 200 Hz in 1 kV MS/MS mode with 3200 laser shots/spectrum (40 laser shots/sub-spectrum) with the following TOF/TOF Series Explorer Stop Conditions: Maximum shots per spectrum 3200; minimum shots per spectrum 1000; number of MS/MS fragments 8; S/N of each fragment 75. Top 30 strongest peaks were selected to do MS/MS analysis. Peptides were identified and quantified via automated database searching on raw data using ProteinPilot 4.5 (SCIEX, CA) against Swiss-Prot Human protein database (40,670 Protein Entries, April 2009). Paragon was selected as processing method. iTRAQ 8plex was selected as sample type. Methyl methanethiosulfonate and trypsin were selected as alkylation reagent and enzyme respectively. In order to achieve a more comprehensive search, through ID was selected in search effort.

Figure16 Multiplexing Isobaric Barcode Labeled Samples for Mass Spec Analysis



MS and MS/MS spectra of peptide VLEGVWIFYELPNY from menstrual blood marker protein glycodelin labeled with isobaric iTRAQ tagging reagent. **a**, MS spectrum of the peptide; **b**, MS/MS spectrum which is used to determine the peptide's sequence; **c**, expanded MS/MS spectrum of iTRAQ barcodes showing the relative amounts of peptides from each sample. Barcode Key: 113 = menstrual blood standard used in all experiments (volunteer 280, day 2); 114-117 = volunteer 135, days 1, 3-5; 118, 119 & 121 = volunteer 158, days 1-3. (Volunteer 135 day 2 had insufficient protein for MS analysis.)

• MENSTRUAL BLOOD / VENOUS BLOOD SAMPLE ANALYSIS: All menstrual and venous blood samples processed through CPLC with sufficient protein were analyzed by mass spectrometry regardless of the amount of protein applied to CPLC columns (Table 7). For menstrual blood, a total of 127 samples (~77%) of the 164 samples processed by CPLC were analyzed by MS/MS. This represented 39 of the 45 women from whom samples were collected. Of these, 102 (~80%) had come from samples with ≥7,500 µg protein applied to the CPLC columns; the remaining 25 samples (~20%) had less than 7,500 µg protein applied to CPLC columns. For venous blood, 40 samples (89%) of the 45 samples processed by CPLC were analyzed by MS. Here, 32 samples (~71%) had ≥7,500 µg protein applied to the CPLC columns, while the remaining 13 samples (~29%) had less than 7,500 µg protein applied to CPLC columns.

As may be seen in Table 7 (also discussed above), the amount of protein eluted from samples with less than 7,500 µg protein applied to CPLC columns was typically less than those samples that achieved saturation. For menstrual blood, eluates from samples with <7,500 µg applied protein were on average ~61% of those that had ≥7,500 µg applied. For venous blood, eluates with <7,500 µg were ~85% of those that reached saturation. While it is likely that protein dynamic range reduction is less in samples that did not saturate CPLC columns, these samples were brought through HPLC MS/MS in order to determine if menstrual blood markers could still be identified when sample size is limited.

Table 7 Statistical Analyses of Menstrual and Venous Blood Samples Processed through CPLC and Analyzed by Mass Spectrometry

Body Fluid/ µg Protein *Applied to CPLC*	Number of Samples (% Total)	Eluted Protein Range (µg)	Mean Protein Range (µg)	Std Deviation	Median Protein Range (µg)
Menstrual Blood <7,500	25 (~20%)	60-239	137	57	130
Menstrual Blood ≥ 7,500	102 (~80%)	81-324	225	49	227
Venous Blood <7,500	13 (~29%)	23-222	167	70	183
Venous Blood ≥ 7,500	32 (~71%)	123-311	196	39	202

A total 1,551 proteins were identified from the 167 samples (127 menstrual blood and 40 venous blood) that were processed through CPLC, HPLC and MALDI TOF/TOF. Of these 889 were found only in menstrual blood. However, not all of these proteins were found in all samples and many were housekeeping proteins found in other body fluids. Consequently, criteria were established for selecting menstrual blood marker candidates based on the following: i) frequency – i.e. the number of women in which candidate markers were identified and ii) specificity – i.e. their abundance in menstrual blood compared with other body fluids.

For frequency, proteins found in half or more of all women were examined as potential markers as it is possible that "missing" marker expression in some women might be due to specific conditions (e.g. age, ethnicity or oral contraception.) Specificity was determined by comparing candidate menstrual blood marker proteins against the proteomes of 12 body fluids (saliva, seminal fluid, serum/plasma, vaginal fluid, amniotic fluid, bronchoalveolar lavage, cerebrospinal fluid, milk, nipple aspirate, synovial fluid, tear and urine) using the *Sys-BodyFluid* database (http://lifecenter.sgst.cn/bodyfluid/home.jsp). Using these criteria, twenty-two unique or menstrual blood enriched marker proteins were identified and are listed in Table 8.

Table 8 Unique or Enriched Menstrual Blood Marker Proteins

No.	Protein ID	Protein Name	No. Peptides	No. Women	Saliva	Seminal Fluid	Serum/ Plasma	Vaginal Fluid	Amniotic fluid	Broncho-alveolar Lavage	Cerebro spinal Fluid	Milk	Nipple Aspirate	Synovial Fluid	Tear	Urine
1	P09466	Progestagen-associated endometrial protein (PAEP aka Glycodelin)	2	ALL			*									
2	P06737	Glycogen phosphorylase, liver form (PYGL)	5	ALL												
3	Q9UBC9	Small proline-rich protein 3 (SPRR3)	2	ALL												
4	P05141	ADP/ATP translocase 2 (ADT2)	4	ALL			*									
5	P08729	Keratin, type II cytoskeletal 7 (K2C7)	4	ALL												
6	P14314	Glucosidase 2 subunit beta (GLU2B)	1	37												
7	Q9UIV8	Serpin B13 (SPB13)	8	35												
8	Q00325	Phosphate carrier protein, mitochondrial (MPCP)	2	35												
9	P22749	Granulysin (GNLY)	1	33												
10	P43307	Translocon-associated protein subunit alpha (SSRA)	1	32												
11	P39656	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit (OST48)	3	32												
12	P12236	ADP/ATP translocase 3 (SLC25A6)	4	31												
13	Q9UBG3	Cornulin (CRNN)	2	29												
14	P30536	Translocator protein (TSPOA)	1	28												
15	P00403	Cytochrome c oxidase subunit 2 (COX2)	1	27												
16	P21397	Amine oxidase [flavin-containing] A (AOFA)	1	24												
17	P20962	Parathymosin (PTMS)	1	24												
18	P51571	Translocon-associated protein subunit delta (SSRD)	1	22												
19	P49755	Transmembrane emp24 domain-containing protein 10 (TMEDA)	2	22												
20	A8K2U0	Alpha-2-macroglobulin-like protein 1 (A2ML1)	5	21												
21	Q9NRM1	Enamelin (ENAM)	1	20			*									
22	P24158	Myeloblastin (PRTN3)	2	20												

* Found in plasma/serum in very low amounts.

= Unique to menstrual blood. Not found in Sys-BodyFluid database.

= Identified by NYC OCME.

= Found in Sys-BodyFluid database

= Identified by NYC OCME & found in Sys-BodyFluid database

These markers may be divided into three broad catagories: i) markers found only in menstrual blood but not found in all women, ii) markers found in all women, but also detected in

other body fluids and iii) markers enriched in menstrual blood but detected in other body fluids and not found in all women. Each category is discussed below.

I) *MARKERS FOUND ONLY IN MENSTRUAL BLOOD BUT NOT FOUND IN ALL WOMEN*: When checked against the *Sys-BodyFluid* database five menstrual blood proteins were identified that had not been reported in any of the 12 body fluids in the database. These were: 1) granulysin (#9) found in 33 women (85%); 2) dolichyl-diphosphooligosaccharide-protein glycosyltransferase 48 kDa subunit (#11) found in 32 women (82%); 3) translocator protein (#14) found in 28 women (72%); 4) amine oxidase [flavin-containing] A (#16) found in 24 women (62%); and 5) parathymosin (#17) also found in 24 women (62%). Of these only dolichyl-diphospho-oligosaccharide-protein glycosyltransferase (#11) was identified by more than one peptide. While not found in all women, when identified these markers give the greatest confidence that menstrual blood is present. However, it should be noted that while not possible on a MALDI mass spectrometer, using multiple reaction monitoring (MRM) or SWATH (a SCIEX proprietary data-independent acquisition method - *Sequential Window Acquisition of all THeoretical fragment-ion spectra*) additional peptides are likely to be detected in positive samples and, because of the sensitivity of these methods, additional samples are likely to be identified as positive.

II) *MARKERS FOUND IN ALL WOMEN, BUT ALSO DETECTED IN OTHER BODY FLUIDS*: Five proteins were consistently and robustly identified in all menstrual blood samples and can easily distinguish menstrual blood from venous blood. These proteins are: 1) progestagen-associated endometrial protein (PAEP aka Glycodelin); 2) glycogen phosphorylase, liver form; 3) small proline-rich protein 3; 4) ADP/ATP translocase 2 and 5) keratin, type II cytoskeletal 7.

Two of these proteins, progestagen-associated endometrial protein and ADP/ATP translocase 2, have been reported in comprehensive proteomic analyses of venous blood at low levels. However, we have not detected these proteins in the 45 venous blood samples used in this study (although we were looking at only one time point - i.e. the first day of menstruation). Progestagen-associated endometrial protein is an excellent marker for distinguishing menstrual blood from venous blood as it is a well documented endometrial protein and known to increase in the secretory phase of the menstrual cycle (Strauss & Barbieri 2013). However, we, and others (Pilch & Mann 2006; Chiu et al. 2007) have also detected a similar isoform of PAEP in seminal fluid. Nevertheless, the presence of PAEP in a blood stain in the absence of the specific and abundant seminal fluid markers semenogelin 1 and 2 would suffice to exclude the possibility of a mixture of semen and menstrual blood.

Except for PAEP, the four remaining proteins (Nos. 2-5 Table 8) have been identified in the Sys-BodyFluid database to be present at varying levels in saliva. (We routinely find only small proline-rich protein 3 in saliva, likely due to its high abundance.) However, here again, the absence of other saliva specific and abundant markers (e.g. alpha-amylase and cystatin-SA) would be sufficient to determine that the source of these proteins are not saliva. Finally, it should be noted that small proline-rich protein 3 is also present in and we consider a marker for vaginal fluid. However, a mixture of *venous* blood and vaginal fluid could not be mistaken for

menstrual blood as it would not have the remaining nine menstrual blood markers (Nos. 1-3, 4, 5, 9, 11, 14, 16 & 17 Table 8).

III) *MARKERS ENRICHED IN MENSTRUAL BLOOD BUT DETECTED IN OTHER BODY FLUIDS AND NOT FOUND IN ALL WOMEN*: In addition to the ten markers described above, 12 additional proteins were identified in menstrual blood that were not detected in venous blood (Table 8 Nos. 6-8, 10, 12, 13, 15, 18-22). These 12 proteins were found in 51% - 95% of all women. Nearly all were identified in the Sys-BodyFluid database in saliva. However, as described above, in the absence of alpha-amylase and cystatin-SA the contribution of these markers could not be confused for saliva. We have also identified four of these 12 proteins in vaginal fluid, two of which, small proline-rich protein 3 (No.3) and alpha-2-macroglobulin-like protein 1 (No. 20) are consistent markers for vaginal fluid. It is not possible to say whether these vaginal fluid markers are normal constituents of menstrual blood or whether they become mixed with menstrual blood as it passes through vaginal canal. However, the presence of these markers in a mixture of vaginal fluid and venous blood could not be confused for menstrual blood since the five menstrual blood specific markers (9, 11, 14, 16 and 17) would be absent.

It is important to point out that the markers identified in this study are proteins inherent in uterine function and consistent with the dynamic processes of the menstrual cycle – in particular the secretory phase (the period just prior to menstruation) and the beginning of menstruation which today is regarded as an inflammatory response (Evans & Salamonsen 2012). In the secretory phase, the endometrium has reached its full thickness and cells are preparing for blastocyst implantation through increased metabolic activity, protein synthesis and secretion. Increased levels of PAEP as well as ADP/ATP translocase 2 & 3, phosphate carrier protein and cytochrome c oxidase subunit 2 are reflective of this stage of the menstrual cycle. Lack of implantation leads to increases in proteolytic enzymes, proteins involved in apoptosis as well as proteins from the increased numbers of immune cells that are an integral part of menstruation (6–9 (Evans & Salamonsen 2012)), and, of course, proteins from the breakdown of the uterine wall. These proteins include serpin B13, granulysin, glucosidase 2 subunit beta, parathymosin and keratin, type II cytoskeletal 7.

Finally, for all menstrual blood markers identified, both those found in all women and those found only in some women, when present, there was no difference between their expression on different days of a woman's period, or on a woman's age, ethnicity or use of hormonal contraception – meaning these markers are commonly expressed during a woman's entire menstrual cycle.

• MULTIVARIATE ANALYSIS OF MENSTRUAL BLOOD: Similar to next generation DNA sequencing, mass spectrometry can produce vast amounts of complex data. For example, a single sample for proteomic analysis can yield thousands of mass spectra and identify hundreds of peptides. Consequently, hundreds of samples generate an amount of data that is no longer easily evaluated by hand. Additionally, when samples represent a variety of different conditions, e.g. menstrual blood samples taken on different days from different individuals of different ages and ethnicity using different methods of contraception (hormonal and non-hormonal), multivariate bioinformatic approaches must be applied to thoroughly interrogate the data to observe

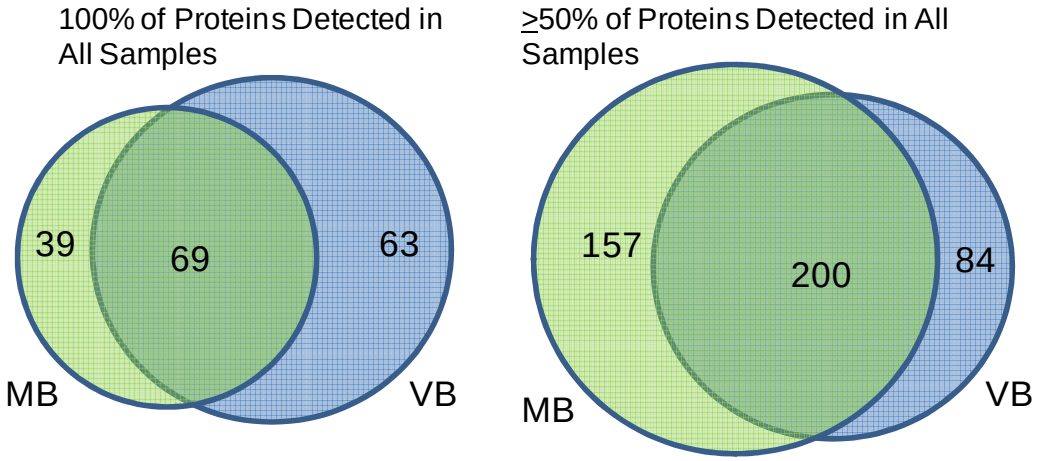
patterns, evaluate hypotheses and make discoveries. Toward this end we have employed Qlucore Omics Explorer multivariate bioinformatics software to evaluate our menstrual blood samples with respect to not only distinguishing them from venous blood, but to determine how, if at all, they differ within individuals over course of their periods and between groups with respect to oral contraception, age and ethnicity. The analysis is divided into three parts: i) comparison of menstrual blood and venous blood, ii) analysis menstrual blood over the course of menses, and iii) evaluation of menstrual blood with respect to demographic differences. A short description of the methods precedes these sections.

I) *METHODS*: As mentioned previously, menstrual blood samples were labeled with isobaric barcodes (iTRAQ reagent) for two reasons: i) to multiplex samples for increased throughput and ii) to include a common menstrual blood standard in all runs so that results could be compared and quantified to this standard across all volunteers. Consequently, data analysis of the relative amounts of all proteins identified in each sample can be compared to all other samples. This allows us to follow relative levels of expression of each protein during the course of menses as well as their expression in different demographic groups.

II) *COMPARISON OF MENSTRUAL AND VENOUS BLOODS:* Mass spectrometry analysis of each sample of menstrual and venous blood identified hundreds of proteins. Not all proteins were identified in all samples. Consequently, we evaluated data two ways: i) considering only those proteins expressed in all samples (100%) and ii) considering proteins that are expression at least in 50% of samples. Figure 17 shows that when restricting analysis to only those proteins found in all samples (100%) there are 108 proteins common to all menstrual blood samples and 132 proteins common to all venous blood samples. However, when including proteins that are present in at least 50% of menstrual blood and venous blood samples, there are 357 menstrual blood proteins and 284 venous blood proteins.

Figure 17 Numbers of unique and shared proteins identified in menstrual and venous blood when 100% or 50% of proteins are found in each sample group.

	Number of Samples	Total Proteins Identified	Proteins Detected in 100% of Samples	Proteins Detected in ≥50% of Samples
Menstrual Blood	126	1347	108	357
Venous Blood	44	738	132	284
Total Compared	170	2085	240	641



Data consists of quant measurements (ratio of the amount of specific sample proteins to the amount of the same proteins in the universal standard) in 126 menstrual blood samples and 44 venous blood samples analyzed when only all proteins are present in every sample (100%), and when ≥50% of proteins are present in every sample. Data are $\log_2$ transformed with a threshold of 0.001 and normalized to mean = 0, variance = 1.

When samples were analyzed by unsupervised hierarchical clustering using weighted average linkage menstrual and venous blood clearly segregated into distinct groups at both 100% and 50% shared protein levels. Figure 18 shows a heat map of all 170 samples with proteins detected in 100% of menstrual blood samples compared with proteins detected in 100% of venous blood samples. The order of samples was then colored yellow for venous blood and blue for menstrual blood (top row), demonstrating segregation of the two groups.

Figure 18 Unsupervised hierarchical clustering heat map of proteins detected in 100% of menstrual blood samples and 100% of venous blood samples.

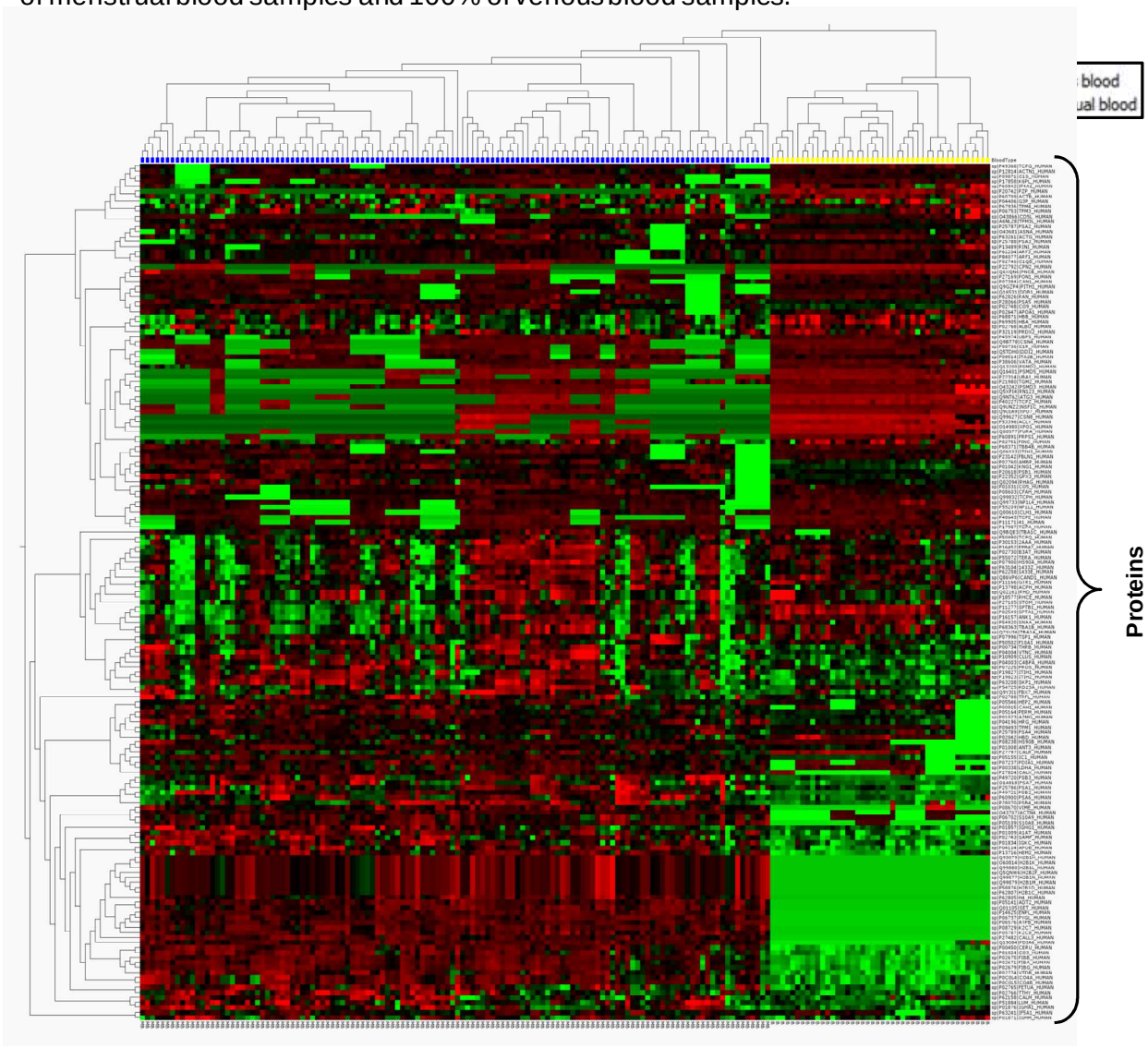
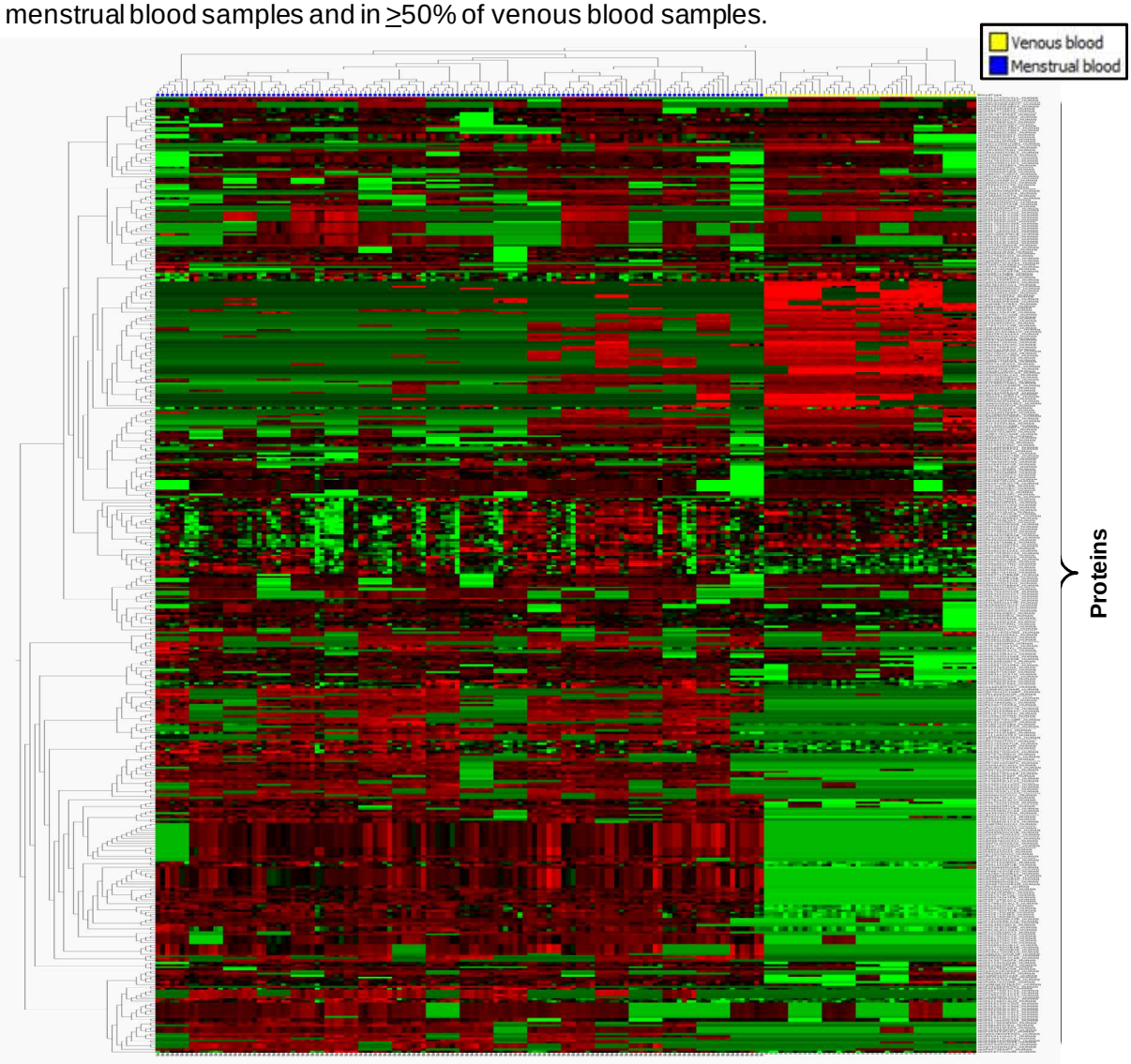


Figure 19 shows a similar heat map of all 170 samples and the set of proteins that were detected in at least 50% of menstrual blood samples or at least 50% venous blood samples analyzed by unsupervised hierarchical clustering (red show increased expression, green show decreased expression). Again menstrual and venous blood clearly segregated.

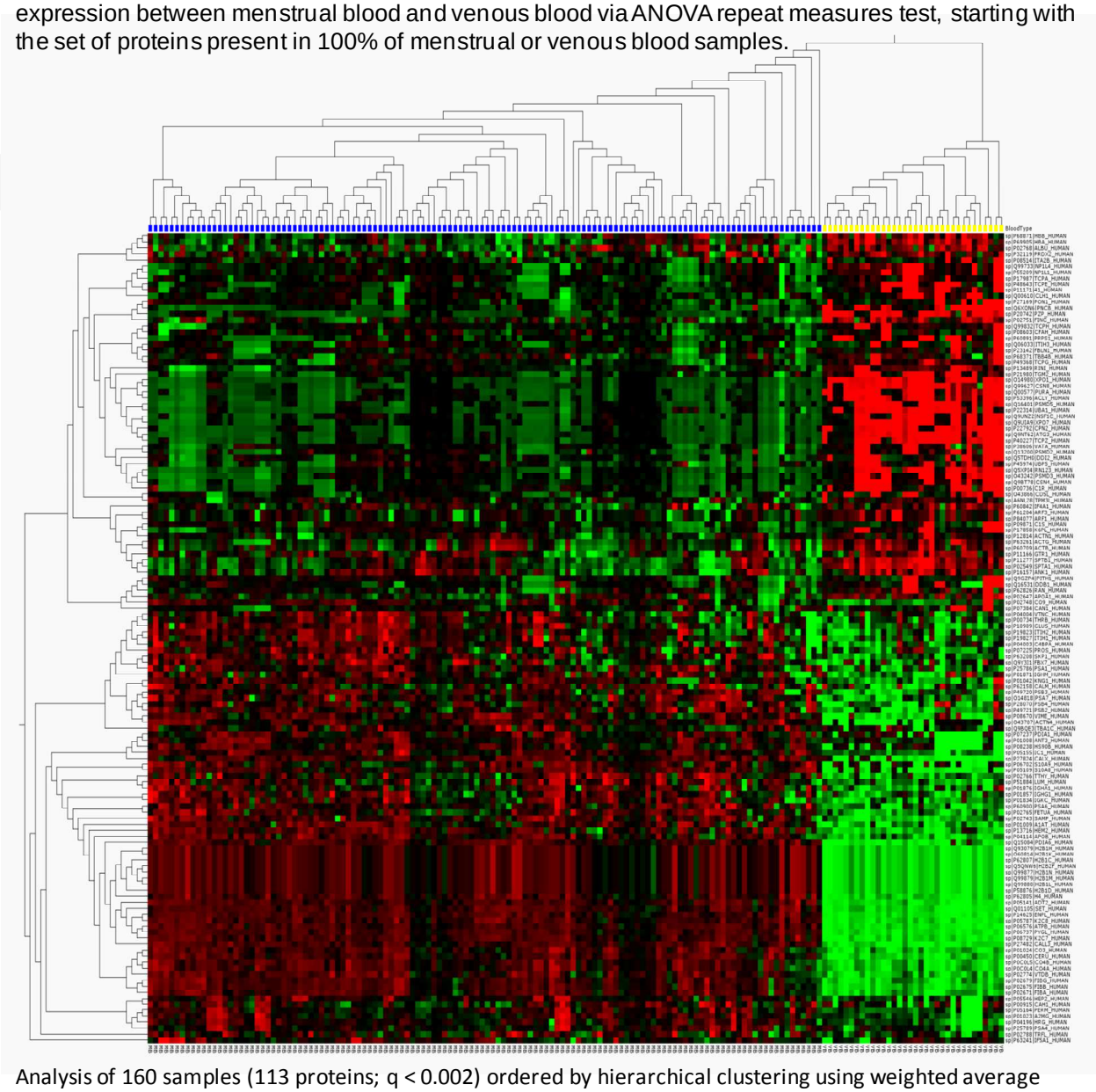
Figure 19 Unsupervised hierarchical clustering heat map of proteins detected in ≥50% of menstrual blood samples and in ≥50% of venous blood samples.



Both sample sets (100% and 50%) were further analyzed by Qlucore Omics Explorer to identify proteins with significant increases or decreases in expression to effectively distinguish the different blood samples. Analysis was performed as a two group comparison (menstrual blood and venous blood) using ANOVA to identify the most significant difference between the groups with a defined q cut off (false discovery rate at a given p value) (Benjamin & Hochberg 1995). This necessitated the removal of ten subjects for which only venous blood was analyzed either because no menstrual blood was collected (postmenopausal and one volunteer on Depo-Provera) or because menstrual blood samples did not yield sufficient protein to proceed with HPLC/MALDI analysis.

Figure 20 shows the results of this analysis performed on menstrual and venous blood proteins present in 100% of each sample. Again, unsupervised hierarchical clustering reveals a natural segregation between venous and menstrual blood (yellow and blue row at top).

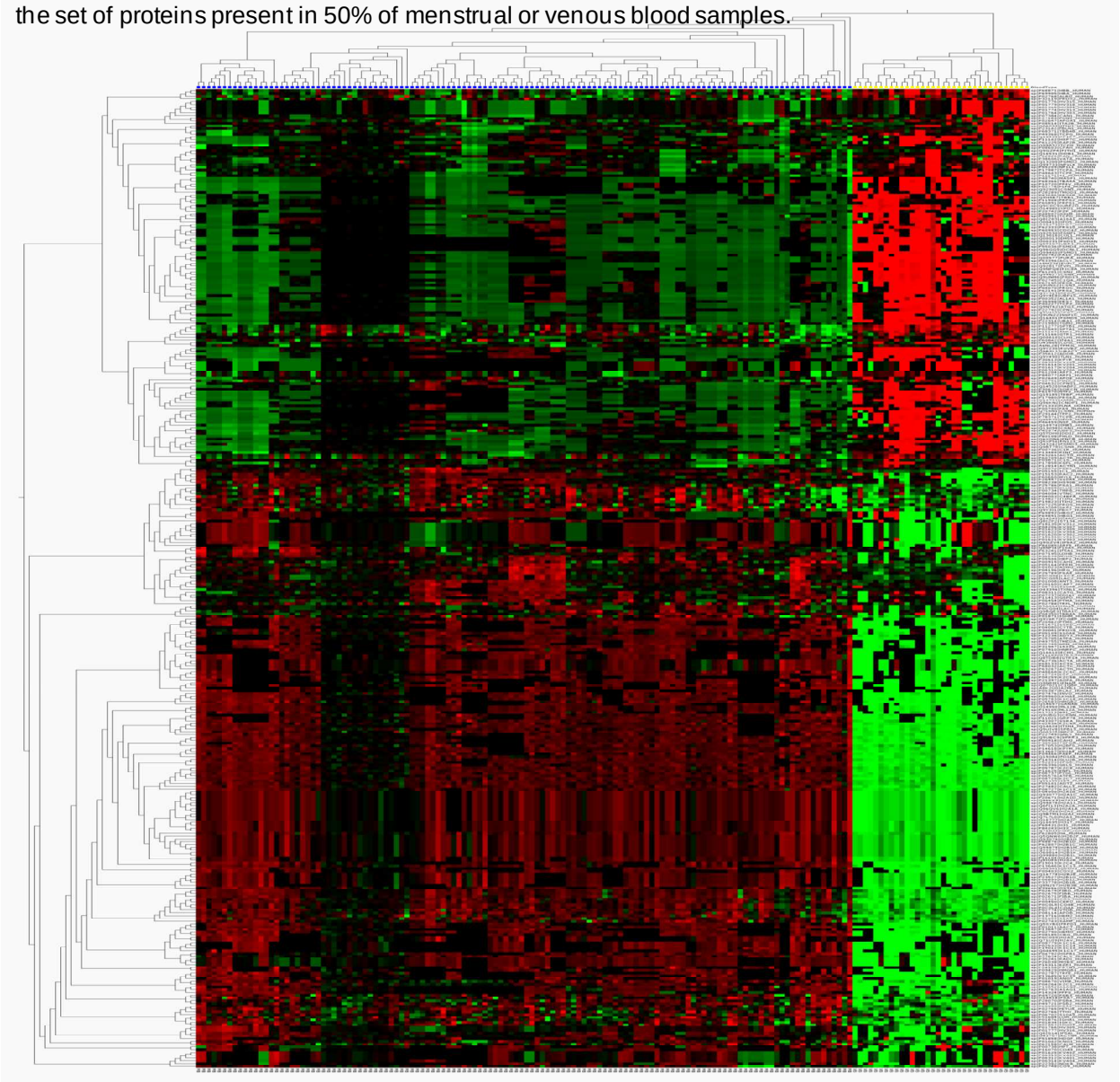
Figure 20 Heat map of an unsupervised hierarchical clustering of protein with significant difference in expression between menstrual blood and venous blood via ANOVA repeat measures test, starting with the set of proteins present in 100% of menstrual or venous blood samples.



Analysis of 160 samples (113 proteins; $q < 0.002$) ordered by hierarchical clustering using weighted average linkage. Red elevated expression. Green reduced expression.

Similarly, Figure 21 shows the results of this analysis performed on menstrual and venous blood proteins present in $\geq$50% of each sample. Again, unsupervised hierarchical clustering reveals a natural segregation between venous and menstrual blood (yellow and blue row at top; red are proteins that show a relative increase in expression and green a relative decrease in expression).

Figure 21 Heat map of unsupervised hierarchical clustering of protein with significant difference in expression between menstrual blood and venous blood via ANOVA repeat measures test, starting with the set of proteins present in 50% of menstrual or venous blood samples.



Analysis of 160 samples (289 proteins; $q < 0.002$) ordered by hierarchical clustering using weighted average linkage. Red elevated expression. Green reduced expression.

In a single direction test (Figure 21) of proteins found in more than 50% of menstrual or venous blood, 222 proteins have significantly greater expression in menstrual blood than venous blood with $q < 0.002$. These proteins are listed in Table 9 along with specific p and q values. Thirty-nine of the 40 potential menstrual blood markers identified (see Figure 23) are included in this list (marked in blue) and all have a q value less than 10^{-16}. The one exception is Mucin-5B because it is not included in the set of proteins present in 50% or more of menstrual blood samples. Mucin-5B was detected in 147 out of 378 (approximately 39%) menstrual blood

samples, in 16 of the 44 subjects (36%). It was not detected in any venous blood samples, hence its inclusion as a potentially useful marker, despite falling below the 50% cut-off.

Table 9 List of 222 proteins enriched in menstrual blood compared to venous blood (q < 0.002) using analyses of proteins present at ≥50% in each group.

Accession	Name	p-value	q-value
sp|P06576|ATPB_HUMAN	ATP synthase subunit beta, mitochondrial	9.86E-93	4.35E-90
sp|P06737|PYGL_HUMAN	Glycogen phosphorylase, liver form	3.56E-92	7.86E-90
sp|P14625|ENPL_HUMAN	Endoplasmin	2.08E-84	3.05E-82
sp|P27482|CALL3_HUMAN	Calmodulin-like protein 3	3.65E-78	4.02E-76
sp|P08729|K2C7_HUMAN	Keratin, type II cytoskeletal 7	7.22E-77	6.37E-75
sp|P05787|K2C8_HUMAN	Keratin, type II cytoskeletal 8	1.93E-74	1.42E-72
sp|P05141|ADT2_HUMAN	ADP/ATP translocase 2	1.09E-71	6.86E-70
sp|Q01105|SET_HUMAN	Protein SET	8.11E-70	4.47E-68
sp|P08727|K1C19_HUMAN	Keratin, type I cytoskeletal 19	2.13E-68	1.04E-66
sp|P13667|PDIA4_HUMAN	Protein disulfide-isomerase A4	3.74E-67	1.65E-65
sp|P14618|KPYM_HUMAN	Pyruvate kinase isozymes M1/M2	4.55E-65	1.83E-63
sp|P14314|GLU2B_HUMAN	Glucosidase 2 subunit beta	4.22E-64	1.55E-62
sp|P62807|H2B1C_HUMAN	Histone H2B type 1-C/E/F/G/I	1.02E-61	2.26E-60
sp|P58876|H2B1D_HUMAN	Histone H2B type 1-D	1.02E-61	2.26E-60
sp|Q93079|H2B1H_HUMAN	Histone H2B type 1-H	1.02E-61	2.26E-60
sp|O60814|H2B1K_HUMAN	Histone H2B type 1-K	1.02E-61	2.26E-60
sp|Q99880|H2B1L_HUMAN	Histone H2B type 1-L	1.02E-61	2.26E-60
sp|Q99879|H2B1M_HUMAN	Histone H2B type 1-M	1.02E-61	2.26E-60
sp|Q99877|H2B1N_HUMAN	Histone H2B type 1-N	1.02E-61	2.26E-60
sp|Q5QNW6|H2B2F_HUMAN	Histone H2B type 2-F	1.02E-61	2.26E-60
sp|P68431|H31_HUMAN	Histone H3.1	1.40E-61	2.58E-60
sp|Q16695|H31T_HUMAN	Histone H3.1t	1.40E-61	2.58E-60
sp|Q71DI3|H32_HUMAN	Histone H3.2	1.40E-61	2.58E-60
sp|P84243|H33_HUMAN	Histone H3.3	1.40E-61	2.58E-60
sp|Q14624|ITIH4_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H4	4.18E-61	7.37E-60
sp|P0C0S8|H2A1_HUMAN	Histone H2A type 1	8.52E-60	1.04E-58
sp|Q96QV6|H2A1A_HUMAN	Histone H2A type 1-A	8.52E-60	1.04E-58
sp|P04908|H2A1B_HUMAN	Histone H2A type 1-B/E	8.52E-60	1.04E-58
sp|Q93077|H2A1C_HUMAN	Histone H2A type 1-C	8.52E-60	1.04E-58
sp|P20671|H2A1D_HUMAN	Histone H2A type 1-D	8.52E-60	1.04E-58
sp|Q96KK5|H2A1H_HUMAN	Histone H2A type 1-H	8.52E-60	1.04E-58
sp|Q99878|H2A1J_HUMAN	Histone H2A type 1-J	8.52E-60	1.04E-58
sp|Q6FI13|H2A2A_HUMAN	Histone H2A type 2-A	8.52E-60	1.04E-58
sp|Q16777|H2A2C_HUMAN	Histone H2A type 2-C	8.52E-60	1.04E-58
sp|Q7L7L0|H2A3_HUMAN	Histone H2A type 3	8.52E-60	1.04E-58
sp|Q9BTM1|H2AJ_HUMAN	Histone H2A.J	8.52E-60	1.04E-58
sp|P02538|K2C6A_HUMAN	Keratin, type II cytoskeletal 6A	2.73E-59	3.25E-58
sp|P13646|K1C13_HUMAN	Keratin, type I cytoskeletal 13	9.39E-58	1.09E-56
sp|P52209|6PGD_HUMAN	6-phosphogluconate dehydrogenase, decarboxylating	1.26E-55	1.43E-54
sp|P62805|H4_HUMAN	Histone H4	3.02E-55	3.33E-54
sp|P00450|CERU_HUMAN	Ceruloplasmin	9.93E-55	1.07E-53
sp|P09466|PAEP_HUMAN	Glycodelin	2.59E-53	2.72E-52
sp|P06396|GELS_HUMAN	Gelsolin	9.18E-52	9.42E-51
sp|Q15084|PDIA6_HUMAN	Protein disulfide-isomerase A6	1.37E-51	1.37E-50
sp|P00918|CAH2_HUMAN	Carbonic anhydrase 2	6.65E-50	6.52E-49
sp|Q8IUE6|H2A2B_HUMAN	Histone H2A type 2-B	1.12E-49	1.05E-48

sp|P16104|H2AX_HUMAN	Histone H2A.x	1.12E-49	1.05E-48
sp|P02675|FIBB_HUMAN	Fibrinogen beta chain	5.24E-49	4.82E-48
sp|Q9UBC9|SPRR3_HUMAN	Small proline-rich protein 3	7.69E-48	6.92E-47
sp|Q9UIV8|SPB13_HUMAN	Serpin B13	1.32E-47	1.17E-46
sp|Q00325|MPCP_HUMAN	Phosphate carrier protein, mitochondrial	7.54E-47	6.52E-46
sp|P22749|GNLY_HUMAN	Granulysin	2.52E-45	2.14E-44
sp|P57053|H2BFS_HUMAN	Histone H2B type F-S	2.52E-43	2.10E-42
sp|P19105|ML12A_HUMAN	Myosin regulatory light chain 12A	2.93E-43	2.35E-42
sp|O14950|ML12B_HUMAN	Myosin regulatory light chain 12B	2.93E-43	2.35E-42
sp|P25705|ATPA_HUMAN	ATP synthase subunit alpha, mitochondrial	8.10E-43	6.38E-42
sp|P11021|GRP78_HUMAN	78 kDa glucose-regulated protein	1.23E-41	9.52E-41
sp|P0C0L5|CO4B_HUMAN	Complement C4-B	1.86E-40	1.41E-39
sp|P02671|FIBA_HUMAN	Fibrinogen alpha chain	2.88E-40	2.16E-39
sp|P43307|SSRA_HUMAN	Translocon-associated protein subunit alpha	5.87E-40	4.32E-39
sp|P01011|AACT_HUMAN	Alpha-1-antichymotrypsin	1.82E-39	1.31E-38
sp|P01024|CO3_HUMAN	Complement C3	2.61E-39	1.72E-38
sp|P02679|FIBG_HUMAN	Fibrinogen gamma chain	2.52E-39	1.72E-38
sp|P33778|H2B1B_HUMAN	Histone H2B type 1-B	2.61E-39	1.72E-38
sp|P06899|H2B1J_HUMAN	Histone H2B type 1-J	2.61E-39	1.72E-38
sp|P23527|H2B1O_HUMAN	Histone H2B type 1-O	2.61E-39	1.72E-38
sp|Q16778|H2B2E_HUMAN	Histone H2B type 2-E	2.61E-39	1.72E-38
sp|P12236|ADT3_HUMAN	ADP/ATP translocase 3	4.25E-39	2.75E-38
sp|P30536|TSPOA_HUMAN	Translocator protein	9.72E-39	6.21E-38
sp|P39656|OST48_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	2.10E-38	1.32E-37
sp|P02774|VTDB_HUMAN	Vitamin D-binding protein	1.22E-37	7.58E-37
sp|P0C0L4|CO4A_HUMAN	Complement C4-A	2.26E-37	1.38E-36
sp|P27824|CALX_HUMAN	Calnexin	2.62E-36	1.58E-35
sp|P07476|INVO_HUMAN	Involucrin	4.92E-35	2.93E-34
sp|P02790|HEMO_HUMAN	Hemopexin	6.90E-35	4.06E-34
sp|P17213|BPI_HUMAN	Bactericidal permeability-increasing protein	3.03E-34	1.76E-33
sp|P05109|S10A8_HUMAN	Protein S100-A8	4.93E-33	2.82E-32
sp|P19013|K2C4_HUMAN	Keratin, type II cytoskeletal 4	5.97E-32	3.38E-31
sp|P09960|LKHA4_HUMAN	Leukotriene A-4 hydrolase	6.27E-32	3.50E-31
sp|Q9UBG3|CRNN_HUMAN	Cornulin	6.89E-32	3.80E-31
sp|P04792|HSPB1_HUMAN	Heat shock protein beta-1	2.86E-31	1.56E-30
sp|P12724|ECP_HUMAN	Eosinophil cationic protein	4.02E-31	2.16E-30
sp|Q14697|GANAB_HUMAN	Neutral alpha-glucosidase AB	2.08E-30	1.11E-29
sp|P05783|K1C18_HUMAN	Keratin, type I cytoskeletal 18	3.49E-30	1.83E-29
sp|Q8N257|H2B3B_HUMAN	Histone H2B type 3-B	6.43E-30	3.33E-29
sp|P00403|COX2_HUMAN	Cytochrome c oxidase subunit 2	1.84E-29	9.44E-29
sp|P68032|ACTC_HUMAN	Actin, alpha cardiac muscle 1	5.26E-29	2.58E-28
sp|P68133|ACTS_HUMAN	Actin, alpha skeletal muscle	5.26E-29	2.58E-28
sp|P62736|ACTA_HUMAN	Actin, aortic smooth muscle	5.26E-29	2.58E-28
sp|P63267|ACTH_HUMAN	Actin, gamma-enteric smooth muscle	5.26E-29	2.58E-28
sp|P01009|A1AT_HUMAN	Alpha-1-antitrypsin	2.39E-28	1.16E-27
sp|P02763|A1AG1_HUMAN	Alpha-1-acid glycoprotein 1	8.20E-28	3.93E-27
sp|P31947|1433S_HUMAN	14-3-3 protein sigma	2.42E-27	1.15E-26
sp|P04114|APOB_HUMAN	Apolipoprotein B-100	3.09E-27	1.45E-26
sp|P30041|PRDX6_HUMAN	Peroxiredoxin-6	5.86E-27	2.72E-26
sp|P48668|K2C6C_HUMAN	Keratin, type II cytoskeletal 6C	1.01E-26	4.63E-26
sp|P26583|HMGB2_HUMAN	High mobility group protein B2	1.68E-26	7.62E-26
sp|Q8TDB8|GTR14_HUMAN	Solute carrier family 2, facilitated glucose transporter member 14	2.86E-26	1.27E-25

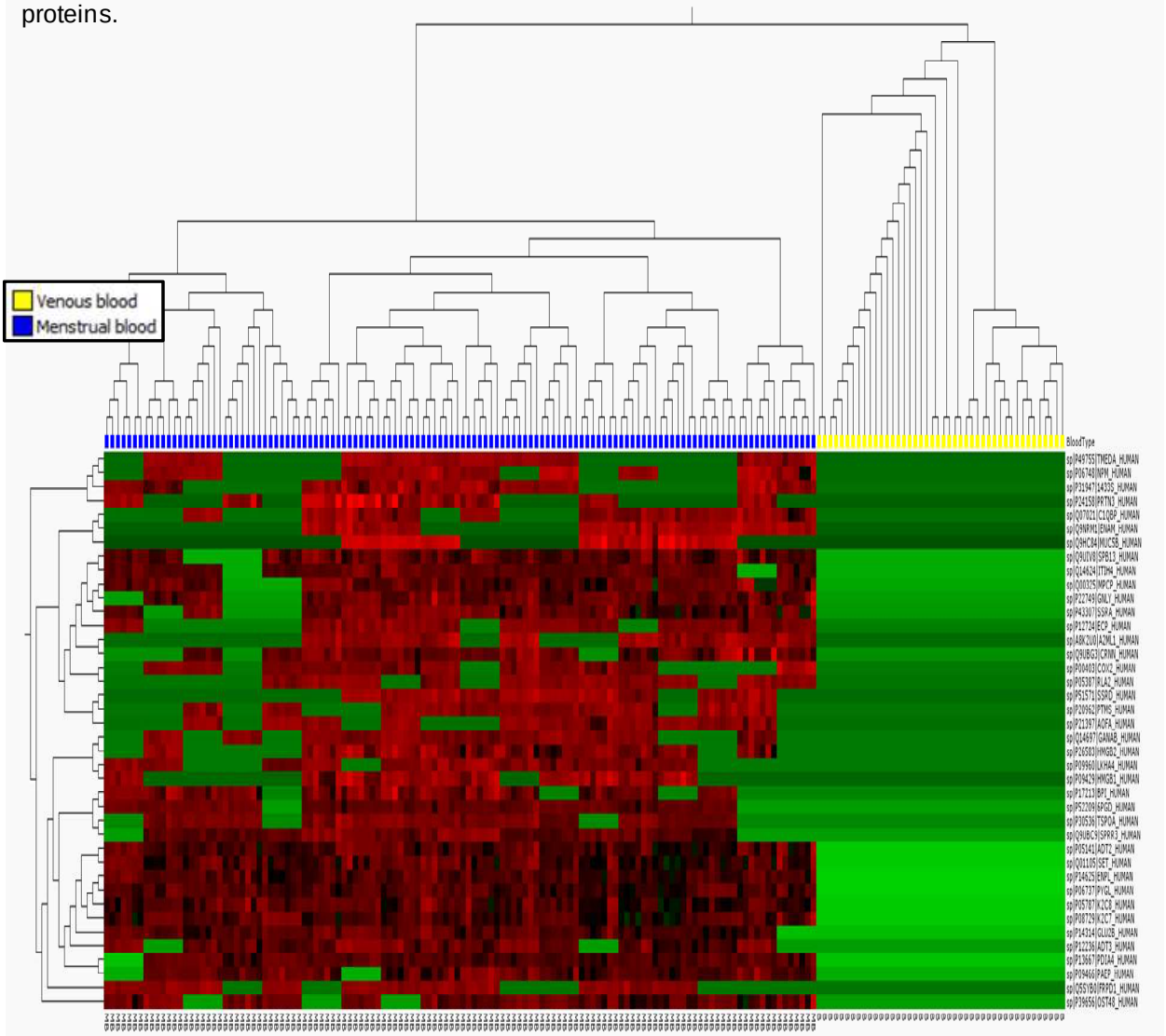
sp|P11169|GTR3_HUMAN	Solute carrier family 2, facilitated glucose transporter member 3	2.86E-26	1.27E-25
sp|Q5SYB0|FRPD1_HUMAN	FERM and PDZ domain-containing protein 1	5.72E-26	2.52E-25
sp|P13716|HEM2_HUMAN	Delta-aminolevulinic acid dehydratase	6.49E-26	2.83E-25
sp|P13645|K1C10_HUMAN	Keratin, type I cytoskeletal 10	1.45E-25	6.28E-25
sp|O00391|QSOX1_HUMAN	Sulfhydryl oxidase 1	1.57E-25	6.73E-25
sp|P08185|CBG_HUMAN	Corticosteroid-binding globulin	2.34E-25	9.94E-25
sp|P49755|TMEDA_HUMAN	Transmembrane emp24 domain-containing protein 10	2.04E-24	8.58E-24
sp|P02743|SAMP_HUMAN	Serum amyloid P-component	4.26E-24	1.77E-23
sp|P07910|HNRPC_HUMAN	Heterogeneous nuclear ribonucleoproteins C1/C2	4.44E-24	1.83E-23
sp|P20962|PTMS_HUMAN	Parathymosin	5.54E-24	2.26E-23
sp|P21397|AOFA_HUMAN	Amine oxidase [flavin-containing] A	5.70E-24	2.31E-23
sp|P13647|K2C5_HUMAN	Keratin, type II cytoskeletal 5	8.65E-24	3.47E-23
sp|Q9Y6R7|FCGBP_HUMAN	IgGFc-binding protein	1.54E-23	6.11E-23
sp|Q07021|C1QBP_HUMAN	Complement component 1 Q subcomponent-binding protein, mitochondrial	2.17E-23	8.53E-23
sp|P02787|TRFE_HUMAN	Serotransferrin	1.61E-22	6.29E-22
sp|Q71UI9|H2AV_HUMAN	Histone H2A.V	1.89E-22	7.26E-22
sp|P0C0S5|H2AZ_HUMAN	Histone H2A.Z	1.89E-22	7.26E-22
sp|P04264|K2C1_HUMAN	Keratin, type II cytoskeletal 1	1.98E-22	7.54E-22
sp|P05387|RLA2_HUMAN	60S acidic ribosomal protein P2	8.43E-22	3.18E-21
sp|P02533|K1C14_HUMAN	Keratin, type I cytoskeletal 14	8.66E-22	3.21E-21
sp|P08779|K1C16_HUMAN	Keratin, type I cytoskeletal 16	8.66E-22	3.21E-21
sp|P06748|NPM_HUMAN	Nucleophosmin	1.10E-21	4.05E-21
sp|P19012|K1C15_HUMAN	Keratin, type I cytoskeletal 15	1.26E-21	4.60E-21
sp|P02765|FETUA_HUMAN	Alpha-2-HS-glycoprotein	1.86E-21	6.71E-21
sp|P15311|EZRI_HUMAN	Ezrin	2.42E-21	8.55E-21
sp|P26038|MOES_HUMAN	Moesin	2.42E-21	8.55E-21
sp|P35241|RADI_HUMAN	Radixin	2.42E-21	8.55E-21
sp|P51571|SSRD_HUMAN	Translocon-associated protein subunit delta	2.95E-21	1.03E-20
sp|A8K2U0|A2ML1_HUMAN	Alpha-2-macroglobulin-like protein 1	2.97E-21	1.03E-20
sp|P09429|HMGB1_HUMAN	High mobility group protein B1	4.81E-21	1.66E-20
sp|P05155|IC1_HUMAN	Plasma protease C1 inhibitor	6.50E-21	2.22E-20
sp|Q04695|K1C17_HUMAN	Keratin, type I cytoskeletal 17	7.24E-21	2.46E-20
sp|P60900|PSA6_HUMAN	Proteasome subunit alpha type-6	2.60E-20	8.74E-20
sp|P04259|K2C6B_HUMAN	Keratin, type II cytoskeletal 6B	2.71E-20	9.05E-20
sp|Q16610|ECM1_HUMAN	Extracellular matrix protein 1	3.17E-20	1.05E-19
sp|O43707|ACTN4_HUMAN	Alpha-actinin-4	5.22E-20	1.72E-19
sp|P01019|ANGT_HUMAN	Angiotensinogen	3.72E-19	1.22E-18
sp|P19652|A1AG2_HUMAN	Alpha-1-acid glycoprotein 2	4.50E-19	1.46E-18
sp|P60660|MYL6_HUMAN	Myosin light polypeptide 6	9.23E-19	2.97E-18
sp|P08670|VIME_HUMAN	Vimentin	9.95E-19	3.18E-18
sp|P04080|CYTB_HUMAN	Cystatin-B	1.83E-18	5.79E-18
sp|P49721|PSB2_HUMAN	Proteasome subunit beta type-2	2.23E-18	7.02E-18
sp|P24158|PRTN3_HUMAN	Myeloblastin	6.55E-18	2.05E-17
sp|Q9NRM1|ENAM_HUMAN	Enamelin	7.84E-18	2.44E-17
sp|P01008|ANT3_HUMAN	Antithrombin-III	1.04E-15	3.22E-15
sp|P01834|IGKC_HUMAN	Ig kappa chain C region	2.16E-15	6.61E-15
sp|Q9GZV4|IF5A2_HUMAN	Eukaryotic translation initiation factor 5A-2	2.99E-15	9.08E-15
sp|Q6IS14|IF5AL_HUMAN	Eukaryotic translation initiation factor 5A-1-like	5.59E-15	1.69E-14
sp|P49720|PSB3_HUMAN	Proteasome subunit beta type-3	5.98E-15	1.79E-14
sp|P02766|TTHY_HUMAN	Transthyretin	1.14E-14	3.39E-14
sp|P08238|HS90B_HUMAN	Heat shock protein HSP 90-beta	1.34E-14	3.98E-14
sp|O14818|PSA7_HUMAN	Proteasome subunit alpha type-7	2.40E-14	7.05E-14

sp|P26447|S10A4_HUMAN	Protein S100-A4	3.15E-14	9.20E-14
sp|P20160|CAP7_HUMAN	Azurocidin	3.47E-14	1.01E-13
sp|P14324|FPPS_HUMAN	Farnesyl pyrophosphate synthase	8.32E-14	2.40E-13
sp|P51884|LUM_HUMAN	Lumican	2.45E-13	7.02E-13
sp|P04196|HRG_HUMAN	Histidine-rich glycoprotein	4.61E-13	1.31E-12
sp|P05164|PERM_HUMAN	Myeloperoxidase	1.17E-12	3.31E-12
sp|P05546|HEP2_HUMAN	Heparin cofactor 2	1.60E-12	4.49E-12
sp|P35579|MYH9_HUMAN	Myosin-9	1.63E-12	4.55E-12
sp|P01023|A2MG_HUMAN	Alpha-2-macroglobulin	2.19E-12	6.06E-12
sp|P51858|HDGF_HUMAN	Hepatoma-derived growth factor	9.76E-12	2.69E-11
sp|P07237|PDIA1_HUMAN	Protein disulfide-isomerase	1.51E-11	4.15E-11
sp|P06702|S10A9_HUMAN	Protein S100-A9	2.88E-11	7.85E-11
sp|P15153|RAC2_HUMAN	Ras-related C3 botulinum toxin substrate 2	6.88E-11	1.86E-10
sp|P01857|IGHG1_HUMAN	Ig gamma-1 chain C region	9.67E-11	2.60E-10
sp|P16070|CD44_HUMAN	CD44 antigen	1.54E-10	4.12E-10
sp|P63208|SKP1_HUMAN	S-phase kinase-associated protein 1	1.56E-10	4.14E-10
sp|P10909|CLUS_HUMAN	Clusterin	2.06E-10	5.44E-10
sp|P28070|PSB4_HUMAN	Proteasome subunit beta type-4	2.10E-10	5.51E-10
sp|P25786|PSA1_HUMAN	Proteasome subunit alpha type-1	2.38E-10	6.22E-10
sp|P01876|IGHA1_HUMAN	Ig alpha-1 chain C region	4.34E-10	1.13E-09
sp|Q8IZP2|ST134_HUMAN	Putative protein FAM10A4	9.45E-10	2.44E-09
sp|P19823|ITIH2_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H2	1.07E-08	2.75E-08
sp|Q14240|IF4A2_HUMAN	Eukaryotic initiation factor 4A-II	1.35E-08	3.44E-08
sp|P25789|PSA4_HUMAN	Proteasome subunit alpha type-4	3.90E-08	9.89E-08
sp|P04206|KV307_HUMAN	Ig kappa chain V-III region GOL	9.59E-08	2.35E-07
sp|P18135|KV312_HUMAN	Ig kappa chain V-III region HAH	9.59E-08	2.35E-07
sp|P18136|KV313_HUMAN	Ig kappa chain V-III region HIC	9.59E-08	2.35E-07
sp|P01620|KV302_HUMAN	Ig kappa chain V-III region SIE	9.59E-08	2.35E-07
sp|P01622|KV304_HUMAN	Ig kappa chain V-III region Ti	9.59E-08	2.35E-07
sp|P01623|KV305_HUMAN	Ig kappa chain V-III region WOL	9.59E-08	2.35E-07
sp|P11413|G6PD_HUMAN	Glucose-6-phosphate 1-dehydrogenase	1.25E-07	3.05E-07
sp|P07195|LDHB_HUMAN	L-lactate dehydrogenase B chain	1.65E-07	3.99E-07
sp|P04003|C4BPA_HUMAN	C4b-binding protein alpha chain	4.33E-07	1.04E-06
sp|Q8NFI4|F10A5_HUMAN	Putative protein FAM10A5	5.82E-07	1.39E-06
sp|P01042|KNG1_HUMAN	Kininogen-1	8.65E-07	2.06E-06
sp|P01871|IGHM_HUMAN	Ig mu chain C region	9.23E-07	2.19E-06
sp|P06454|PTMA_HUMAN	Prothymosin alpha	1.77E-06	4.17E-06
sp|P07225|PROS_HUMAN	Vitamin K-dependent protein S	2.04E-06	4.78E-06
sp|P00738|HPT_HUMAN	Haptoglobin	2.31E-06	5.39E-06
sp|P22392|NDKB_HUMAN	Nucleoside diphosphate kinase B	2.46E-06	5.70E-06
sp|P01766|HV305_HUMAN	Ig heavy chain V-III region BRO	2.55E-06	5.86E-06
sp|P01777|HV316_HUMAN	Ig heavy chain V-III region TEI	2.55E-06	5.86E-06
sp|P0CG04|LAC1_HUMAN	Ig lambda-1 chain C regions	4.02E-06	9.19E-06
sp|P0CG05|LAC2_HUMAN	Ig lambda-2 chain C regions	4.20E-06	9.54E-06
sp|B9A064|IGLL5_HUMAN	Immunoglobulin lambda-like polypeptide 5	5.65E-06	1.28E-05
sp|P0CG06|LAC3_HUMAN	Ig lambda-3 chain C regions	9.52E-06	2.14E-05
sp|P19827|ITIH1_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H1	1.02E-05	2.29E-05
sp|P06314|KV404_HUMAN	Ig kappa chain V-IV region B17	1.28E-05	2.86E-05
sp|P28072|PSB6_HUMAN	Proteasome subunit beta type-6	1.31E-05	2.90E-05
sp|Q9Y3I1|FBX7_HUMAN	F-box only protein 7	1.64E-05	3.61E-05
sp|P69891|HBG1_HUMAN	Hemoglobin subunit gamma-1	1.65E-05	3.61E-05
sp|P69892|HBG2_HUMAN	Hemoglobin subunit gamma-2	1.65E-05	3.61E-05

sp|P04350|TBB4A_HUMAN	Tubulin beta-4A chain	2.12E-05	4.60E-05
sp|P62158|CALM_HUMAN	Calmodulin	2.28E-05	4.92E-05
sp|P01621|KV303_HUMAN	Ig kappa chain V-III region NG9 (Fragment)	2.67E-05	5.75E-05
sp|P00734|THRB_HUMAN	Prothrombin	2.73E-05	5.85E-05
sp|P00915|CAH1_HUMAN	Carbonic anhydrase 1	4.41E-05	9.40E-05
sp|P02788|TRFL_HUMAN	Lactotransferrin	6.42E-05	0.000136
sp|Q9BQE3|TBA1C_HUMAN	Tubulin alpha-1C chain	6.94E-05	0.0001464
sp|O43396|TXNL1_HUMAN	Thioredoxin-like protein 1	7.20E-05	0.0001512
sp|P06703|S10A6_HUMAN	Protein S100-A6	0.000106	0.0002216
sp|P84085|ARF5_HUMAN	ADP-ribosylation factor 5	0.0001359	0.0002826
sp|P63241|IF5A1_HUMAN	Eukaryotic translation initiation factor 5A-1	0.0002028	0.0004199
sp|P06312|KV401_HUMAN	Ig kappa chain V-IV region (Fragment)	0.0003051	0.0006229
sp|P06313|KV403_HUMAN	Ig kappa chain V-IV region JI	0.0003051	0.0006229
sp|P01625|KV402_HUMAN	Ig kappa chain V-IV region Len	0.0003051	0.0006229
sp|P08311|CATG_HUMAN	Cathepsin G	0.0003204	0.0006512
sp|P04004|VTNC_HUMAN	Vitronectin	0.0006151	0.0012443
sp|P02748|CO9_HUMAN	Complement component C9	0.0007614	0.0015331
sp|P27348|1433T_HUMAN	14-3-3 protein theta	0.0008381	0.0016801
sp|P59665|DEF1_HUMAN	Neutrophil defensin 1	0.0009091	0.001806
sp|P59666|DEF3_HUMAN	Neutrophil defensin 3	0.0009091	0.001806

Figure 22 shows a heat map of the set of 40 candidate menstrual blood specific marker proteins expressed within the 170 samples. Red to black reflect level of expression; green proteins not expressed. The names of these proteins are highlighted in blue above (Table 9).

Figure 22 Heat map of all 170 samples and the set of 40 candidate menstrual blood specific marker proteins.



Samples are ordered by hierarchical clustering using weighted average linkage.

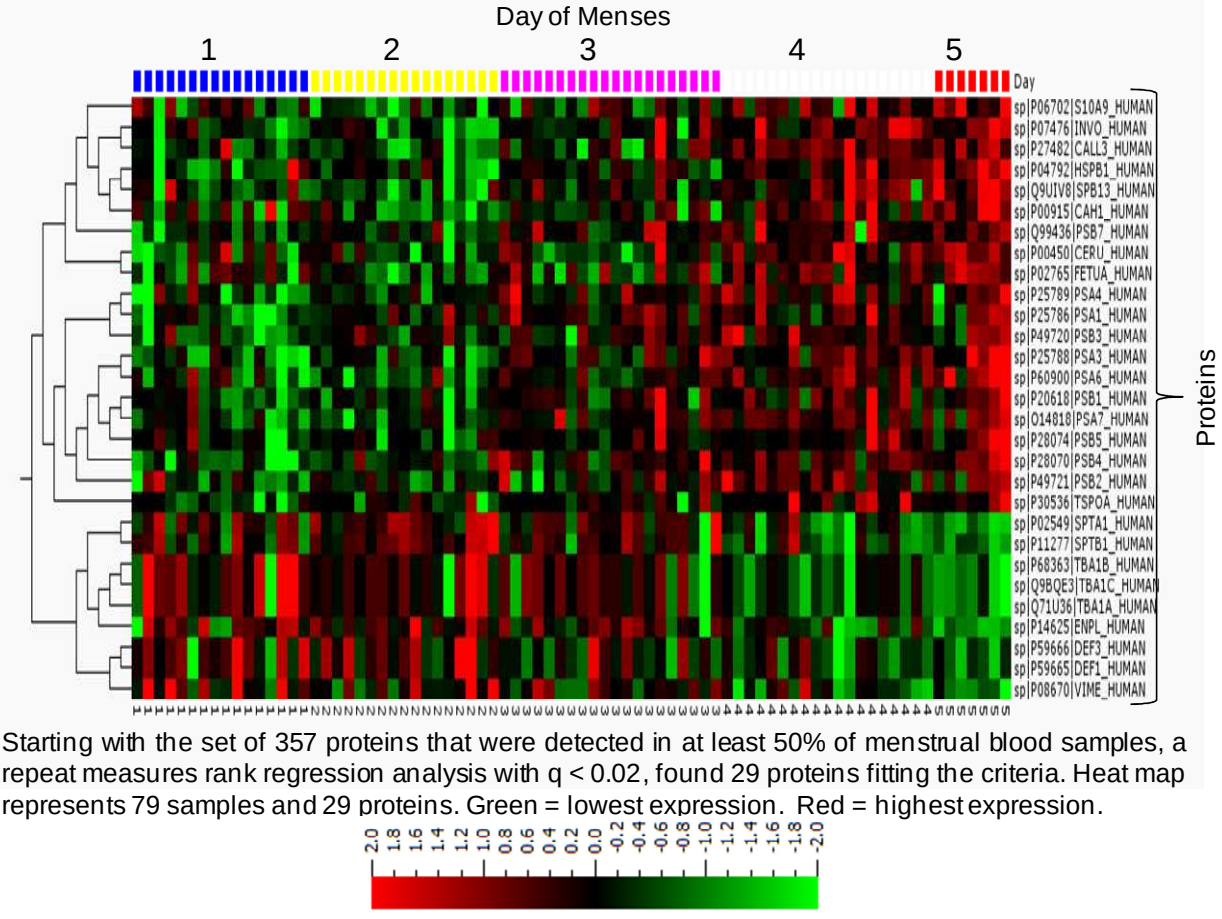
III) *FURTHER DEMOGRAPHIC ANALYSIS:* The primary goal of this study was to identify candidate marker proteins that could consistently distinguish between menstrual and venous blood in all women regardless of age, ethnicity and whether or not they use oral contraception (estrogen may affect protein expression). Results presented above demonstrate the feasibility of this method for menstrual blood identification for all women in all categories. Additional analyses (below) confirm these results but also suggest some interesting trends that might prove useful in forensic investigations. However, it should be stressed that the number of participants is relatively small (45 women) to draw solid conclusions about demographic data.

IV) *ANALYSIS OF MENSTRUAL BLOOD OVER THE COURSE OF MENSES*: As discussed previously, analysis of menstrual blood over the course of menses is not simple, as the lengths of women's

periods vary over several days. Consequently, while one might expect to see similar protein expression patterns and levels on the first day of all women's periods, the third and last day of one woman's period may more resemble the fifth and last day of another's woman period rather than the third and middle day of that women's period. These differences are accounted for as follows. Analyses were performed with a set of 357 proteins detected in at least 50% of menstrual blood samples. Since almost all subjects contributed menstrual blood on more than one day, the analysis is more powerful and better able to find relevant trends if the data is "paired", i.e. variations within each subject is taken into account. Some or all samples, depending on the number of days of menses, were used depending on the type of analysis performed.

i) ***Days of Menses: Rank Regression –*** Rank regression analysis was used on subjects whose periods were four or five days in order to remove possible confounding effects from subjects whose periods lasted only two or three days. A total of 79 samples remained for analysis.

Figure 23 Rank regression (heat map) of protein expression over the course of menses – limited to women with periods of four to five days.



Starting with the set of 357 proteins that were detected in at least 50% of menstrual blood samples, a repeat measures rank regression analysis with $q < 0.02$, found 29 proteins fitting the criteria. Heat map represents 79 samples and 29 proteins. Green = lowest expression. Red = highest expression.

A heat map of these data (Figure 23) clearly shows significant trends of increased and decreased expression over the course of menses in 29 proteins - nine decreased over the length of the period (Table 10) and 20 increased (Table 11).

Table 10 List of 9 proteins which decrease over the course of menses (q < 0.02)

Candidate menstrual blood specific marker

One of 222 proteins enriched in menstrual blood over venous blood

Accession	Name	p-value	q-value	R-statistic
sp|P68363|TBA1B_HUMAN	Tubulin alpha-1B chain	3.18E-05	0.0011345	-0.5097796
sp|Q71U36|TBA1A_HUMAN	Tubulin alpha-1A chain	3.18E-05	0.0011345	-0.5097796
sp|Q9BQE3|TBA1C_HUMAN	Tubulin alpha-1C chain	3.18E-05	0.0011345	-0.5097796
sp|P02549|SPTA1_HUMAN	Spectrin alpha chain, erythrocytic 1	0.0002826	0.0059804	-0.4525918
sp|P08670|VIME_HUMAN	Vimentin	0.0003015	0.0059804	-0.4507309
sp|P14625|ENPL_HUMAN	Endoplasmin	0.000759	0.0118658	-0.4230182
sp|P11277|SPTB1_HUMAN	Spectrin beta chain, erythrocytic	0.0013583	0.0181593	-0.4042514
sp|P59665|DEF1_HUMAN	Neutrophil defensin 1	0.0015317	0.0188563	-0.4002385
sp|P59666|DEF3_HUMAN	Neutrophil defensin 3	0.0015317	0.0188563	-0.4002385

Table 11 List of 15 proteins which increase over the course of menses (q < 0.02)

Candidate menstrual blood specific marker

One of 222 proteins enriched in menstrual blood over venous blood

Accession	Name	p-value	q-value	R-statistic
sp|P25788|PSA3_HUMAN	Proteasome subunit alpha type-3	2.44E-08	8.70E-06	0.6463171
sp|P28070|PSB4_HUMAN	Proteasome subunit beta type-4	1.58E-07	2.83E-05	0.6163317
sp|P60900|PSA6_HUMAN	Proteasome subunit alpha type-6	9.55E-07	0.0001136	0.5842457
sp|P04792|HSPB1_HUMAN	Heat shock protein beta-1	2.80E-06	0.0002496	0.5632603
sp|P07476|INVO_HUMAN	Involucrin	5.98E-06	0.0004273	0.5474804
sp|P27482|CALL3_HUMAN	Calmodulin-like protein 3	1.01E-05	0.0006031	0.5360646
sp|Q9UIV8|SPB13_HUMAN	Serpin B13	2.09E-05	0.0010657	0.5196779
sp|O14818|PSA7_HUMAN	Proteasome subunit alpha type-7	9.57E-05	0.0026894	0.4821848
sp|P00450|CERU_HUMAN	Ceruloplasmin	9.79E-05	0.0026894	0.4815899
sp|P25789|PSA4_HUMAN	Proteasome subunit alpha type-4	9.23E-05	0.0026894	0.483143
sp|P25786|PSA1_HUMAN	Proteasome subunit alpha type-1	0.0002576	0.0059804	0.4552372
sp|P30536|TSPOA_HUMAN	Translocator protein	0.0002528	0.0059804	0.4557692
sp|P49720|PSB3_HUMAN	Proteasome subunit beta type-3	0.0002892	0.0059804	0.4519266
sp|Q99436|PSB7_HUMAN	Proteasome subunit beta type-7	0.0004424	0.0083129	0.4395047

sp|P02765|FETUA_HUMAN	Alpha-2-HS-glycoprotein	0.0006314	0.01127	0.4287349
sp|P49721|PSB2_HUMAN	Proteasome subunit beta type-2	0.0006724	0.0114312	0.4267895
sp|P06702|S10A9_HUMAN	Protein S100-A9	0.0007645	0.0118658	0.4227919
sp|P00915|CAH1_HUMAN	Carbonic anhydrase 1	0.0009045	0.0134545	0.4174744
sp|P20618|PSB1_HUMAN	Proteasome subunit beta type-1	0.0013631	0.0181593	0.404135
sp|P28074|PSB5_HUMAN	Proteasome subunit beta type-5	0.0013734	0.0181593	0.4038845

ii) ***First Day/Last Day Sample Comparison –*** To determine if there were any major differences in protein expression between the first and last days of menses, a paired comparison was performed irrespective of how long a particular subject's period lasted. This analysis excluded any samples that were not first or last, leaving 71 samples in the analysis. With $q < 0.02$, only 5 proteins showed significant differences between first and last days: two were higher on the last day and three lower on the last day (Table 12). All five also appear as significant in the rank regression analysis described above.

Table 12 List of five proteins with significant change in expression between first and last days of menses (q < 0.02).

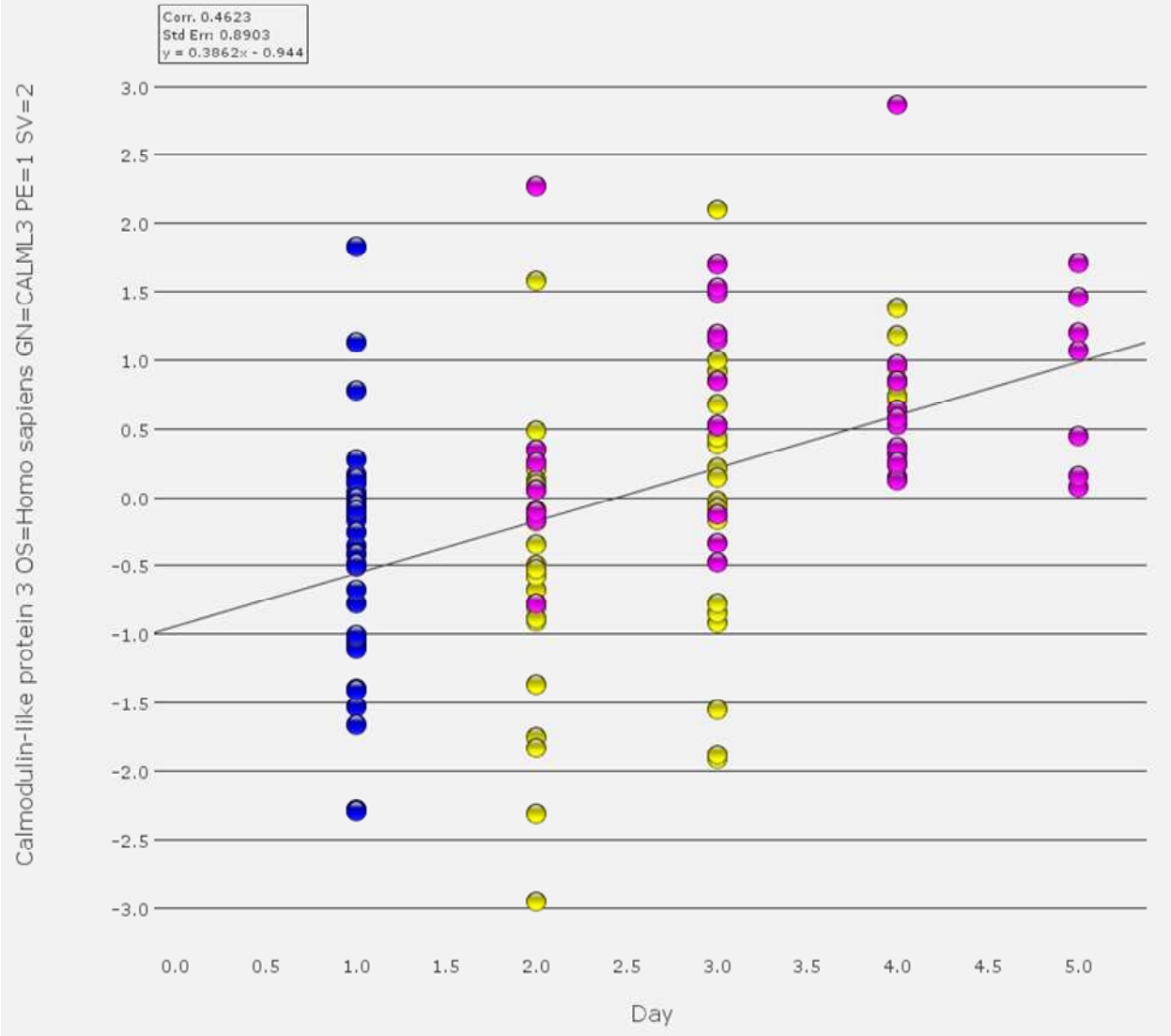
A positive R-statistic indicates that the protein is higher on the last day than the first, and vice versa for a negative R-statistic.

One of 222 proteins enriched in menstrual blood over venous blood

Accession	Name	p-value	q-value	R-statistic
sp|P27482|CALL3_HUMAN	Calmodulin-like protein 3	6.79E-05	0.0160342	0.6289124
sp|P25788|PSA3_HUMAN	Proteasome subunit alpha type-3	0.0002246	0.0160342	0.5922063
sp|P68363|TBA1B_HUMAN	Tubulin alpha-1B chain	0.0002029	0.0160342	-0.5954936
sp|Q71U36|TBA1A_HUMAN	Tubulin alpha-1A chain	0.0002029	0.0160342	-0.5954936
sp|Q9BQE3|TBA1C_HUMAN	Tubulin alpha-1C chain	0.0002029	0.0160342	-0.5954936

Figure 24 shows a scatter plot of calmodulin-like protein 3 expression vs day of period. First day samples are colored blue, last day samples are colored pink, and all others are colored yellow. Samples from days two or three which are also the last days of a woman's period tend to have higher expression values than samples from the same days which are not last, suggesting that calmodulin-like protein 3's expression is more consistent with the end of menses than the second or third day.

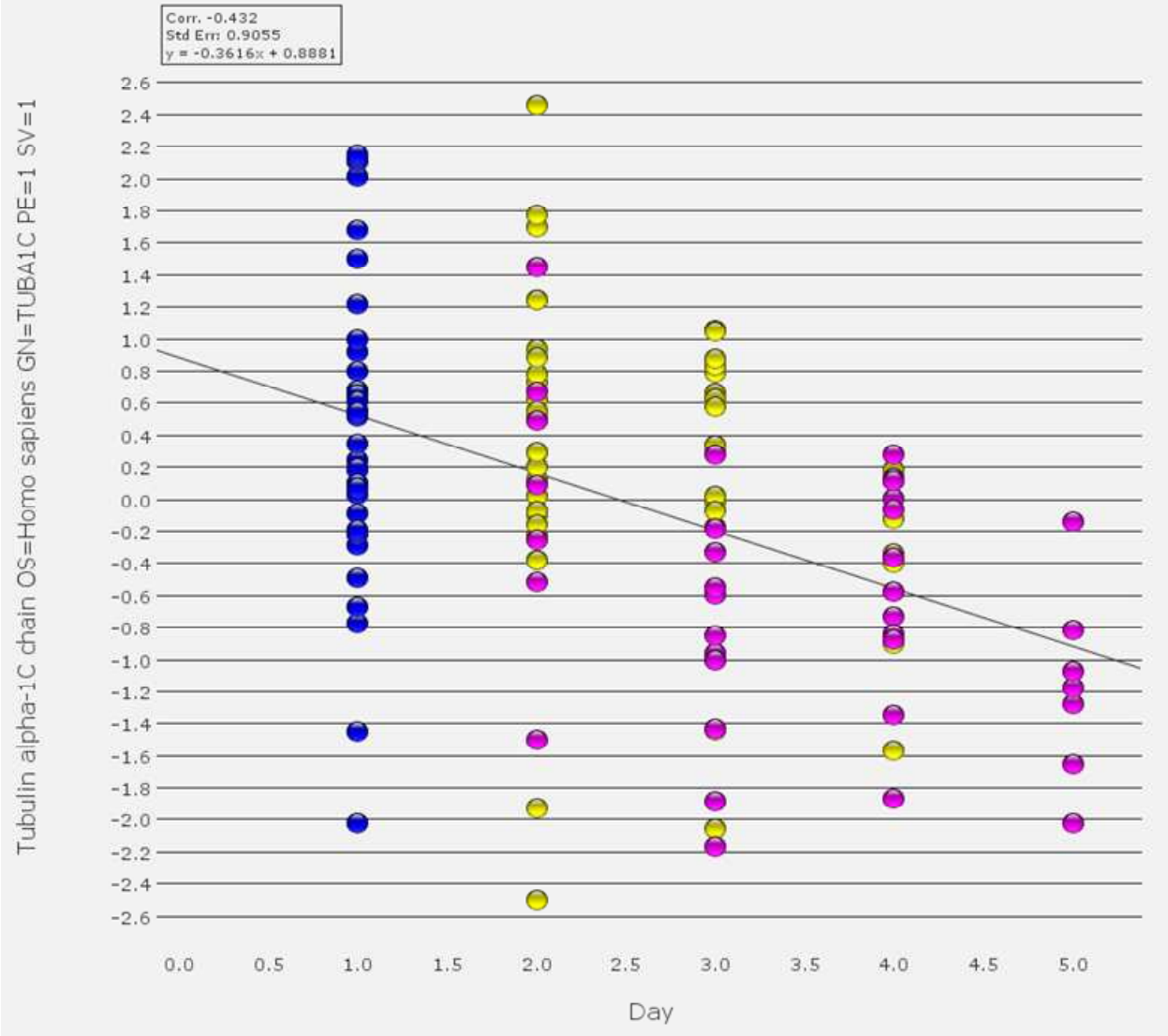
Figure 24 Scatter plot of the amount of calmodulin-like protein 3 vs day of period.



Samples from day 2 or 3 which are also the last day tend to have higher expression than samples from the same day which are not last.

Similarly, Figure 25 shows a scatter plot of the expression of tubulin alpha-1C chain which declines over menses. Again, first day samples are colored blue, last day samples are colored pink, and all others are colored yellow. Here, however, samples from day two, and especially day three, which are also the last day of women's periods tend to have lower expression than samples from the same day which are not last, suggesting that tubulin alpha-1C chain's expression is more consistent with the end of menses than days in between.

Figure 25 Scatter plot of the amount of tubulin alpha-1 C chain vs. day of period.



Samples from day two or three which are also the last day tend to have lower expression than samples from the same day which are not last.

v) *EVALUATION OF MENSTRUAL BLOOD WITH RESPECT TO DEMOGRAPHIC DIFFERENCES:* Menstrual blood samples were analyzed to determine if there were differences in protein expression with respect to i) oral contraception use, ii) ethnicity and iii) age.

i) ***Oral Contraception*** **–** Hormonal contraception functions to prevent pregnancy through multiple mechanisms including suppression of ovulation as well as changes in cervical mucus (Rivera et al., 1999) both of which are likely to result from and contribute to changes in protein expression. ANOVA comparison of women using oral contraception and those not using oral

contraception was performed to determine whether differences in protein expression could be identified.

Analysis was performed on 357 proteins detected in 50% or more of menstrual blood samples (for a total of 126 samples). A two-group comparison carried out on subjects who use oral birth control and those who do not (with $q < 0.02$ and filtered for a fold change of two or more) found 30 proteins fitting the criteria.

Table 13 List of 15 proteins which increase with use of oral birth control (q < 0.02).

Accession	Name	p-value	q-value	R-statistic
sp|P02647|APOA1_HUMAN	Apolipoprotein A-I	3.36E-08	5.09E-06	0.4989021
sp|P08185|CBG_HUMAN	Corticosteroid-binding globulin	3.90E-08	5.09E-06	0.4968708
sp|P00450|CERU_HUMAN	Ceruloplasmin	2.81E-07	1.10E-05	0.4683719
sp|P02790|HEMO_HUMAN	Hemopexin	9.87E-07	3.20E-05	0.4487807
sp|P68871|HBB_HUMAN	Hemoglobin subunit beta	1.09E-06	3.25E-05	0.4471185
sp|P02766|TTHY_HUMAN	Transthyretin	1.13E-05	0.000287335	0.4069478
sp|P02042|HBD_HUMAN	Hemoglobin subunit delta	1.58E-05	0.00035159	0.4007236
sp|P05164|PERM_HUMAN	Myeloperoxidase	2.88E-05	0.000571626	0.3891923
sp|P02743|SAMP_HUMAN	Serum amyloid P-component	4.14E-05	0.000739488	0.3820505
sp|P00915|CAH1_HUMAN	Carbonic anhydrase 1	4.55E-05	0.000774133	0.3801614
sp|P02768|ALBU_HUMAN	Serum albumin	0.00010408	0.001429112	0.3631349
sp|P69905|HBA_HUMAN	Hemoglobin subunit alpha	0.00041844	0.003556768	0.332121
sp|P02763|A1AG1_HUMAN	Alpha-1-acid glycoprotein 1	0.00053005	0.004205068	0.3265149
sp|P01009|A1AT_HUMAN	Alpha-1-antitrypsin	0.00101373	0.006462517	0.3105659
sp|P13646|K1C13_HUMAN	Keratin, type I cytoskeletal 13	0.00199484	0.011486417	0.2929187

One of 222 proteins enriched in menstrual blood over venous blood

Of these, 15 showed an increased expression with use of oral birth control (Table 13), and 15 showed a decrease in expression (Table 14). Figure 26 is a heat map of these 30 proteins ordered by use of oral birth control (blue/yellow top row). Proteins with elevated expression in women using oral contraceptives are seen in red in the upper right of the figure; proteins with reduced expression (green) are seen on the lower right. Two proteins, keratin type II cytoskeletal 7 and 8, are among the candidate markers for menstrual blood. Fourteen proteins are part of the set that show higher expression in menstrual blood than venous blood (there were no differences in protein expression in venous blood between women on oral contraception and those not on oral contraception). Of particular note is the relative increase and tighter spread in corticosteroid-binding globulin (CBG) in people using oral birth control ($q = 5.08e\text{-}06$) (Figure 27). CBG is a major transport protein for glucocorticoids and progestins in the blood of almost all vertebrate species. Progestins are commonly used in oral birth control (Zhou, 2008).

Table14 List of 15 proteins which decrease with use of oral birth control (q < 0.02).

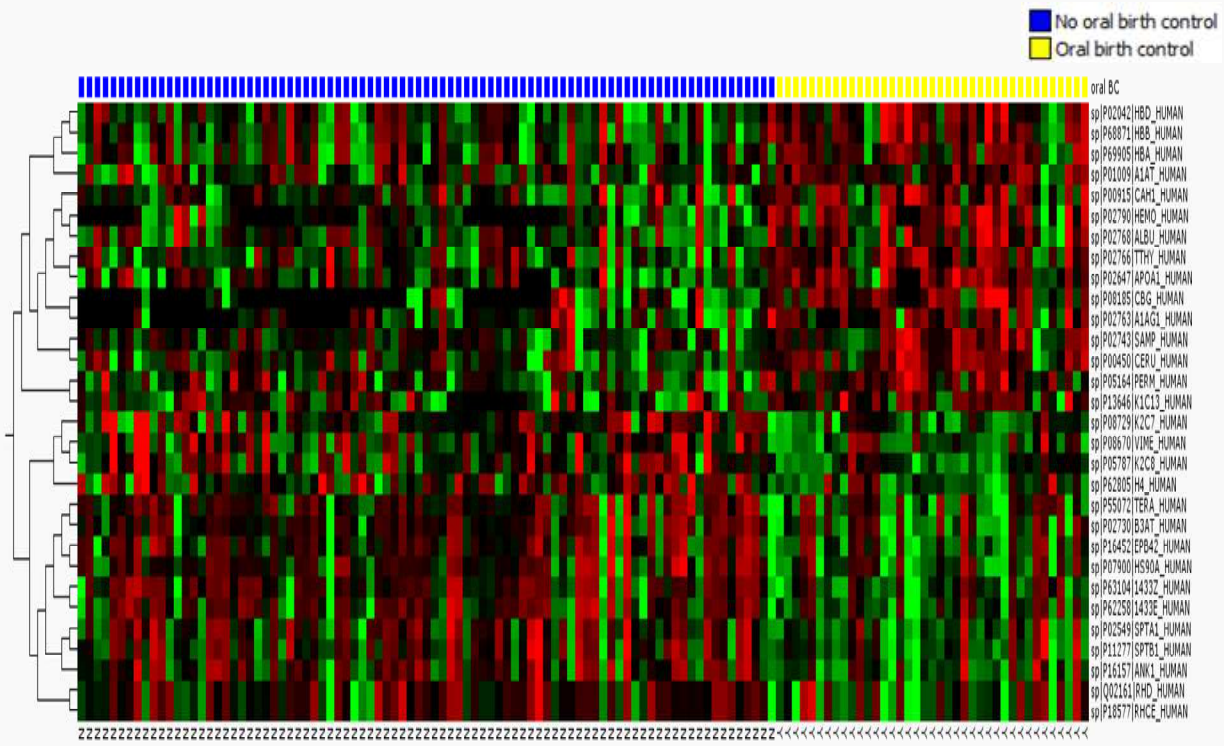
Accession	Name	p-value	q-value	R-statistic
sp|P08729|K2C7_HUMAN	Keratin, type II cytoskeletal 7	2.65E-07	1.10E-05	-0.4692313
sp|P05787|K2C8_HUMAN	Keratin, type II cytoskeletal 8	1.94E-06	5.33E-05	-0.4376928
sp|P16157|ANK1_HUMAN	Ankyrin-1	3.39E-05	0.0006375	-0.3860012
sp|P63104|1433Z_HUMAN	14-3-3 protein zeta/delta	6.31E-05	0.0009787	-0.3735704
sp|P62805|H4_HUMAN	Histone H4	8.98E-05	0.001336	-0.3662414
sp|P16452|EPB42_HUMAN	Erythrocyte membrane protein band 4.2	0.000112	0.0014805	-0.3615839
sp|P02730|B3AT_HUMAN	Band 3 anion transport protein	0.0002757	0.0025655	-0.341761
sp|P02549|SPTA1_HUMAN	Spectrin alpha chain, erythrocytic 1	0.000292	0.0026059	-0.3404556
sp|P07900|HS90A_HUMAN	Heat shock protein HSP 90-alpha	0.0003782	0.0032932	-0.3344861
sp|P08670|VIME_HUMAN	Vimentin	0.0004477	0.0037171	-0.3305284
sp|P11277|SPTB1_HUMAN	Spectrin beta chain, erythrocytic	0.0004958	0.0040229	-0.3281086
sp|P55072|TERA_HUMAN	Transitional endoplasmic reticulum ATPase	0.0007276	0.0054115	-0.318832
sp|P18577|RHCE_HUMAN	Blood group Rh(CE) polypeptide	0.0008138	0.0056968	-0.3160666
sp|Q02161|RHD_HUMAN	Blood group Rh(D) polypeptide	0.0008138	0.0056968	-0.3160666
sp|P62258|1433E_HUMAN	14-3-3 protein epsilon	0.0041878	0.0191671	-0.2722272

Candidate menstrual blood specific marker

One of 222 proteins enriched in menstrual blood over venous blood

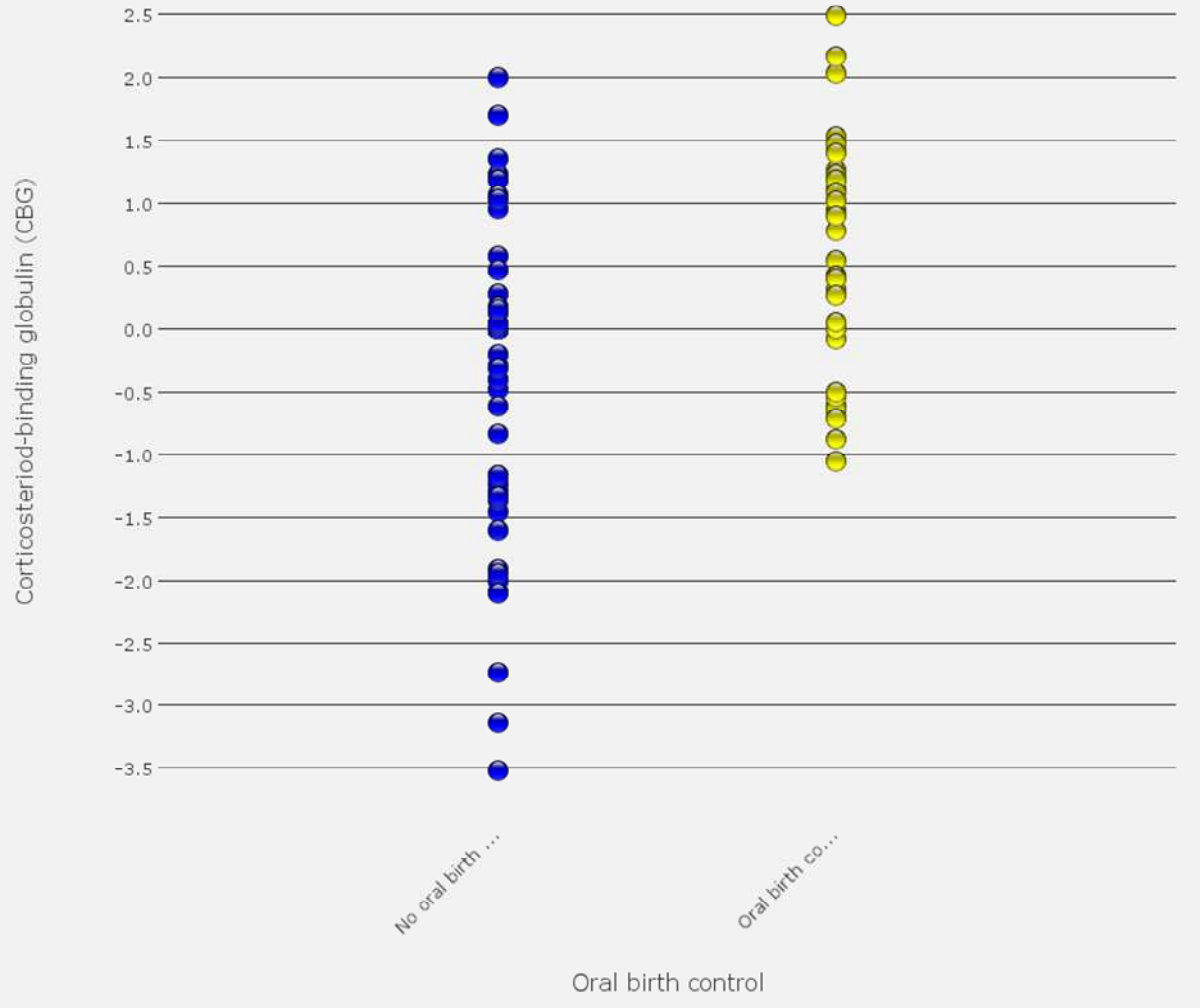
Presently the numbers of samples are too small to draw a final conclusion; however, these data suggest that differences in protein expression levels may be able to distinguish menstrual blood samples from women using hormonal contraception from those who do not. (Jesperson J 1983; Jesperson & Nielson 1989; Ganesh et al. 2009). Regardless, the data clearly demonstrate that oral contraceptives do not interfere with the ability to identify menstrual blood in women using them.

Figure 26 Comparison of 30 proteins found to have a greater than two fold change in expression (up or down) between women using hormonal contraception and those not using hormonal contraception.



Rank regression of protein expression over the course of menses – limited to women with periods of four to five days - ordered by hierarchical clustering and samples ordered by use of oral birth control (blue/yellow line top).

Figure 27 Relative increase in corticosteroid-binding globulin (CBG) in individuals using oral birth control.



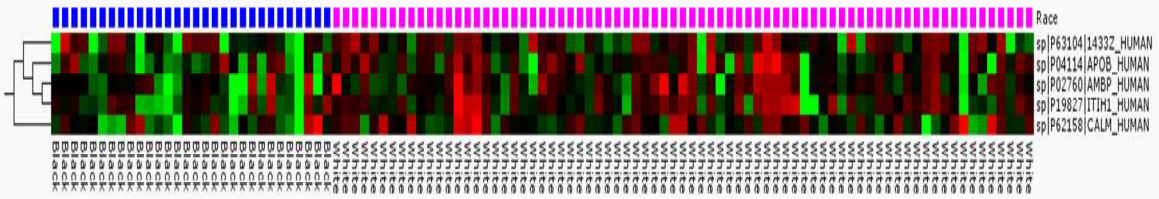
ii) ***Ethnicity –*** Analysis of differences in protein expression between ethnic groups was confined to Blacks and Whites due to the comparatively small sample size of Asians and Hispanics. Analysis was performed on a set of 357 proteins that were detected in 50% or more of menstrual blood samples. Use of oral birth control was corrected for by setting oral birth control as an eliminated factor. A two group comparison with q <0.02, found five proteins that fit the criteria, all of which show lower expression in Blacks compared to Whites (Table 15). Figure 28 is a heat map of these five proteins ordered by hierarchical clustering and samples ordered by ethnicity. Of these five proteins, three are part of the set that show higher expression in menstrual blood than venous blood. Again, however, it must be stressed that this data set is too small to draw any conclusions about differences in menstrual blood protein expression by ethnicity; nonetheless they clearly demonstrate that the proteomic menstrual blood assay is not limited in its ability to detect menstrual blood by ethnicity.

Table 15 Differential expression of menstrual blood proteins by ethnicity.

Accession	Name	p-value	q-value	R-statistic
sp|P02760|AMBP_HUMAN	Protein AMBP	9.11E-06	0.0032527	-0.4535066
sp|P62158|CALM_HUMAN	Calmodulin	0.0001315	0.0143548	-0.3964398
sp|P19827|ITIH1_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H1	0.0001519	0.0143548	-0.3930393
sp|P63104|1433Z_HUMAN	14-3-3 protein zeta/delta	0.0001624	0.0143548	-0.3914489
sp|P04114|APOB_HUMAN	Apolipoprotein B-100	0.000201	0.0143548	-0.3863286

One of 222 proteins enriched in menstrual blood over venous blood

Figure 28 Heat map of five proteins ordered by hierarchical clustering and samples ordered by ethnicity.

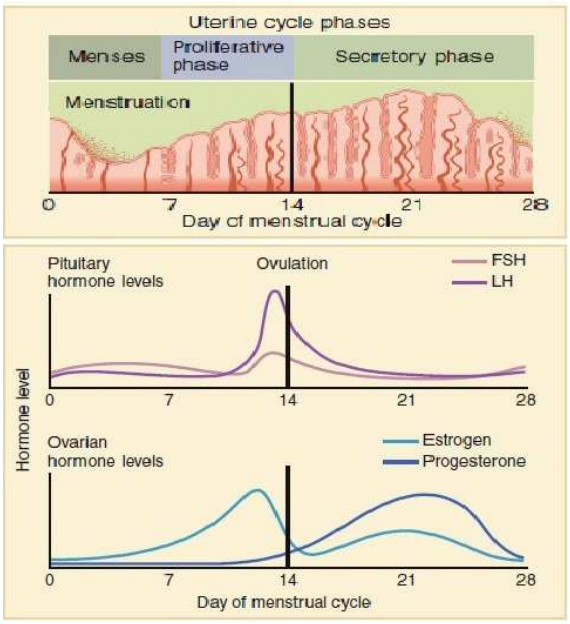


iii) ***Age –*** Multivariate linear regression analysis by age of all samples and 357 proteins found in 50% or more of menstrual blood samples revealed no proteins with significant correlation of expression to age at $q < 0.02$.

SECTION 2 – **MALDI TOF/TOF ANALYSIS OF VAGINAL FLUID***:* The aims were to identify and assess vaginal fluid marker candidates in a large number of women of varying ages, ethnicities and use of hormonal contraceptives.

• VAGINAL FLUID SAMPLE COLLECTION & PROCESSING METHODS:

Figure 29 – 28 day menstrual cycle showing endometrial changes with hormone expression.



Source: OpenStax College - Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/, Jun 19, 2013

I) *SAMPLE COLLECTION:* Vaginal fluid samples were obtained from the same 45 women who donated menstrual blood samples. Like menstrual blood, whole vaginal fluid samples were collected using sterile, single use Softcup menstrual cups on the 13th and 19th days following the first day of each woman's period. These time points correspond to approximately one week following the end of their last period and one week prior to their next period respectively (assuming a 28 day menstrual cycle). This timing captured the peak proliferative phase of the menstrual cycle (maximum levels of estrogen, luteinizing and follicle stimulating hormone) and the peak secretory phase (near maximum level of progesterone) (see Figure 29) and, consequently, represents distinct times of protein expression. Additionally, two vaginal fluid samples were collected from five post-menopausal women one week apart thus giving a total of 100 vaginal fluid samples. Volunteers were asked to keep Softcups in place for a minimum of four hours but no longer than 12 hours. When removed, cups were immediately put in sterile 50 ml conical tubes, placed in small black plastic bags and stored at -20^{o}C. When returned to the laboratory, samples were kept at -80^{o}C until processed. Demographic information on these women was described above (Tables 2 and 3). Six of the 100 samples could not be processed (typically because of contamination with blood likely due to irritation caused by the Softcup). Four were from menstruating women (180 D19, 724 D13, 568D13 & D19) and two from one postmenopausal volunteer (679 D1 & D7).

II) *PROCESSING*: After thawing, samples were removed from Softcups using a 1 ml pipette tip with the tip cut to increase orifice size. For viscous samples, known volumes of PBS were added to the Softcup and mixed with the vaginal fluid using a rubber policeman. All sample

volumes were measured. Samples were homogenized in ten volumes of PBS using a bead beater (FastPrep-24, MP Biomedical, Santa Ana, CA) with 1.4 mm ceramic beads with the instrument set at 6 m/s for 30 s. Samples were then spun at 20,000 g x 30 min at 4^{o}C. The supernatant was carefully removed. Protein concentration was assayed using the BCA method with BSA as standard. The solubilized pellet was diluted 4 times with PBS and both pellet and supernatant digested with trypsin at a 20:1 protein:trypsin ratio. Peptides were dried by SpeedVac and resuspended in 2% ACN/0.1% TFA. As with menstrual blood, samples were tagged with iTRAQ isobaric labels for multiplexing and relative quantitation.

III) *HPLC MS/MS*: For HPLC (Dionex 3000, ThermoFisher, Sunnyvale, CA) the equivalent of 2 µg protein were loaded onto a C-18 reverse phase column and separated using a two phase non-linear ACN/TFA gradient over 3 hrs. Gradient: Phase A = 2% ACN/0.1% TFA, Phase B = 98% ACN/0.1% TFA.

0-10 min 2% ACN
10-13 min 2%-12% ACN
13-179 min 12%-40% ACN
179-184 min 40%-95% ACN
184-190 min 95% ACN
190-200 min 2% ACN

HPLC eluate was mixed in a 1:1 ratio with CHCA matrix (a three X dilution of saturated CHCA prepared in 0.1% TFA, 50% ACN/H_2O) just prior to spotting on a 384 position MALDI plate. MS data was acquired at a laser repetition rate of 200 Hz with 600 laser shots/spectrum (50 laser shots/sub-spectrum). MS/MS data was acquired at 200 Hz in 1 kV MS/MS mode with 3200 laser shots/spectrum (40 laser shots/sub-spectrum) with the following TOF/TOF Series Explorer Stop Conditions: Maximum shots per spectrum, 3200; minimum shots per spectrum, 1000; number of MS/MS fragments, 8; S/N of each fragment 75. The top 30 strongest peaks were selected for MS/MS analysis. Peptides were identified and quantified via automated database searching of raw data using ProteinPilot 4.5 (SCIEX, CA) against Swiss-Prot Human protein database (40,670 Protein Entries, April 2009). Paragon was selected as processing method. iTRAQ 8plex was selected as sample type. Methyl methanethiosulfonate and trypsin were selected as alkylation reagent and enzyme respectively.

• VAGINAL FLUID RESULTS: The same two criteria used to identify menstrual blood candidate marker proteins were applied to vaginal fluid markers. These were: i) specificity (i.e. markers unique to, or enriched in, vaginal fluid) and ii) frequency (i.e. markers found in all or most vaginal fluid samples). Unlike the menstrual blood proteome, however, the vaginal fluid proteome has been well described (Tang et al. 2007). Consequently, to identify marker candidates the 94 vaginal fluid samples were labeled with iTRAQ barcodes (see iTRAQ labeling method in menstrual blood section above, Figure 15), processed by HPLC-MS/MS and all identified proteins compared against the 13 body fluid proteomes in the *Sys-BodyFluid database* (Li et al. 2009).

Typical MS and MS/MS data from an iTRAQ labeling experiment are shown in Figure 30. Here, D13 vaginal fluid samples from four volunteers were individually labeled with iTRAQ isobaric tags containing reporter groups 113, 114, 115 and 116 Da. After labeling, samples were combined, peptides separated by nano-HPLC and analyzed by MALDI TOF/TOF. Figure 30a shows a single strong signal with a mass of 2202.1492 Da. MS/MS of this peptide (Figure 30b) reveals it is from marker protein Beta-crystallin S with the amino acid sequence VLEGVWIFYELPNYR. The bracket below the MS/MS peak at ~115 Da is expanded in Figure 30c and shows each of the reporter ions 113, 114, 115 and 116. The relative intensities of these ions are proportional to the amount of this peptide in each sample. These relative peptide quantities are reported by Protein Pilot software (AB Sciex) as ratios relative to a single labeled sample in the run. Protein ratios are obtained using a weighted average of the ratios of all of the peptides contributing to the protein identification. For analysis of all vaginal fluid samples, day 13 and day 19 samples were labeled with isobaric tags 113-116 and 117-121 (exclusive of 120 Da which is not used) respectively, and analyzed by HPLC-MALDI TOF/TOF.

Figure 30 MS & MS/MS Spectra of an iTRAQ Labeled Vaginal Fluid Sample

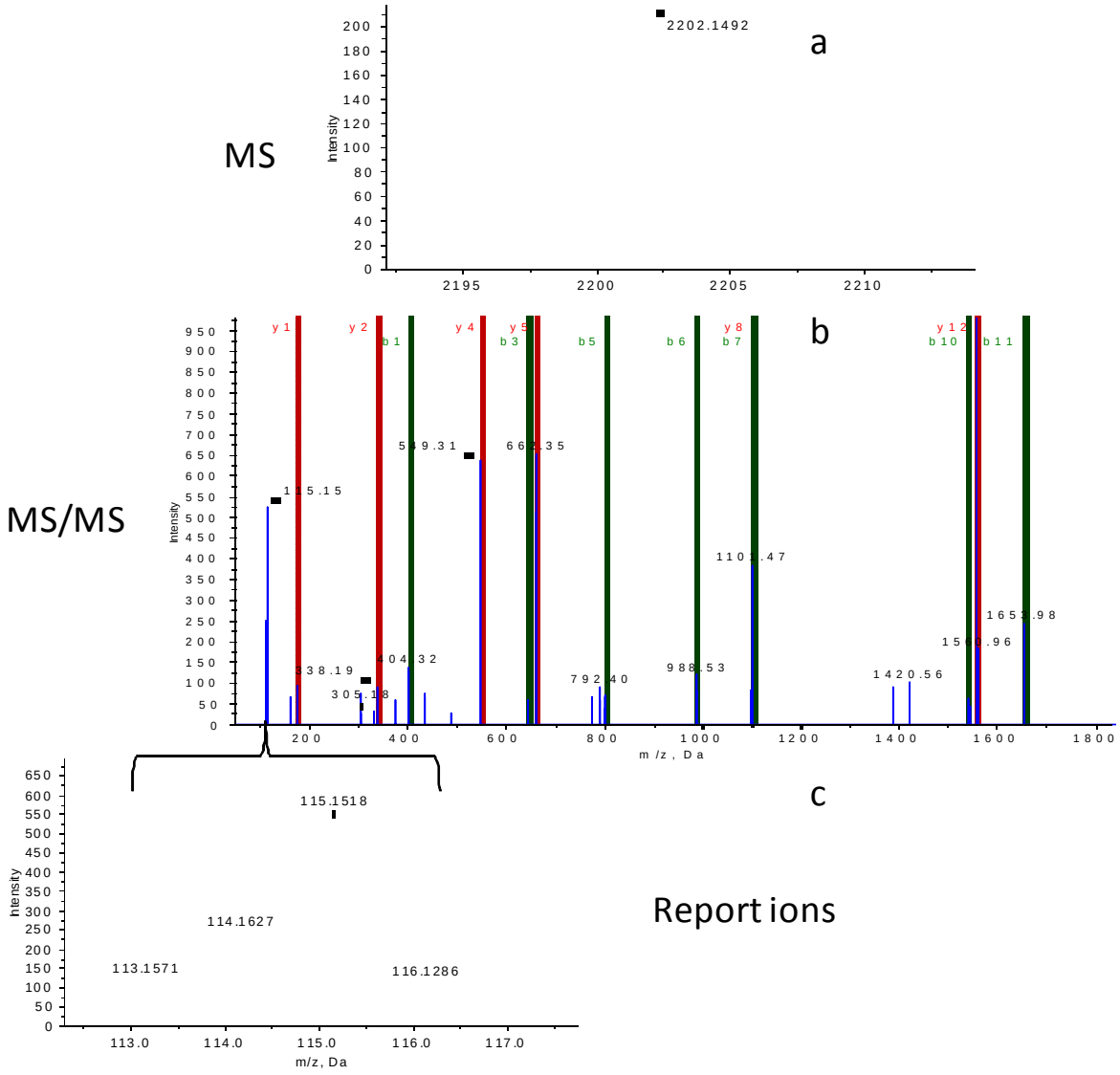
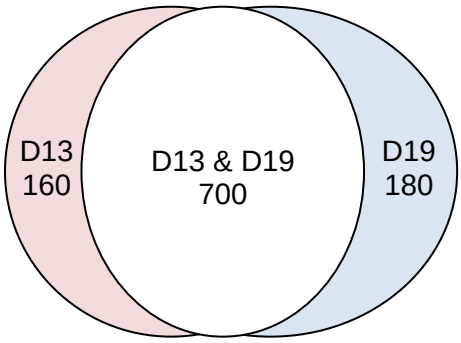


Figure 30 Spectra of MS, MS/MS of peptide VLEGVWIFYELPNYR from marker CRYS labeled with iTRAQ reagents (113, 114, 115 116). a, MS spectrum of the peptide; b, MS/MS spectrum which is used to identify peptide sequence; c, expanded MS/MS spectrum of iTRAQ labels showing the relative amounts of peptides from each of the four samples.

A total of 1,040 proteins were identified in the 94 vaginal fluid samples tested; not all proteins were found in every sample. Of the 1,040 proteins, 860 were found only in day 13 samples, and 880 only in day 19 samples, 700 proteins were common to both days 13 and 19 (Figure 31). However, not all of these 700 proteins were shared equally between the days, i.e. some proteins are found more often in day 13 and some in day 19 samples. Tables 16 and 17 show

Figure 31 – Total number of vaginal fluid proteins found and their distribution between days 13 and 19 collections.



Total proteins identified = 1,040; D13 = 860; D19 = 880

vaginal fluid proteins shared between days 13 and 19 and the percentage of women they are found in on each day. For example, 48 proteins were found in all women on days 13 and 19. However, an additional 32 proteins were found in 40 women (91%) at day 13 but found in 64%–100% of women on day 19 (see Table 9 top shaded rows). Similarly, 53 proteins were found in 40 women (91%) on day 19, while only 24%-40% of women expressed these proteins at day 13. While it is possible that these differences represent actual variations in protein expression, it is also possible that they represent differences in sample collection time (e.g. morning, evening etc.) as well as in the length of time over which sample collection occurred (4-12 hrs). This is also a reasonable explanation as to why there are differences in protein expression within each collection day – i.e. not all proteins are found in all women on day 13 or day 19.

Table 16 Percent Proteins Shared Between D-13 and D-19 Ranked by D-13

Total Number of Proteins	No. of Women Positive D-13	D-13 %	No. of Women Positive D19	D-19 %
48	44	100	44	100
3	44	100	40	91
5	40	91	44	100
15	40	91	40	91
10	40	91	36	82
1	40	91	32	73
1	40	91	28	64
14	36	82	44	100

Table 17 Percent Proteins Shared Between D-13 and D-19 Ranked by D-19

Total Number of Proteins	No. of Women Positive D-13	D-13 %	No. of Women Positive D19	D-19 %
48	44	100	44	100
14	36	82	44	100
15	40	91	40	91
13	36	82	40	91
8	32	73	40	91
6	28	64	40	91
1	24	55	40	91
10	36	82	40	91

13	36	82	40	91
10	36	82	36	82
3	36	82	32	73
2	36	82	28	64
1	36	82	24	55
8	32	73	40	91
5	32	73	36	82
5	32	73	32	73
4	32	73	28	64
2	32	73	24	55
1	32	73	16	36
Total 136				

10	36	82	36	82
5	32	73	36	82
4	28	64	36	82
3	24	55	36	82
1	20	45	36	82
1	40	91	32	73
3	36	82	32	73
5	32	73	32	73
8	28	64	32	73
2	24	55	32	73
3	20	45	32	73
3	16	36	32	73
Total 160				

• IDENTIFICATION OF VAGINAL FLUID: To identify vaginal fluid markers, proteins identified in all or most vaginal fluid samples (i.e. high frequency) were compared to the proteomes of the 13 body fluids found in the Sys-BodyFluid database to determine relative specificity. Table 18 identifies five vaginal fluid candidate marker proteins, four of which, small proline-rich protein 3 (SPRR3), alpha-2-macroglobulin-like 1 precursor (A2ML1), transmembrane protease, serine 11B (TMPRSS11B) and Transmembrane protease, serine 11D (TMPRSS11D) are found in nearly all samples at days 13 and 19. One protein, beta-crystallin S (CRYS), is found in 40 of 48 D-13 vaginal fluid samples (83%) and 16 of 48 D-19 samples (33%).

Table 18 Vaginal Fluid Marker Protein Expression on Days 13 & 19 of the Menstrual Cycle

PROTEIN MARKERS	D-13 EXPRESSION (%)	D-19 EXPRESSION (%)	BODY FLUID EXPRESSION	COMMENTS
SPRR3	48/48 (100)	48/48 (100)	found in saliva, reported in amniotic fluid and urine	Sys-BodyFluid reported in vaginal fluid
A2ML1	48/48 (100)	48/48 (100)	reported in saliva, and urine	Sys-BodyFluid reported in vaginal fluid
T**MPRSS11B**	48/48 (100)	47/48 (98)	**not reported in other body fluids**	Sys-BodyFluid reported in vaginal fluid
TMPRSS11D	44/48 (92)	46/48 (96)	reported in saliva	Sys-BodyFluid reported in vaginal fluid
CRYGS	40/48 (83)	16/48 (33)	**not reported in other body fluid**	structural components of the vertebrate eye lens
PLCD1	24/48 (50)	0/48 (0)	reported in plasma/serum	not found by us in serum/plasma
PZP	4/48 (8)	24/48 (50)	reported in bronchoalveolar lavage, cerebral spinal fluid, plasma/ serum, saliva, synovial fluid	not found by us in serum/plasma or saliva, we did not analyze other body fluids mentioned

SPRR3 = small proline-rich protein 3
A2ML1 = alpha-2-macroglobulin-like 1 precursor
TMPRSS11B = Transmembrane protease, serine 11B

TMPRSS11D = Transmembrane protease, serine 11D
CRYS = Beta-crystallin S
PLCD1 = phospholipase C delta-1
PZP = pregnancy zone protein

Of these markers, TMPRSS 11B and CRYGS have not been reported in other body fluids in the Sys-BodyFluid database, and consequently can be considered highly specific for vaginal fluid. TMPRSS 11D is also found in nearly all D-13 and D-19 vaginal fluid samples, but has been reported in saliva – although we have not been able to detect it there. Two other proteins that are found consistently in all vaginal fluid samples on D-13 and D-19 are SPRR3 and A2ML1; both are expressed in saliva and we have detected them there. SPRR3 has been reported inconsistently in amniotic fluid and urine proteomes and A2ML1 in urine. We found SPRR3 in all menstrual blood samples we examined and A2ML1 in approximately half of those samples. Neither has been reported in endometrium, making the vaginal canal their most likely source in menstrual blood.

Tissue expression of alpha 2-macroglobulin-like 1 (A2ML1) has been reported in human epidermis, testis, thymus, and placenta. It is a 180 kDa protein secreted by granular keratinocytes (Galliano et al. 2006) believed to bind to KLK7 (a protease that regulates epidermal desquamation by disrupting corneocyte adhesion (Caubet et al. 2004). Galliano et al. have also suggested that A2ML1 may function in defending against pathogens because it is known to inhibit proteases papain and subtilisin that are expressed in invading bacteria and aid surface adhesion. TMPRSS11B is a functional serine protease. This protein localizes in differentiating epithelial cells in stratified squamous epithelia from diverse tissue (Miller et al. 2014). Both proteins have functions consistent with their location in vaginal fluid.

Two other proteins of interest are phospholipase C delta-1 (PLCD1) and pregnancy zone protein (PZP). PLCD1 is not found in any of the 13 body fluids in the Sys-BodyFluid database. Interestingly, it is found in half of all D-13 samples, but not in any women's D-19 samples. If this expression pattern holds up in future studies, PLCD1 could become a useful marker for distinguishing pre- and post-ovulation time points in vaginal fluid samples. Nakamura & Fukami 2005 suggest that PLCD1 is the most abundant phospholipase C isozyme in the epidermis. Its expression increases when keratinocytes undergo differentiation, and as mentioned, keratinocytes function to fight pathogenic invasions (Swamy et al. 2010) by secreting cytokines that initiate local inflammatory responses. In mice, Nakamura et al., 2005 found that a decrease in PLCD1 results in placental vascular defects due to a decrease in number of vessels in the labyrinth layer.

PZP appears to have the opposite expression pattern, i.e. it is present in half of women's samples at D-19, but only 4 of 48 samples at D-13 (8%). Although PZP has been reported in several body fluids (Table 11), if it is found in a pure vaginal fluid sample (i.e. not mixed with another body fluid) it might also serve as a marker of where a woman is in her cycle. PZP is a trace protein found in the plasma of non-pregnant women and males; it is found in much higher levels in women than men (Folkersen et al., 1981). It is functionally and structurally related to A2ML (Sottrup-Jensen et al. 1984), both act as glycoprotein proteinase inhibitors in human

plasma. During pregnancy, levels of PZP increase substantially, reaching its peak in the last trimester. (Sand et al., 1985)

It is worth noting that all post-menopausal women showed expression of markers SPRR3, A2ML1, TMPRSS11B and TMPRSS11D, demonstrating that these markers can be identified in vaginal fluid of both women of childbearing age as well as postmenopausal women. Interestingly, neither PZP nor PLCD1 were identified in postmenopausal women. However, the sample size is too small to make a definitive conclusion about these markers.

ı) *THE VAGINAL MICROBIOME:* Several forensic studies have proposed using the vaginal microbiome as a means of identifying vaginal fluid as some of these microorganisms appear to be specific to, or enriched in, vaginal fluid (Fleming & Harbison 2010b; Doi et al. 2014; Benschop et al. 2012) Complete proteomic analysis of our samples identified numerous bacteria several of which have also been proposed as forensic vaginal fluid markers.

As shown in Table 19, and consistent with the literature (Tang et al. 2007), *Lactobacillus* is the most abundant bacteria found in vaginal fluid where it functions to maintain a low (acidic) vaginal pH by producing lactic acid and thereby preventing foreign bacterial pathogens from colonizing the vagina/cervix (Doderlein 1892). Recent studies have shown that a number of different bacteria are usually found to coexist in the vagina, and have been named the Community State Types (CST) (Ravel et al. 2010). Most of these are different strains of *Lactobacillus*: *L. crispatus*, *L. iners*, *L. jensenii*, *L. gasseri*, *Prevotella, Megsphaera*, *Gardnerella vaginalis*, *Sneathia* and *Atopobium vaginae* (Romero et al. 2014). Of these, our samples show the presence of *L. jensenii/gasseri* and *Gardnerellla vaginalis* (Table 19). Akutsu et al. (2012) reported *L. gasseri* and *Gardnerella vaginalis* are also found in sperm and therefore cannot be considered as markers solely for vaginal fluid, although Fleming et al. (2010) argues that *L. gasseri* along with *L. crispatus* are vaginal-specific bacteria and should therefore be considered for vaginal secretion identification.

Benschop et al. conducted a study comparing the microbiota of other body sites and bodily fluids and concluded that *Lactobacillus* is the only abundant microbe to be vagina-specific (see caveat above regarding *L. gasseri* in sperm), although they also noted that it was not found in all vaginal fluid samples tested. However, failure to detect *Lactobacillus* in some women could be the result of the time of month subjects were tested during the menstrual cycle. Steroidal hormones have been shown to affect the stability of the vaginal microbiota (Gupta et al. 2006) as well as differences between women whose CST remain stable and those whose CST levels vary (Zhou et al. 2007). Gajer et al. 2012 have shown that CSTs remain stable in the secretory phase of the menstrual cycle. The stability of the vaginal microbiome of pregnant women was shown to be greater than those of non-pregnant women (Romero et al. 2014) again suggesting that changes in CSTs may follow hormonal changes during the menstrual cycle. Finally, Ravel et al., 2010 have shown that African American and Hispanic women have a higher (more alkaline) vaginal pH, with CSTs dominated by anaerobic bacteria, and a lower abundance of *Lactobacillus*. These data may help to explain Benschop's finding that while *Lactobacillus* is a good marker for vaginal fluid, it is not found consistently in all women.

A second bacterium we identified, *Gardnerella vaginalis*, is commonly found in healthy women, although it is usually associated with bacterial vaginosis (BV). BV occurs when the

population of *Lactobacillus* in vagina declines and a mixture of other microorganisms, including *Gardnerella vaginalis*, supersede it (Albert et al. 2015).

Our detection of *Lactobacillus crispatus* in vaginal fluid samples is consistent with both the clinical and forensic literature of the vaginal microbiome, and while our data serve to confirm current efforts to identify vaginal fluid by its microbiome, we believe that the human vaginal fluid protein markers we identified give greater confidence for forensic testing.

Table 19 Bacterial Species found in Vaginal Fluid Samples

Species	Protein Accession #	Protein Name	No. Peptides 95% Confidence
Leishmania major strain Friedlin	gi|157866567	fatty acid elongase [Leishmania major strain Friedlin]	1
	gi|157879097	hypothetical protein [Leishmania major strain Friedlin]	1
Leishmania braziliensis MHOM/BR/75/M2904	gi|322504645	conserved hypothetical protein [Leishmania braziliensis MHOM/BR/75/M2904]	1
Lactobacillus sanfranciscensis TMW 1.1304	gi|347534272	glucose-6-phosphate isomerase [Lactobacillus sanfranciscensis TMW 1.1304]	1
Lactobacillus kefiranofaciens ZW3	gi|336053527	translation initiation factor IF-1 [Lactobacillus kefiranofaciens ZW3]	1
	gi|336053707	fructose-bisphosphate aldolase [Lactobacillus kefiranofaciens ZW3]	4
	gi|336054712	protein-N(pi)-phosphohistidine--sugar phosphotransferase [Lactobacillus kefiranofaciens ZW3]	1
	gi|336054882	NAD(P)H dehydrogenase [Lactobacillus kefiranofaciens ZW3]	1
	gi|336054906	beta-phosphoglucomutase [Lactobacillus kefiranofaciens ZW3]	2
	gi|336055394	L-lactate dehydrogenase 1 [Lactobacillus kefiranofaciens ZW3]	2
Lactobacillus johnsonii pf01/gasseri JV-V03/Lactobacillus gasseri 202-4	gi|42518797	glyceraldehyde 3-phosphate dehydrogenase [Lactobacillus johnsonii NCC 533]	1
	gi|8489169	GroEL [Lactobacillus johnsonii]	1
Lactobacillus helveticus MTCC 5463	gi|328461797	phosphoglycerate kinase [Lactobacillus helveticus MTCC 5463]	1
	gi|328464294	hypothetical protein AAULH_10822 [Lactobacillus helveticus MTCC 5463]	1
	gi|58337250	50S ribosomal protein L32 [Lactobacillus acidophilus NCFM]	1
Lactobacillus helveticus DSM 20075	gi|58336690	hypothetical protein LBA0352 [Lactobacillus acidophilus NCFM]	3
Lactobacillus fructivorans KCTC 3543	gi|339448758	MscS Mechanosensitive ion channel [Lactobacillus fructivorans KCTC 3543]	1
Lactobacillus curvatus CRL 705	gi|81429356	translation initiation factor IF-1 [Lactobacillus sakei subsp. sakei 23K]	1
Lactobacillus crispatus ST1	gi|227877217	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) [Lactobacillus crispatus JV-V01]	9
	gi|227878467	adenylosuccinate synthetase [Lactobacillus crispatus JV-V01]	3
	gi|295692022	endopeptidase o [Lactobacillus crispatus ST1]	2
	gi|295692086	hypothetical protein LCRIS_00224 [Lactobacillus crispatus ST1]	3
	gi|295692353	aggregation promoting factor [Lactobacillus crispatus ST1]	1
	gi|295692373	NADH peroxidase [Lactobacillus crispatus ST1]	1
	gi|295692830	citrate lyase ligase [Lactobacillus crispatus ST1]	1
	gi|295693047	hypothetical protein LCRIS_01185 [Lactobacillus crispatus ST1]	1
	gi|295693520	esterase [Lactobacillus crispatus ST1]	4

	gi|295693745	aggregation promoting protein [Lactobacillus crispatus ST1]	2
	gi|295693881	pyruvate oxidase [Lactobacillus crispatus ST1]	2
Lactobacillus crispatus MV-3A-US	gi|262045726	conserved hypothetical protein [Lactobacillus crispatus MV-3A-US]	4
	gi|262045885	30S ribosomal protein S20 [Lactobacillus crispatus MV-3A-US]	1
Lactobacillus crispatus CTV-05	gi|227876958	conserved hypothetical protein [Lactobacillus crispatus JV-V01]	3
	gi|227877122	PTS enzyme II, ABC component [Lactobacillus crispatus JV-V01]	2
	gi|227877219	triosephosphate isomerase [Lactobacillus crispatus JV-V01]	8
	gi|227877266	glucose-6-phosphate isomerase [Lactobacillus crispatus JV-V01]	7
	gi|227877346	conserved hypothetical protein [Lactobacillus crispatus JV-V01]	1
	gi|227877611	chaperone GroES [Lactobacillus crispatus JV-V01]	1
	gi|227878115	conserved hypothetical protein [Lactobacillus crispatus JV-V01]	1
	gi|227878359	ribosomal protein L31 [Lactobacillus crispatus JV-V01]	1
	gi|256843015	conserved hypothetical protein [Lactobacillus crispatus 125-2-CHN]	1
	gi|256848963	phosphotransferase system enzyme II [Lactobacillus crispatus MV-1A-US]	4
	gi|256850441	conserved hypothetical protein [Lactobacillus crispatus MV-1A-US]	1
	gi|262046701	thermostable pullulanase [Lactobacillus crispatus MV-3A-US]	4
	gi|312976672	putative esterase [Lactobacillus crispatus CTV-05]	1
	gi|312977004	conserved hypothetical protein [Lactobacillus crispatus CTV-05]	1
	gi|312977040	D-lactate dehydrogenase [Lactobacillus crispatus CTV-05]	12
	gi|312977149	phosphopyruvate hydratase [Lactobacillus crispatus CTV-05]	9
	gi|312977155	Dps family protein [Lactobacillus crispatus CTV-05]	1
	gi|312977162	putative membrane protein [Lactobacillus crispatus CTV-05]	1
	gi|312977181	thiol peroxidase [Lactobacillus crispatus CTV-05]	1
	gi|312978135	fructose-bisphosphate aldolase class-II [Lactobacillus crispatus CTV-05]	9
	gi|312983924	bacterial surface layer protein [Lactobacillus crispatus CTV-05]	5
	gi|312983946	aminopeptidase N [Lactobacillus crispatus CTV-05]	1
	gi|312983960	beta-phosphoglucomutase [Lactobacillus crispatus CTV-05]	9
	gi|312984022	conserved hypothetical protein [Lactobacillus crispatus CTV-05]	1
	gi|312984023	conserved hypothetical protein [Lactobacillus crispatus CTV-05]	1
Lactobacillus crispatus 214-1	gi|256844482	4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis protein [Lactobacillus crispatus 125-2-CHN]	2
	gi|293379757	pyruvate kinase [Lactobacillus crispatus 214-1]	5
	gi|293379782	glycosyl hydrolase family 25 [Lactobacillus crispatus 214-1]	3
	gi|293381809	multicopper oxidase [Lactobacillus crispatus 214-1]	2
	gi|293381879	pyridine nucleotide-disulfide oxidoreductase [Lactobacillus crispatus 214-1]	4
	gi|293382008	conserved hypothetical protein [Lactobacillus crispatus 214-1]	2
Lactobacillus amylovorus GRL1118	gi|315037882	PTS system, glucitol/sorbitol-specific IIA component [Lactobacillus amylovorus GRL 1112]	1
	gi|315037927	triosephosphate isomerase [Lactobacillus amylovorus GRL 1112]	4
	gi|327183582	Multicopper oxidase [Lactobacillus amylovorus GRL1118]	1
	gi|327184386	pyruvate oxidase [Lactobacillus amylovorus GRL1118]	1
Lactobacillus amylolyticus DSM 11664	gi|295425185	multidrug ABC superfamily ATP binding cassette transporter, ABC protein [Lactobacillus amylolyticus DSM 11664]	1
	gi|295426026	conserved hypothetical protein [Lactobacillus amylolyticus DSM 11664]	1

Lactobacillus acidophilus NCFM	gi|58336860	aspartyl/glutamyl-tRNA amidotransferase subunit B [Lactobacillus acidophilus NCFM]	1
	gi|58336965	p-enolpyruvate-protein p-transferase PTSI [Lactobacillus acidophilus NCFM]	1
	gi|58337164	aspartate-semialdehyde dehydrogenase [Lactobacillus acidophilus NCFM]	1
	gi|58337544	ribosome recycling factor Rrf [Lactobacillus acidophilus NCFM]	1
Lactobacillus acidophilus ATCC 4796	gi|58336361	30S ribosomal protein S6 [Lactobacillus acidophilus NCFM]	1
	gi|58336628	elongation factor G [Lactobacillus acidophilus NCFM]	1
	gi|58336644	50S ribosomal protein L6 [Lactobacillus acidophilus NCFM]	1
	gi|58336646	30S ribosomal protein S5 [Lactobacillus acidophilus NCFM]	1
	gi|58336674	4-methyl-5(b-hydroxyethyl)-thiazole monophosphatebiosynthesis [Lactobacillus acidophilus NCFM]	1
	gi|58336861	lipid kinase [Lactobacillus acidophilus NCFM]	1
	gi|58337021	triosephosphate isomerase [Lactobacillus acidophilus NCFM]	1
	gi|58337129	hypothetical protein LBA0821 [Lactobacillus acidophilus NCFM]	1
	gi|58337999	cell wall-associated hydrolase [Lactobacillus acidophilus NCFM]	1
	gi|58338082	oxidoreductase [Lactobacillus acidophilus NCFM]	1
	gi|58338169	D-alanine--poly(phosphoribitol) ligase subunit 1 [Lactobacillus acidophilus NCFM]	1
Gardnerella vaginalis HMP9231	gi|333394130	chaperone protein DnaK [Gardnerella vaginalis HMP9231]	2
	gi|308235801	putative phosphoketolase [Gardnerella vaginalis ATCC 14018]	1
	gi|333393776	translation elongation factor Ts [Gardnerella vaginalis HMP9231]	1
	gi|333602000	ABC transporter, solute-binding protein [Gardnerella vaginalis 315-A]	2
	gi|333602391	transaldolase [Gardnerella vaginalis 315-A]	6
	gi|333603302	LPXTG-motif cell wall anchor domain protein [Gardnerella vaginalis 315-A]	1

SECTION 3 -- SPECIES IDENTIFICATION FROM BLOOD USING MASS SPECTROMETRY: The goal of this aim was to determine whether species information can be recovered during routine body fluid testing by mass spectrometry. We proposed to examine blood from 11 species including seven domestic animals (cow, pig, sheep, chicken, turkey, dog and cat) as well as rabbit, deer, ferret and human for species-specific differences in protein sequences in abundantly expressed proteins. In addition to these evolutionary diverse animals we also examined blood from a primate close to humans - *Gorilla gorilla*.

- ***METHODS:*** Whole blood samples were used for human, sheep, rabbit, dog, cat, ferret and deer. Samples for cow, pig and chicken were obtained by collecting fluid following the freeze/thaw of skeletal muscle. One µl of samples were mixed with 9 µl of RIPA buffer (50 mM Tris-HCl pH7.4, 150 mM NaCl, 1% NP40, 0.1% SDS, 0.5% sodium deoxycholate) and homogenized using a FastPrep homogenizer. Following extraction, samples were centrifuged and supernatants were collected. Protein concentrations were measured by Bradford assay. Samples were digested with trypsin, cleaned by C18 reverse phase ZipTip chromatography, mixed with an equal volume of matrix and directly spotted on MALDI plates for detection. Amino acid sequence data was analyzed by ProteinPilot 3 software and blasted against protein databases.

• ***RESULTS:*** As seen in Table 21, all species can be identified by detection of multiple species-specific peptides from more than one protein, demonstrating that: i) specific protein purification prior to MALDI MS is not necessary, and ii) proteins other than hemoglobin can be used for species identification, and iii) such proteins, e.g. albumin, can further distinguish closely related species which could not be distinguished by hemoglobin alone. For example, sheep was identified along with its ancestor Mouflon (OVIMU) by using only hemoglobin, while albumin peptides could distinguish these two closely related species.

Cow and pig liquid samples were collected from frozen/thawed skeletal muscle (similar to possible stains present at a crime scene (kitchen) and therefore of forensic importance) and also evaluated by direct MALDI MS analysis. Although visually these samples looked like blood, direct MALDI MS analysis (i.e. without prior protein purification) did not identify hemoglobin or albumin, but instead, abundant muscle specific proteins such as: myoglobin, β-enolase, creatine kinase M-type (muscle type), carbonic anhydrase 3 (muscle specific), and phosphoglucomutase-1. As seen in Table 21, all of these proteins produced species-specific peptides easily detected by MALDI MS. (It should be noted that hemoglobins, albumin and other blood specific proteins were easily identified from these samples with prior SDS PAGE protein purification, or by LC/MALDI.) These results demonstrate that species detection by MALDI MS is not confined to blood, but can also be obtained from muscle. Indeed, it has been reported that species were identified from bone by analysis of collagen using MALDI MS (Buckley, 2009). Our data and the published results suggest that MALDI analysis can be a universal method for species identification without restriction of sample types.

It should also be pointed out that species identification by direct-spotting MALDI MS is a sensitive method. The data in Table 21 were collected from 60 ng of protein which is equivalent to approximately 0.2 nl of blood.

Table 21 Species-specific peptides identified from animal blood/muscle samples

Blood sample	Protein	Total Peptides Detected	Species Specific Peptides
***Homo sapiens* (human)**	HBB	9	EFTPPVQAAYQK FFESFGDLSTPDAVMGNPK GTFATLSELHCDKLHVDPENFR
	HBA	4	VGAHAGEYGAEALER
	ALBU	3	RPCFSALEVDETYVPK VFDEFKPLVEEPQNLIK
***Bovine* (cow) <u>skeletal muscle</u>**	ENOB	5	HIADLAGNPELILPVPAFNVINGGSHAGNK
	MYG	3	VEADVAGHGQEVLIR GLSDGEWQLVLNAWGK YLEFISDAIIHVLHAK

	PGM1	1	VFQSSSNYAENFIQSIISTVEPAQR
	CAH3	1	EAPFNNFNPSCLFPACR
	KCRM	2	DLFDPIIQDR
***Porcine* (pig) <u>skeletal muscle</u>**	ENOB	3	NYPVVSIEDPFDQDDWK
	KCRM	6	LSVEALNSLTGEFK
	CAH3	1	EAPFTNFNPSCLFPACR
	MYG	1	
***Gallus* (chicken)**	HBB	2	NTFSQLSELHCDKLHVDPENFR
	HBA	3	IAGHAEEYGAETLER KVVAALIEAANHIDDIAGTLSK
***Feline* (cat)**	HBB	7	LLGNVLVCVLAHHFGHDFNPQVQAAFQK
	HBA	3	IGSHAGEYGAEALER VADALTQAVAHMDDLPTAMSALSDLHAYK
***Canine* (dog)**	HBB	9	FFDSFGDLSTPDAVMSNAK
	HBA	6	TFQSFPTTK IGGHAGDYGGEALDR KVADALTTAVAHLDDLPGALSALSDLHAYK
***Odocoileus virginianus* (deer)**	HBA	6	LLSHSLLVTLATHLPDNFTPAVHASLDK VGGNAGAYGAEALER
	HBB	8	VNVDVVGAEALGR NFGGEFTPLVQADFQK GAFAELSELHCNKLHVDPENFR VLDAFSDGLK FFDSFGDLSSADAVMGNPK
***Leporidae* (rabbit)**	HBA	4	IGSHGGEYGAEAVER TYFPHFDFTHGSEQIK LLSHCLLVTLANHHPSEFTPAVHASLDK
	HBB	2	VLAAFSEGLSHLDNLK
	ALBU	9	KVPQVSTPTLVEISR LPCVEDYLSVVLNR RPCFSALGPDETYVPK DDKPDLPPFARPEADVLCK EFNAETFTFHADICTLPETER

			QNCELYEQLGDYNFQNALLVR
***Ovine* (sheep)**	HBA	4	VGGNAGAYGAEALER LLSHTLLVTLACHLPNDFTPAVHASLDK
	HBB	8	HHGNEFTPVLQADFQK FFEHFGDLSNADAVMNNPK
	ALBU	4	FFTFHADICTLPDTEK RPCFSDLTLDETYVPKPFDEK
***Mustela putorius furo* (ferret)**	HBB	7	VNVDEVGGETLGR FFDSFGDLSSPDAVMSNPK
	HBA	6	VADALTNAVAHMDDLPGALSALSDLHAYK
	ALBU	3	

HBB, Hemoglobin-β; HBA, Hemoglobin-α; ALBU, Albumin; ENOB, β-enolase; MYG, Myoglobin; CAH3, Carbonic anhydrase 3; KCRM, Creatine kinase M-type; PGM1, Phosphoglucomutase-1.

Finally, distinguishing humans from other primates by hemoglobin alone is not always possible as many Hb protein sequences are identical or nearly identical between primates (Table 22). However many other abundant blood proteins can be identified by our MALDI MS assay including: albumin, serotransferrin, carbonic anhydrase and apolipoprotein. As *in silico* sequence alignments of these proteins show differences between primates, they make excellent targets for distinguishing humans from other primates.

Table 22 Sequence Alignments of Primates Hemoglobins

α-Hemoglobin				
Human	Length (AA)	Species	Length (AA)	Score
HUMAN	142	PANTR	142	100
HUMAN	142	PANPA	142	100
HUMAN	142	PONAB	142	100
HUMAN	142	GORGO	141	99

β Hemoglobin				
Human	Length (AA)	Species	Length (AA)	Score
HUMAN	147	PANTR	147	100
HUMAN	147	PANPA	147	100
HUMAN	147	GORGO	147	99
HUMAN	147	PONAB	147	93

Scores (percentage) are calculated as the number of identities in the best alignment divided by the number of residues compared. HUMAN, human; PANTR, chimpanzee; PANPA, pygmy chimpanzee; GORGO, lowland gorilla; PONAB, Sumatran orangutan.

I) *PRIMATE SPECIES IDENTIFICATION:*

i) ***Materials & Methods*** **--** A sample of gorilla blood, obtained from the World Wildlife Society, was processed as was done for the 11 species described above. In brief: One volume of whole gorilla blood was mixed with nine volumes of RIPA buffer (50 mM Tris-HCl pH7.4, 150 mM NaCl, 1% NP40, 0.1% SDS, 0.5% Sodium deoxycholate) and homogenized using a FastPrep homogenizer. Following solubilization, the samples were centrifuged and supernatants were collected. Protein concentrations were measured by the Bradford assay using BSA as standard. Thirty µg of supernatant protein were reduced, alkylated and digested with trypsin. Two µg total peptides were loaded to a Dionex 3000 HPLC and separated on a C18 reverse phase nano-column using a 60 minute, 5% to 45 % continuous gradient of acetonitrile. Eluted fractions were mixed with matrix and spotted on MALDI plates for analysis by an AB Sciex 4800 MALDI TOF/TOF. Data were analyzed by AB Sciex ProteinPilot 4.0 software, and compared with the UniProtKB protein database with no species selected during database search.

ii)***Results*** **--** As expected, sequence homologies between human and gorilla proteins are quite high. Figure 32A, for example, shows that only one amino acid (R to K substitution at position 105) differentiates human and gorilla hemoglobin-β (**Hb B**); ~99.3% homology. Similarly, Figure 32B reveals only two amino acids (the initial methionine removed after translation in gorilla, and an E to D substitution at position 24) differentiate human and gorilla hemoglobin-α (**Hb A**); ~98.6% homology. However, these differences are sufficient to allow species differentiation by mass spectrometry.

Figure 32A Alignment of Human and Gorilla Hb B Protein Sequences

```
HBB_HUMAN MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK 60
HBB_GORGO MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK 60
          ************************************************************

HBB_HUMAN VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
HBB_GORGO VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFKLLGNVLVCVLAHHFG 120
          ********************************************:***************

HBB_HUMAN KEFTPPVQAAYQKVVAGVANALAHKYH 147
HBB_GORGO KEFTPPVQAAYQKVVAGVANALAHKYH 147
          *************************
```

Figure 32B Alignment of Human and Gorilla Hb A Protein Sequences

```
HBA_HUMAN MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG 60
HBA_GORGO -VLSPADKTNVKAAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG 59
          ***********************:************************************

HBA_HUMAN KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 120
HBA_GORGO KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 119
          ************************************************************

HBA_HUMAN AVHASLDKFLASVSTVLTSKYR 142
HBA_GORGO AVHASLDKFLASVSTVLTSKYR 141
          **********************
```

Figure 33A shows MS/MS spectra of gorilla Hb B peptide LHVDPENF**K** which differentiates it from human Hb B peptide (LHVDPENF**R**).

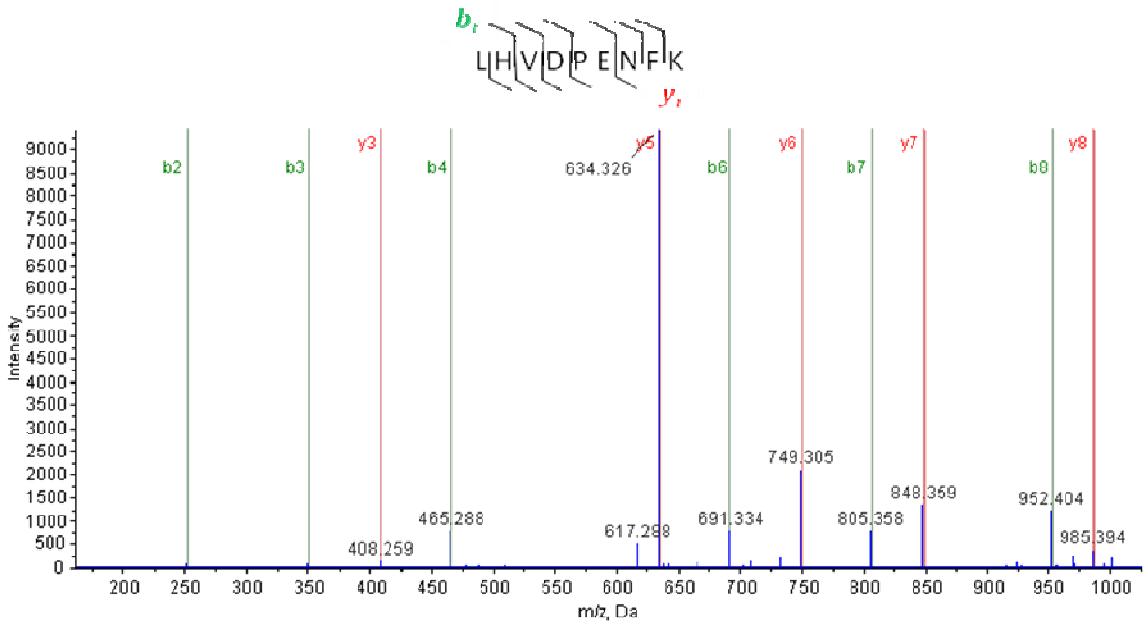


FIG 33A MS/MS OF GORILLA HB B PEPTIDE LHVDPENFK

Figure 33B shows MS/MS spectra of gorilla Hb A peptide VGAHAG**D**YGAEALER which differentiates it from human (VGAHAG**E**YGAEALER).

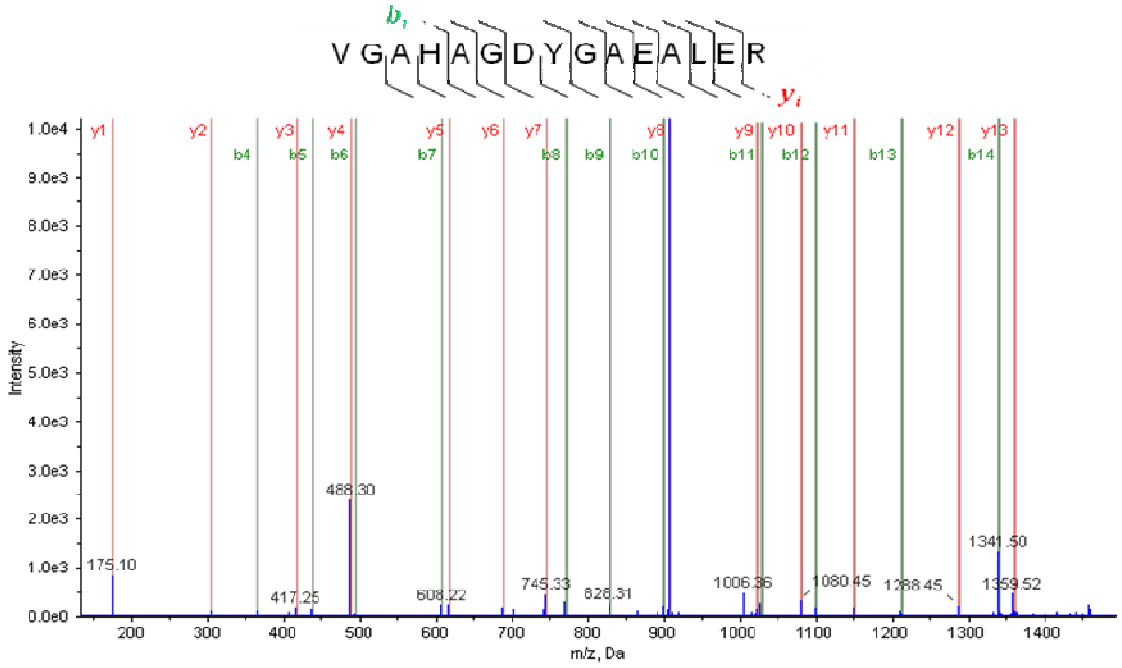


FIG 33B MS/MS OF GORILLA HB B PEPTIDE VGAHAGDYGAEALER

We also identified several other proteins that could distinguish human blood form gorilla (e.g. catalase; carbonic anhydrase 1).

• CONCLUSION: These results demonstrate that even closely related species (including primates) can be identified from very small volumes of whole blood (~0.2 nl) by mass spectrometry analysis. These data are part of data acquisition during the mass spec process and do not require any additional sample preparation to achieve.

DISCUSSION AND CONCLUSIONS –

• DISCUSSION OF FINDINGS: The major goals of this grant were to use mass spectrometry to identify menstrual blood and vaginal fluid protein markers that would be useful for the forensic identification of these body fluids, and determine if the identified markers were present in all women regardless of age, ethnicity, use of hormonal contraception and, for menstrual blood, present on each day of women's periods, while for vaginal fluid, present on the two most distinct days of the menstrual cycle – the peak of the proliferative phase (Day 13) and peak of the secretory phase (Day 19) – when hormonal and endometrial cell changes are greatest. These goals have been achieved.

Five proteins were consistently and robustly identified in all menstrual blood samples and can easily distinguish menstrual blood from venous blood. These proteins are: 1) progestagen-associated endometrial protein (PAEP aka Glycodelin); 2) glycogen phosphorylase, liver form; 3) small proline-rich protein 3; 4) ADP/ATP translocase 2 and 5) keratin, type II cytoskeletal 7.

Two of these proteins, progestagen-associated endometrial protein and ADP/ATP translocase 2, have been reported in comprehensive proteomic analyses of venous blood at low levels. However, we have not detected these proteins in the 45 venous blood samples used in this study. Progestagen-associated endometrial protein is an excellent marker for distinguishing menstrual blood from venous blood as it is a well documented endometrial protein and known to increase in the secretory phase of the menstrual cycle (Strauss & Barbieri 2013) However, we, and others (Pilch & Mann 2006; Chiu et al. 2007), have also detected PAEP in seminal fluid. Nevertheless, the presence of PAEP in a blood stain in the absence of the specific and abundant seminal fluid markers semenogelin 1 and 2 would suffice to exclude the possibility of a mixture of semen and menstrual blood.

Except for PAEP, the four remaining proteins have been identified in the Sys-BodyFluid database to be present at varying levels in saliva. (We routinely find only small proline-rich protein 3 in saliva, likely due to its high abundance.) However, here again, the absence of other saliva specific and abundant markers (e.g. alpha-amylase and cystatin-SA) would be sufficient to determine that the source of these proteins are not saliva. Finally, it should be noted that small proline-rich protein 3 is also present in and we consider a marker for vaginal fluid (see below). However, a mixture of *venous* blood and vaginal fluid could not be mistaken for menstrual blood as it would not have the remaining nine menstrual blood markers.

It is important to point out that the markers identified in this study are proteins inherent in uterine function and consistent with the dynamic processes of the menstrual cycle – in particular the secretory phase (the period just prior to menstruation) and the beginning of menstruation which today is regarded as an inflammatory response (Evans & Salamonsen 2012)– and accounts for the significant levels of natural killer cell specific proteins we identify. In the secretory phase, the endometrium has reached its full thickness and cells are preparing for

blastocyst implantation through increased metabolic activity, protein synthesis and secretion. Increased levels of PAEP as well as ADP/ATP translocase 2 & 3, phosphate carrier protein and cytochrome c oxidase subunit 2 are reflective of this stage of the menstrual cycle. Lack of implantation leads to increases in proteolytic enzymes, proteins involved in apoptosis as well as proteins from the increased numbers of immune cells that are an integral part of menstruation (6–9 (Evans & Salamonsen 2012)), and, of course, proteins from the breakdown of the uterine wall. These proteins include serpin B13, granulysin, glucosidase 2 subunit beta, parathymosin and keratin, type II cytoskeletal 7.

Finally, for all menstrual blood markers identified there was no difference between their expression on different days of a woman's period, or on a woman's age, ethnicity or use of hormonal contraception – meaning these markers are commonly expressed during a woman's entire period. Perhaps the most significant finding of this work, however, was the discovery that multivariate unsupervised hierarchical clustering analysis of menstrual and venous blood naturally separated them into their respective types. These data suggest a new method for serology analysis not restricted to menstrual blood but to all body fluids and body fluid mixtures.

For vaginal fluid, four markers were identified that are consistently found in all women on days 13 and 19 of their menstrual cycle regardless of age, ethnicity or use of hormonal contraception. Of these, TMPRSS11B is not reported in any other body fluids, and as such is an excellent candidate for development of a targeted assay for forensic identification of vaginal fluid. Additionally, three markers were identified that appear to be differentially expressed, i.e. found more frequently on day 13 or 19. If these data are confirmed by further studies it may possible that they could be used to roughly determine the point in the menstrual cycle from which a vaginal stain was derived. Finally, by searching our data against the entire proteomic database (all species) we were able to confirm the presence of vaginal microbiome specific organisms that have been proposed as forensic identifiers for vaginal fluid.

We believe that a validated menstrual blood and vaginal fluid assay can be established from these findings.

Tandam mass spectrometry effectively sequences all peptides it detects. As with DNA, inherant in protein sequence information is species identity. Consequently, without any additional assay requirements (sample preparation or instrument programing) species are determined. Using blood from 11 different animals (cat, chicken, cow, deer, dog, ferret, pig, rabbit, sheep, turkey and human) we were easily able to identify species by differences in multiple common blood proteins (albumin, β-enolase, carbonic anhydrase 3, creatine kinase M-type, hemoglobin-α, hemoglobin-β, myoglobin and phosphoglucomutase-1) using HPLC-MS/MS with as little as 0.2 nl of blood. Additionally, because the abundance of these proteins in blood, identifications can be made by MS/MS alone (i.e. without HPLC purification) saving significant amounts of time. Finally, comparing human and gorilla blood, we demonstrate the ability to distinguish primates by protein sequence differences in the common blood proteins hemoglobin-α, hemoglobin-β, carbonic anhydrase 1 and catalase.

● IMPLICATIONS FOR POLICY AND PRACTICE: The goal of this work was to discover protein markers that could be used to identify menstrual blood and vaginal fluid. This goal has been achieved. The ultimate purpose of this work is to use this knowledge to establish robust, confirmatory and routine forensic assays to aid in criminal investigations, and provide evidence of unquestioned reliability. Protein analysis by mass spectrometry effectively sequences proteins. These data then are similar in their specificity, quality and reliability to DNA analysis of STRs and, consequently, may be used to establish confirmatory serology tests of unparalleled reliability. Nor should such testing be limited to menstrual blood and vaginal fluid. We have already demonstrated (Yang et al. 2013) that serology testing by mass spectrometry for blood, saliva and semen are robust, confirmatory, rapid and more cost effective than other methods currently

in the market. Consequently, we believe that policy makers should press for confirmatory serology assays for all relevant forensic samples and continue funding the basic and applied research in order to bring these assays into routine use in public laboratories. As DNA analysis has helped balance the scales of justice by providing impartial evidence – both incriminatory and exculpatory - so too will the establishment of confirmatory serology testing.

• IMPLICATIONS FOR FURTHER RESEARCH: The discovery of menstrual blood and vaginal fluid protein markers in over 300 samples from fifty women demonstrates their reliability to accurately identify these body fluids. However, additional basic and applied research needs to be done to take these discoveries from the research lab to routine forensic laboratory testing. Additional samples should be evaluated to confirm these data. Robust extraction and downstream processing methods need to be established along with SOPs. Limits of detection need to be determined from both "pure" samples and mock forensic samples applied to multiple substrates to determine if such substrates affect recovery. Finally, mixtures of body fluids need to be examined to determine the limits of detection in mixed samples.

REFERENCES –

Adkins, R. et al., 2011. Racial Differences in Gene Specific DNA methylation levels are present at birth. *Birth Defects Res A Clin Mol Teratol*, 91, pp.728–736.

Akutsu, T. et al., 2012. Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid. *Legal Medicine*, 14(3), pp.160–162.

Albert, A.Y.K. et al., 2015. A study of the vaginal microbiome in healthy Canadian women utilizing cpn60-based molecular profiling reveals distinct Gardnerella subgroup community state types. *PLoS ONE*.

Andraos, C. et al., 2011. Vitamin D receptor gene methylation is associated with ethnicity, tuberculosis, and Taq1 polymorphism. *Hum Immunol*, 72, pp.262–268.

Anon, RNase and DEPC Treatment: Fact or Laboratory Myth. *Ambion Technical Bulletin 178*.

Avci, R. et al., 2005. Preservation of Bone Collagen from the Late Creataceous period studied by immunological techniques and atomic force microscopy. *Langmuir*, 21(8), pp.3584-3590.

Bada, J. et al., 1999. Preservation of key biomolecules in the fossil record: current knowledge and future challenges. *Phil Trans R Soc Lond B*, 354(1379), pp.77-87.

Bauer, M. & Patzelt, D., 2008. Identification of menstrual blood by real time RT-PCR: Technical improvements and the practical value of negative test results. *Forensic Science International*, 174(1), pp.55–59.

Bencharit, S. et al., 2011. Defining salivary biomarkers using mass spectrometry based proteomics: a systematic review. *OMICS: A Journal of Integrative Biology*, 15, p.353+.

Benjamin, Y. & Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B*, pp.289–300.

Benschop, C.C.G. et al., 2012. Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? *International Journal of Legal Medicine*, 126(2), pp.303–310.

Berezikov, E. et al., 2005. Phylogenetic Shadowing and Computational Identification of Human microRNA genes. *Cell*, 120, pp.21–24.

Blank, A. & Dekker, C., 1981. Human body fluid ribonucleases: detection, interrelationships, and significance. *Nucleic Acids Symposium Series*, 10, pp.203–9.

Brewis, I. & Gadella, B., 2010. Sperm surface proteomics: from protein lists to biological function. *Molecular Human Reproduction*, 16, pp.68–79.

Brown, T., Mackey, K. & Du, T., 2004. Analysis of RNA by northern and slot blot hybridization. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, Chapter 4, p.Unit 4.9.

Buckley, M., et al., *Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry.* Rapid Commun Mass Spectrom, 2009. **23**(23): p. 3843-54.

Cataldo, F. 2006. Ozone degradation of biological macromolecules: proteins, hemoglobin, RNA, and DNA. Ozone: *Science & Engineering*. 28, pp. 317-328.

Caubet, C. et al., 2004. Degradation of corneodesmosome proteins by two serine proteases of the kallikrein family, SCTE/KLK5/hK5 and SCCE/KLK7/hK7. *Journal of Investigative Dermatology*, 122(5), pp.1235–1244.

Chiu, P.C.N. et al., 2007. Glycodelin-A interacts with fucosyltransferase on human sperm plasma membrane to inhibit spermatozoa-zona pellucida binding. *Journal of cell science*.

Choi, S. & Friso, S., 2010. Epigenetics: A New Bridge between Nutrition and Health. *Adv Nutr*, 1, pp.8–16.

DeSouza, L. et al., 2005. Proteomic analysis of the proliferative and secretory phases of the human endometrium: Protein identification and differential protein expression. *Proteomics*.

Doderlein, A., 1892. Das Scheidensekret und seine Bedeutung fiir Puerperalheber. *Zbl. Bakt*, 11, p.699.

Doi, M. et al., 2014. A simple identification method for vaginal secretions using relative quantification of Lactobacillus DNA. *Forensic Science International: Genetics*, 12, pp.93–99.

Eckhardt, F. et al., 2006. DNA methylation Profiling of Human Chromosomes 6,20, and 22. *Nature Genetics*, 38, pp.1378–1385.

Egger, G. et al., 2004. Epigenetics in Human Disease and Prospects for Epigenetic Therapy. *Nature*, 429, pp.457–463.

Ehrich, M. et al., 2007. A New Method for Accurate Assessment of DNA Quality after Bisulfite Treatment. *Nucleic Acids Research*, 35, p.e29.

El-Maari, O. et al., 2007. Gender Specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Human Genetics*, 122, pp.505–514.

Ende, N., 1961. AMYLASE ACTIVITY I N BODY FLUIDS. *Cancer*, 14(September-October), pp.1109–1114.

Evans, J. & Salamonsen, L.A., 2012. Inflammation, leukocytes and menstruation. *Reviews in Endocrine and Metabolic Disorders*, 13(4), pp.277–288.

Fleming, R.I. & Harbison, S., 2010a. The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. *Forensic Science International: Genetics*, 4(4), pp.244–256.

Fleming, R.I. & Harbison, S., 2010b. The use of bacteria for the identification of vaginal secretions. *Forensic Science International: Genetics*.

Foran, D.R, et al., 2015. A comparison of mitochondrial DNA Amplification strategiesfor species identification. *J Forensic Investigation*, 33(72), pp.1-7.

Frumkin, D. et al., 2011. DNA methylation-based forensic tissue identification. *Forensic science international. Genetics*, 5, pp.517–524.

Gaensslen Ph.D, R.E., 2000. Forensic Analysis of Evidence. In C. Wecht, ed. *Forensic Sciences*. New York: Matthew Bender and Co.

Gajer, P. et al., 2012. Temporal dynamics of the human vaginal microbiota. *Science translational medicine*, 4(132), p.132ra52. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3722878&tool=pmcentrez&rendertype=abstract.

Galliano, M.F. et al., 2006. A novel protease inhibitor of the α2-macroglobulin family expressed in the human epidermis. *Journal of Biological Chemistry*, 281(9), pp.5780–5789.

Ganesh, M. et al., 2009. Plasma Fibrinogen Levels during the Different Phases of Menstrual Cycle - A Scientific Insight. *Sri Ramachandra Journal of Medicine*, 1(1), pp.21–23.

Gilman, M., 2002. Preparation of cytoplasmic RNA from tissue culture cells. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, Chapter 4, p.Unit 4.1.

Gronniger, E. et al., 2010. Aging and Chronic Sun Exposure Cause Distinct Epigenetic Changes in Human Skin. *PLoS genetics*, 6, p.e1000971.

Grunau, C., Clark, S. & Rosenthal, A., 2001. Bisulfite Genomic Sequencing: Systemic Investigation of Critical Experimental Parameters. *Nucleic Acids Research*, 29, pp.E65–65.

Gupta, S. et al., 2006. Vaginal microflora in postmenopausal women on hormone replacement therapy. *Indian journal of pathology & microbiology*, 49(3), pp.457–461. Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=17001923.

Haas, C. et al., 2008. mRNA profiling for body fluid identification. *Forensic Science International: Genetics Supplement Series*, 1(1), pp.37–38.

Haas, C. et al., 2009. mRNA profiling for body fluid identification by reverse transcription

endpoint PCR and realtime PCR. *Forensic Science International: Genetics*, 3(2), pp.80–88.
Haas, C. et al., 2014. RNA/DNA Co-Analysis from Human Menstrual Blood and Vaginal Secretion Stains: Results of a Fourth and Fifth Collaborative EDNAP Exercise. *Forensic science international. Genetics*, 8, pp.203–212.
Hachani, J. et al., 2011. MALDI-TOF MS profiling as the first-tier screen for sickle-cell diseasein neonates: Matching throughput to objectives. *Proteomics-Clinical Applications*, 5, pp.405–414.
Hanson, E., Lubenow, H. & Ballantyne, J., 2009. Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Forensic Science International: Genetics Supplement Series*, 2(1), pp.503–504.
Hernandez - Triana, L. et al., 2015. DNA barcoding of neotropical black flies (Diptera:simulidae): Species identification and discovery of cryptic diversity in Mesoamerica. *Zootaxa, 3936*(I), pp.93-114.
Ibberson, D. et al., 2009. RNA Degradation Compromises the Reliability of microRNA Expression Profiling. *BMC Biotechnol*, 9(102).
Isaaq, H., Waybright, T. & Veenstra, T., 2011. Cancer Biomarker Discovery: Opportunities and pitfalls in analytical methods. *Electrophoresis*, 32, pp.967–975.
Jakubowska, J. et al., 2014. mRNA heptaplex protocol for distinguishing between menstrual and peripheral blood. *Forensic Science International: Genetics*, 13, pp. 53-60.
Jesperson J, 1983. Sequential Study of Plasma Euglobulin Fibrinolytic Activity during the Normal Menstrual Cycle and in Women on Oral Contraceptives Low in Estrogen. *Gynecologic and Obstetric Investigation*, 15(5), pp.266–274.
Jesperson, J. & Nielson, M., 1989. Levels of Protein S during the Normal Menstrual Cycle and in Women on Oral Contraceptives Low in Estrogen. *Gynecologic and Obstetric Investigation*, 28(2), pp.82–86.
Jung, M. et al., 2010. Robust microRNA Stability in Degraded RNA Preparations from Human Tissue and Cell Samples. *Clinical Chemistry*, 56, pp.998–1006.
Juusola, J. & Ballantyne, J., 2007. mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. In *Journal of Forensic Sciences*. pp. 1252–1262.
Juusola, J. & Ballantyne, J., 2005. Multiplex mRNA profiling for the identification of body fluids. *Forensic Science International*, 152(1), pp.1–12.
Kohn, J., 1986. Measurement of amylase activity-- a useful indication of tissue damage after major radiation accidents? *Br Med J (Clin Res Ed)*, 292(6534), pp.1523–4.
Lewis, B., Burge, C. & Bartel, D., 2005. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that thousands of Human Genes are mirco RNA Targets. *Cell*, 120, pp.15–20.
Li, S.J. et al., 2009. Sys-BodyFluid: A systematical database for human body fluid proteome research. *Nucleic Acids Research*, 37(SUPPL. 1).
Lill, J. et al., 2007. Microwave-assisted Proteomics. *Mass Spectrometry Reviews*, 26, pp.657–671.
Lim, L. et al., 2005. Microarray analysis shows that some micro RNAs downregulate large numbers of target mRNAs. *Nature*, 433, pp.769–773.
Lindenbergh, A., et al., 2012. A multiplex m(RNA)-profiling system for forensic identification of body fluids and contact traces. *Forensic Science International: Genetics*, 6, pp/565-577.
Miller, G.S. et al., 2014. HATL5: A cell surface serine protease differentially expressed in epithelial cancers. *PLoS ONE*, 9(2).
Mor, E. et al., 2011. Species-specific microRNA Roles Elucidated Following Astrocyte Activation. *Nucleic Acids Research*, 39, pp.3710–3723.
Nakamura, Y. & Fukami, K., 2005. Physiological functions of phospholipase Cdelta. *Seikagaku. The Journal of Japanese Biochemical Society.*, 77(11), pp.1431–1435.
Negi, R.K., et al., 2016. Application of computational methods in fish species identification

based on mitochondrial DNA sequences. *Current Science*, 110(10), pp.2172-2176.
Nielson Marsh, CM., 2002. Biomolecules in fossil remains - Multidisciplinary approach to endurane. *The Biochemist*, 24(3), pp.12-14
Nussbaumer, C., Gharehbaghi-Schnell, E. & Korschineck, I., 2006. Messenger RNA profiling: A novel method for body fluid identification by Real-Time PCR. *Forensic Science International*, 157(2-3), pp.181–186.
Patzelt, D., 2004. History of Forensic Serology and Molecular Genetics in the Sphere of Activity of the German Society for Forensic Medicine. *Forensic Science International*, 144, pp.185–1991.
Peirson SN, Butler JN., 2007. RNA extraction from mammalian tissues. *Methods Mol Biol,* 362: pp.315-327. Edited by: E Rosato © Human Press Inc., Totowa, NJ
Pigati, L. et al., 2010. Selective Release of MicroRNA Species from Normal and Malignant Mammary Epithelial Cells. *PloS one*, 5, p.e13515.
Pilch, B. & Mann, M., 2006. Large-Scale and High-Confidence Proteomic Analysis of Human Seminal Plasma. *Genome Biology*, 7(5), p.R40.
Ravel, J. et al., 2010. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences*, 108(Supplement_1), pp.4680–4687. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3063603&tool=pmcentrez&rendertype=abstract.
Remily-Wood, E. et al., 2011. A database of reaction monitoring mass spectrometry assays for elucidating therapeutic response in cancer. *Proteomics-Clinical Applications*, 5, pp.383–396.
Richard, M. et al., 2011. Evaluation of mRNA Marker Specificty for the Identification of Five Human Body Fluids by Capillary Electrophoresis. *Forensic science international. Genetics*.
Righetti, P. et al., 2005. Proteome analysis in the clinical chemistry laboratory: Myth or Reality? *Clinica chimica acta; international journal of clinical chemistry*, 357, pp.123–139.
Rivera, R. et al., 1999. The Mechanism of Action of Hormonal Contraceptives and Intrauterine Contraceptive Devices. *American Journal of Obsterics and Gynecology*, 181(5), p.pp1263-1269.
Romero, R. et al., 2014. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome*, 2(1), p.18. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4066267&tool=pmcentrez&rendertype=abstract.
Saferstein, R., 1982. *Forensic Science Handbook*, Eaglewood Cliffs, Prentice Hall Reagents.
Saito, Y. et al., 2006. Specific Activation of microRNA-127 with Downregulation of the Proto-oncogene BCL6 by Chromatin-Modifying Drugs in Human Cancer Cells. *Cancer Cell*, 9, pp.435–443.
San Antonio, J.D. et al., 2011. Dinosaur peptides suggest mechanisms of protein survival. *PLoS One*. 6(6): e20381
Schweitzer, et al.,2005. Molecular preservation in late Cretaceous sauropod dinosaur eggshells.*Proc Biol Sci*, 272, pp775-84.
Scott, G. et al., 2006. Rapid Alteration of microRNA Levels by Histone Deacetylase Inhibition. *Cancer Research*, 66, pp.1277–1281.
Shaw, J.L. V & Diamandis, E.P., 2007. Distribution of 15 human kallikreins in tissues and biological fluids. *Clinical Chemistry*, 53(8), pp.1423–1432.
Shomron, N., Golan, D. & Hornstein, E., 2009. An evolutionary perspective of animal microRNAs and their targets. *J Biomed Biotechnol*, 2009.
Sottrup-Jensen, L. et al., 1984. Partial primary structure of human pregnancy zone protein: extensive sequence homology with human alpha 2-macroglobulin. *Proceedings of the*

National Academy of Sciences of the United States of America, 81(23), pp.7353–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=392144&tool=pmcentrez&rendertype=abstract.

Sparbier, K. et al., 2009. Immuno-MALDI-TOF MS: New perspectives for clinical applications of mass spectrometry. *Proteomics*, 9, pp.1442–1450.

Strauss, J.F. & Barbieri, R.L., 2013. *Yen and Jaffe's Reproductive Endocrinology* 7th ed.,

Swamy, M. et al., 2010. Epithelial decision makers: in search of the "epimmunome". *Nature immunology*, 11(8), pp.656–665. Available at: http://dx.doi.org/10.1038/ni.1905.

Tang, L.-J. et al., 2007. Proteomic Analysis of Human Cervical-Vaginal Fluid. *Journal of Proteome Research*, 6(7), pp.2874–2883.

Tanner, M.A. et al., 2002. Substantial changes in gene expression level due to the storage temperature and storage duration of human whole blood. *Clin Lab Haematol*, 24(6), pp.337–341. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12452813.

Thacker, S. et al., 2011. Evaluation of Sperm Proteins in Infertile Men: A proteomic approach. *Fertility and Sterility*, 95, pp.2745–2748.

Vercelli, D., 2004. Genetics, Epigenetics, and the Environment: Switching, buffering, releasing. *J Allergy Clin Immunol*, 113, pp.381–386.

Wang, Z. et al., 2012. A Model for Data Analysis of microRNA Expression in Forensic Body Fluid Identification. *Forensic science international. Genetics*, 6, pp.419–423.

Westhoff, C., 2003. Depot-medroxyprogesterone acetate injection (Depo-Provera??): A highly effective contraceptive option with proven long-term safety. *Contraception*.

Whitehead, P.H. & Divall, G.B., 1973. Assay of "Soluble Fibrinogen" in Bloodstain Extracts as an Aid to Identification of Menstrual Blood in Forensic Science: Preliminary Finding. *Clinical Chemistry*, 19(7), pp.762–765.

Xu, Y. et al., 2014. Development of highly sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification. *PLoS One*, 9(7), e100123, pp.1 -16

Yang, H. et al., 2013. Body Fluid Identification by Mass Spectrometry. *International Journal of Legal Medicine*, 127, pp.1065–1077.

Yang, L. et al., 2014. Species identification through mitochondrial rRNA genetic analysis. *Scientific Reports*, 4(4089), pp.1-11.

Zegels, G. et al., 2009. Comprehensive proteomic analysis of human cervical-vaginal fluid using colposcopy samples. *Proteome Science*.

Zhou, X. et al., 2007. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. *The ISME Journal*, 112(10), pp.121–133.

Zubakov, D. et al., 2010. MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International Journal of Legal Medicine*, 124(3), pp.217–226.

DISSEMINATION OF RESEARCH FINDINGS –

- PUBLICATIONS:

Yang, H., Zhou, B., Prinz, M., Siegel, D. (2012). Proteomic Analysis of Menstrual Blood. Molecular & Cellular Proteomics 11:1024 – 1035

Yang, H., Zhou, B., Deng, H., Prinz, M., Siegel, D.(2013). Body Fluid Identification by Mass Spectrometry. Int. J of Legal Medicine. 127: 1065-1077

- PRESENTATIONS:

Primate & Other Species Identification by Proteomic Analysis, Yang H, Zhou B, Prinz M, Siegel D, presented in the Forensic Section of the 61st American Society of Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, June 2013 in Minneapolis, MN.

Proteomic Applications in Forensic Science, Yang H, Siegel D, presented at the 62nd Annual Conference of the American Society of Mass Spectrometry and Allied Topics, June 2014 in Baltimore, MD.

Forensic Serology Testing by Mass Spectrometry, Yang H, Siegel D, presented at the 63rd annual meeting of the American Society for Mass Spectrometry and Allied Topics, June 2015 in St Lois, MO.

Species Identification Using Bayesian Modeling and Mass Spectrometry Teuble, J, Yang H, Siegel D, Fenyo D, presented at the 64rd annual meeting of the American Society for Mass Spectrometry and Allied Topics, June 2016 in San Antonio, T**X**.